

**Univerzita Karlova, Přírodovědecká fakulta**

**Charles University, Faculty of Science**

**Doktorský studijní program: Parazitologie**

**Doctoral study program: Parasitology**



**RNDr. Štěpánka Hrdá**

**Evoluce jaderných a plastidových genomů u euglenidů**

**Evolution of nuclear and plastid genomes in euglenids**

**Doktorská práce / Doctoral thesis**

**Školitel / Supervisor: doc. Vladimír Hampl, PhD.**

**Praha 2020**



**Prohlášení autorky:**

Prohlašuji, že jsem tuto práci zpracovala samostatně a že jsem uvedla všechny použité zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

**Declaration of the author:**

I declare that I elaborated this thesis independently. I also proclaim that the literary sources were cited properly and neither this work nor a substantial part of it has been used to obtain the same or any other academic degree.

RNDr. Štěpánka Hrdá



### **Prohlášení vedoucího práce:**

Data prezentovaná v této práci jsou výsledkem týmové spolupráce v Laboratoři evoluční protistologie. Prohlašuji, že zapojení a příspěvek Štěpánky Hrdé k této práci bylo podstatné a dostatečné z hlediska rozsahu a kvality pro udělení doktorského titulu.

### **Declaration of the thesis supervisor:**

Data presented in this thesis are the result of team collaboration at the Laboratory of evolutionary protistology. I declare that the involvement and contribution of Štěpánka Hrdá in this work was substantial and sufficient in terms of scope and quality for the award of doctoral degree.

doc. Vladimír Hampl, Ph.D

školitel



## Předmluva

Biologická studie by měla vyprávět nějaký biologický příběh. Obsahem této práce je příběh sekundární endosymbiózy u krásnooček (Euglenophyceae), skupiny fotosyntetizujících prvoků, se kterými se každý seznámí už na základní škole. Původní záměr byl věnovat se plastidovým genomům zástupců této skupiny a jejich srovnáním. To, že toto téma je opravdu nosné, se projevilo záhy po začátku práce a charakterizaci plastomu *Eutreptiella gymnastica*, který byl v roce 2012 teprve třetí osekvenovaný v rámci euglenofyt. Ukázalo se, že několik vědeckých skupin na tématu též pracuje a záhy bylo publikováno dalších 27 plastidových genomů této skupiny.

Práce se začala ubírat trochu jiným směrem a zkusili jsme nahlédnout na sekundární endosymbiózu z více úhlů pohledu - jednak se zaměřit na prasinofytní řasu *Pyramimonas parkeae*, která je potomkem řasy příbuzné donoru plastidu, a jednak prozkoumat také zástupce nefotosyntetických euglen *Rhabdomonas costata*, který reprezentuje potomka euglenida příbuzného původnímu hostiteli řasy. Srovnání některých vlastností a znaků u těchto organismů a u fotosyntetických euglen tvoří mozaiku poznatků, které snad pomohou poodkrýt obraz sekundární endosymbiózy.

Publikace nesouvisí s mou hlavní pracovní náplní v Laboratoři sekvenace DNA a vznikaly tak v poměrně dlouhém období. Poznatky nabyté při zpracování genomických dat do publikací jsem však při své práci v laboratoři mnohokrát využila.

## ...a poděkování

Na tomto místě bych tak ráda poděkovala svému školiteli Vladimíru Hamplovi, za to, že mně umožnil podílet se na svém výzkumu a společně jsme snad přivedli na svět zajímavé publikace. Dále děkuji své kolegyni Blance Hamplové, která spolehlivě zajišťovala chod sekvenační laboratoře, když jsem si někdy potřebovala "hrát" s disertací. Jsem ráda, že jsem léta součástí společenství na katedře parazitologie a děkuji všem za neustávající pomoc. Na závěr děkuji celé své rodině a zejména manželovi Ivanovi a své mamince Evě za vydatnou podporu.





# OBSAH

Abstrakt	3
Abstract	4
1 Úvod do problematiky endosymbiózy	5
1.1 Primární endosymbióza mitochondrií	5
1.2 Primární endosymbióza plastidů	7
1.3 Druhá primární endosymbióza u rodu <i>Paulinella</i>	9
1.4 Endosymbiózy sekundární, terciární a endosymbiózy vyššího stupně	9
1.4.1 Komplexní plastidy odvozené od skupiny Chloroplastida	10
1.4.2 Komplexní plastidy odvozené od skupiny Rhodophyta	11
2 Cíle práce	15
3 Přehled publikací a podíl autorky	15
4 <i>Eutreptiella gymnastica</i> – zástupce opomíjené čeledi euglenofyt	16
4.1 Charakteristika třídy Euglenida	16
4.2 Charakteristika Euglenophyceae	18
4.3 Sekundárního plastid euglenofyt	21
4.4 Obsah genů v plastidovém genomu euglenofyt	22
4.5 Struktura genomu	23
4.6 Introny v plastomech euglenofyt	26
4.7 Maturázy/reverzní transkriptázy u euglenofyt	29
4.8 Plastidový genom <i>Eutreptiella gymnastica</i>	30
4.9 Publikace I	31
5 <i>Pyramimonas parkeae</i> – donor plastidu	33
5.1 Prasinofytní řasy	33
5.2 Plastidové genomy prasinofyt	34
5.3 Plastidový genom <i>Pyramimonas parkeae</i>	35

5.4	Plastidové geny <i>Pyramimonas parkeae</i> u euglenofyt	35
5.5	Mitochondriální genomy prasinofyt	36
5.6	Mitochondriální genom <i>Pyramimonas parkeae</i>	36
5.7	Mitochondriální geny <i>Pyramimonas parkeae</i> u euglenofyt	38
5.8	Publikace II	39
6	Endosymbiotický genový přenos	41
6.1	EGT při primární endosymbióze	41
6.2	EGT při endosymbióze komplexního plastidu	42
6.3	Navrhovaná vysvětlení různorodého původu přenesených genů	44
6.4	EGT u skupiny Euglenida	45
7	Srovnání osmotrofního euglenida <i>Rhabdomonas costata</i> a fototrofního euglenida <i>Euglena gracilis</i>	48
7.1	<i>Euglena gracilis</i> a <i>Rhabdomonas costata</i> – obecná charakterizace	48
7.2	Velikost a struktura jaderného genomu <i>E. gracilis</i> a <i>R. costata</i>	49
7.3	Analýza celkového transkriptomu	50
7.4	Plastidový genom a proteom <i>E. gracilis</i>	53
7.5	Mitochondriální genom <i>E. gracilis</i> a <i>R. costata</i>	54
7.6	Mitochondriální proteom <i>E. gracilis</i>	55
7.7	Mitochondriální proteom <i>R. costata</i> ( <i>in silico</i> )	57
7.8	Publikace III	59
8	Shrnutí	61
9	Literatura	64
10	Příloha	84

## Abstrakt

Řasy tvoří rozmanitou skupinu jednoduchých fotosyntetizujících eukaryotických organismů s polyfyletickým původem. Dělí se na řasy s primárním plastidem (Archaeplastida), který tato eukaryota získala pozřením sinice, tedy prokaryota, a řasy s komplexním plastidem, které získaly plastid pozřením jiného eukaryota s primárním nebo již komplexním plastidem. Řasy s komplexním plastidem jsou chiméry obsahující geny pocházející z hostitelské buňky, a dále geny pocházející nejen z plastidu, ale také z jádra a někdy i mitochondrie endosymbionta, případně i genetický materiál pocházející od jejich předchozích stálých nebo přechodných endosymbiontů. Jednou ze skupin, která získala plastid od zelených řas, jsou krásnoočka (Euglenophyta). Tato práce se zabývá genomy tří organismů, které zastupují jednotlivé aktéry endosymbiotického procesu u euglenofyt. Jsou to původně heterotrofní hostitel ze třídy Euglenida, fototrofní endosymbiont ze třídy zelených řas Prasinophyceae a výsledný fototrofní euglenid ze skupiny Euglenophyceae. Znalost jejich genomů by měla pomoci osvětlit průběh endosymbiotického přenosu genů (EGT) při formování řas s komplexním plastidem.

U zástupce fototrofních euglenidů *Eutreptiella gymnastica* jsme anotovali plastidový genom, který jsme publikovali jako tehdy teprve třetí ve skupině Euglenophyta, po ikonickém a ekonomicky významném druhu *Euglena gracilis* a jeho druhotně nefotosyntetickém protějšku *E. longa*. Ukázali jsme, že obsah genů v plastidových genomech všech euglenofyt je prakticky stejný a že k redukci genomového obsahu o cca 20 % došlo už u jejich společného předka. Dále jsme zjistili, že v plastomu euglenofyta *Etl. gymnastica*, poměrně vzdáleného *E. gracilis*, je velmi málo intronů a tedy že extrémně vysoký počet intronů v plastomu *E. gracilis* není společným znakem celé skupiny, ale znakem druhově specifickým. V neposlední řadě jsme potvrdili, že zelená řasa *Pyramimonas parkeae* je potomkem řasy příbuzné s původním endosymbiontem euglenofyt. U této řasy již byl znám plastidový genom a my jsme jej doplnili o genom mitochondriální, který má stejně jako většina prasinofytních řas poměrně vysoký obsah genů. Přesto zřejmě žádný z nich současná euglenofyta neobsahují a celý tento genomový pool byl při endosymbióze ztracen. Posledním zkoumaným organismem je heterotrofní (osmotrofní) euglenid *Rhabdomonas costata*, jehož částečný jaderný genom a téměř úplný transkriptom jsme porovnávali s dostupným genomem, transkriptomem a mitochondriálním proteomem *E. gracilis*. V jaderném genomu *R. costata* jsme našli nekonvenční introny typické pro Euglenida. V jeho transkriptomu je velmi nízký podíl genů příbuzných genomů fotosyntetizujících organismů a zároveň chybí některé geny, které *E. gracilis* získala prostřednictvím EGT, což vyvrací hypotézu časného získání plastidu u kmene Euglenozoa.

Jednotlivá zjištění nepochybně přispěla k pochopení evoluce třídy Euglenida a do budoucna shromážděná data ještě jistě poslouží v dalších navazujících studiích.

## Abstract

Algae form a diverse group of simple photosynthetic eukaryotes of polyphyletic origin. Algae with a primary plastid (Archaeplastida) acquired it by ingesting cyanobacterium, a prokaryote; algae with a complex plastid acquired their plastid by ingesting another eukaryote with a primary or already complex plastid. Algae with a complex plastid are chimeras containing genes derived from the host genome, as well as genes derived from the genome of the endosymbiont, and also genetic material derived from genomes of their previous stable or transient endosymbionts. One of the groups with plastid derived from green algae are euglenophytes. This thesis deals with the genomes of three organisms that represent individual actors in the endosymbiotic process in euglenophytes. These are a heterotrophic host from the class Euglenida, a phototrophic endosymbiont from the class of green algae Prasinophyceae and the resulting phototrophic euglenid from the group Euglenophyceae. Knowledge of their genomes should illuminate the course of endosymbiotic gene transfer (EGT) in the formation of algae with a complex plastid.

We annotated the plastid genome of a phototrophic euglenid *Eutreptiella gymnastica* and published it as the third plastome of Euglenophytes after the iconic and economically important *Euglena gracilis* and its secondarily non-photosynthetic relative *E. longa*. We have shown that the content of genes in the plastid genomes of Euglenophytes is almost the same and that the reduction of plastid gene content by about 20 % has already occurred in the common ancestor of all euglenophytes. We further found that the plastome of the basal euglenophyte *Etl. gymnastica* has only few introns and therefore the extremely high number of introns in *E. gracilis* plastome is not a common feature of the whole group but a species-specific feature. Last but not least, we confirmed that the green alga *Pyramimonas parkeae* is related to the ancestor of euglenophytes' endosymbiont. The plastid genome of this alga was already known, and we complemented it by the mitochondrial genome with relatively high gene content. None of these genes is probably present in current euglenophytes' genomes, and the entire mitochondrial gene pool of the endosymbiont has been lost during endosymbiosis. The last organism examined in this study is the osmotrophic euglenid *Rhabdomonas costata* whose partial nuclear genome and almost complete transcriptome were compared with the currently available genome, transcriptome and mitochondrial proteome of *E. gracilis*. We found unconventional introns typical for euglenids in *R. costata* nuclear genome. Its transcriptome contains a very low portion of genes related to the genes of photosynthetic organisms and lacks some genes obtained by *E. gracilis* via EGT which disproves the hypothesis of early plastid acquisition in Euglenozoa.

These individual findings undoubtedly contributed to the understanding of the evolution of Euglenida and the collected data will be used in further studies.

# 1 Úvod do problematiky endosymbiózy

Vznik a evoluce eukaryotických buněk úzce souvisí se vznikem jejich organel. Získání mitochondrií a plastidů vedlo k jejich úplné integraci do buňky, která v případě mitochondrií a primárních plastidů proběhla pouze jednou (Gray, 2017; Roger et al., 2017). Zjevné výhody fototrofní výživy vedly k tomu, že plastidy se dále horizontálně šířily a několikrát nezávisle došlo k jejich převzetí skupinami různých nepříbuzných eukaryotů.

## 1.1 Primární endosymbióza mitochondrií

Jednou z hlavních evolučních událostí, které vedly k masivnímu přesunu genů z jedné říše do druhé, byl vznik eukaryotní buňky. Předek dnešních eukaryotních buněk byl pravděpodobně prokaryot z říše Archea z linie Asgard (Roger et al., 2017; Zaremba-Niedzwiedzka et al., 2017), který pohltil prokaryotickou buňku z říše Bacteria z linie  $\alpha$ -proteobakterií (Gray & Archibald, 2012; Yang et al., 1985). Ta dala vznik dnešním mitochondriím. Tento proces nazývaný primární endosymbióza byl několikrát předpovězen (Margulis, 1967; Mereschkowsky, 1905) a následně odmítnut, aby byl nakonec všeobecně přijat díky přesvědčivým molekulárním a ultrastrukturním dokladům. Přesný mechanismus vzniku mitochondrií je ovšem nadále předmětem diskusí.

Během endosymbiogeneze, tedy transformace autonomní bakterie na semiautonomní organelu, muselo dojít k několika událostem, které pravděpodobně neproběhly najednou, ale postupně. Především došlo ke ztrátě většiny genů endosymbionta, které již nebyly potřebné vůbec nebo byly redundantní vzhledem ke genům hostitelské buňky. Dále došlo k přesunu menší části genů nezbytných pro funkci organely do jádra hostitelské buňky, kterýžto proces se nazývá endosymbiotický genový přenos (EGT – endosymbiotic gene transfer) (Roger et al., 2017). K tomuto procesu docházelo v určitých vlnách a redukce mitochondriálního obsahu se exponenciálně zpomalovala (Janouškovec et al., 2017). Z hlediska fungování chiméry endosymbiont/hostitel bylo nezbytné sladit jejich biochemické dráhy a koordinovat dělení, a dále zajistit oboustranný přenos malých molekul (ATP, aminokyselin, iontů a dalších). Pro zpětný přenos proteinů nutných pro fungování organely, které byly nově transkribovány v jádře a translatovány v cytoplasmě, vznikly komplexy proteinů v membránách mitochondrie translokony ("Translocase of the Outer (TOM) and the Inner (TIM) Mitochondrial membrane"), jejichž některé komponenty jsou prokaryotického a některé eukaryotického původu (Roger et al., 2017). Zároveň proteiny určené pro transfer musely být osazeny krátkou aminokyselinovou sekvencí (tzv. target peptide – TP), dlouhou průměrně 70 aminokyselin, která proteiny navádí přes translokony dovnitř mitochondrie, a poté je odstraněna peptidázami. Je možné, že se vyvinuly z krátkých peptidů, které eukaryotické buňky syntetizují jako obranu proti patogenům a z nichž některé jsou jim strukturně i sekvenčně podobné. Tyto peptidy mají za úkol dezintegrovat membrány prokaryotických patogenů. Jednou z protistrategií prokaryot je využití ABC transportérů, které tyto peptidy převedou bez porušení membrány do lumen prokaryota, kde jsou rozbity peptidázami. Takovéto ABC transportéry tak mohly být předchůdci translokonů (Wollman, 2016).

Má se za to, že primární endosymbióza mitochondrií proběhla pouze jednou a to pravděpodobně před dvěma miliardami let. Ačkoli je nepochybné, že mitochondrie pochází z nějaké  $\alpha$ -proteobakterie, její přesná fylogenetická pozice nebyla dosud určena (Roger et al., 2017). Dlouhou dobu se zdálo, že mitochondriální genom se větví společně s linií dnešních intracelulárně žijících Rickettsiales (Wang & Wu, 2015), nebo s linií volně žijících mořských bakterií Pelagibacterales (SAR11) (Thrash et al., 2011). V nejnovější analýze, která využívá i data z oceánického metagenomového datasetu (Tara projekt) a která se pokouší minimalizovat vliv problematických fylogenetických artefaktů, jsou předchozí nálezy zpochybněny a mitochondrie je postavena jako sesterská linie všem  $\alpha$ -proteobakteriím (Martijn et al., 2018).

Mitochondriální genomy jsou velice variabilní co do počtu genů. Obsahují dva geny u *Chromera velia* až cca sto původních bakteriálních genů u jakobidů (Burger et al., 2013; Flegontov et al., 2015). Nejkompletnější set původně bakteriálních genů (99) byl nalezen u prvoka *Andalucia godoy*, zástupce jakobidů, v jejichž genomu lze vždy nalézt subset ze 101 genů, pocházejících od jejich společného předka, který měl podle předpokladu okolo 3000 genů (Burger et al., 2013; Roger et al., 2017). Různé linie organismů obsahují i další geny nepopsané u jakobidů. K redukci genomu tedy docházelo paralelně a ztráty genů byly víceméně náhodné, přestože určitá konzervovaná část genů je ve všech mitochondriálních genomech přítomná (Janouškovec et al., 2017). Zároveň během vývoje docházelo také k získání genů horizontálním genovým přenosem (HGT) od nepříbuzných organismů, např. ve všech mitochondriálních genomech s výjimkou jakobidů se nachází fágová T4 RNA polymeráza (Roger et al., 2017).

Mitochondriální DNA tvoří většinou kruhová molekula, byť v mnoha skupinách byly nalezeny i lineární formy, které se objevily nezávisle u různých skupin, a i blízce příbuzné druhy se mohou v tomto znaku lišit (Pérez-Brocal et al., 2010). Některé linie eukaryot (Fungi, Stramenopiles, Chlorophytes) také obsahují dlouhé invertované repetice (IR) obsahující ribozomální operon, které jsou jinak typické pro plastidové genomy (viz dále). Přítomnost IR je dáována do souvislosti s vysokým počtem tzv. G-quadruplex struktur, které mohou rozvolňovat DNA helixy a mimo jiné způsobovat zlomy v DNA. Invertované repetice by pak mohly hrát roli opravného mechanismu, snad pomocí homologní rekombinace (Gualberto et al., 2014; Nieuwenhuis et al., 2019).

Učebnicovou funkcí mitochondrie je tvorba ATP oxidativní fosforylací. Mitochondrie však mají mnohem větší škálu funkcí (např. metabolismus aminokyselin, nukleotidů, mastných kyselin a lipidů, biogeneze železosírných klastrů nebo degradace glycinu), které se často liší mezi jednotlivými liniemi (Roger et al., 2017). U některých organismů je mitochondrie redukována a zachovává si jen některé funkce, které ani nemusí souviset s energetickým metabolismem. Obecně jsou mitochondrie a jim podobné organely nazývány "Mitochondrion-related organelles" (MROs) a jsou rozdělovány do několika funkčních typů (Muller et al., 2012), které ale *de facto* tvoří kontinuum organel s různými funkcemi (Roger et al., 2017). Výjimečně mitochondrie zcela chybí (Karnkowska et al., 2016), nebo je sice přítomná, ale postrádá vlastní genom a všechny její geny jsou kódované v jádře (John et al., 2019).

## 1.2 Primární endosymbióza plastidů

Velmi podobným způsobem jako primární endosymbióza mitochondrií proběhla později i primární endosymbióza plastidů. Doklady nasvědčují tomu, že hostitelem již byla vyvinutá eukaryotická buňka s plně funkčním fagocytickým aparátem. Endosymbiontem byla sinice, která si zachovala dvě membrány, které dnes najdeme u všech primárních plastidů, a podařilo se jí uniknout z fagozomu. Přestože dnes známe tři linie organismů s různým typem primárních plastidů (Glaucophyta, Rhodophyta a Chloroplastida), podle posledních fylogenetických studií patří všechny do monofyletické skupiny Archaeplastida, tzn. že k endosymbiotické události došlo u této skupiny pouze jednou (Gray, 2017), a to před 1,2 – 1,6 miliardami let (Löffelhardt, 2014). Chloroplastům je momentálně nejpodobnější dnešní sladkovodní sinice *Gloeomargarita litophora*, což vede i k závěru, že k endosymbiotické události nedošlo v mořském prostředí (Ponce-Toledo et al., 2017).

V hlavních rysech se endosymbiogeneze plastidů neliší od té mitochondriální, také vnější a vnitřní membrány chloroplastu musely být vybaveny komplexem translokáz, zde nazývaných TOC a TIC ("Translocons at the Outer (TOC) and Inner (TIC) envelope membranes of Chloroplasts"). Preproteiny přenášené z cytoplasmy do chloroplastu jsou také osazené specifickou chloroplastovou naváděcí sekvencí (target peptid – TP), která je ovšem o něco kratší než u mitochondrií a je dlouhá průměrně 55 aminokyselin (Wollman, 2016). Systém translokáz i signální peptidy jsou však u mitochondrií a plastidů nehomologní.

Všechny primární plastidy obsahují stejné translokační proteiny TOC a TIC. Některé z nich mají prokazatelně ne-sinicový původ a jejich komplexita prakticky vylučuje možnost paralelního vzniku, což je také hlavní argument pro monofyletický původ všech archaeplastid. Pro jednotný původ archaeplastid jednoznačně hovoří také některé další společné znaky, které u dnešních sinic nenalezneme, např. určité typické genové klastry nebo protein vázající chlorofyl s inovativně ztrojeným transmembránovým helixem (Keeling, 2010; McFadden & Van Dooren, 2004).

Důležitým znakem plastidových genomů obecně je jejich uspořádání do čtyř částí, z nichž dvě jsou shodné, opačně orientované ("Inverted repeats – IR"), a střídají se s jedním delším a jedním kratším unikátním úsekem ("Large and small single copy regions – LSC, SSC"). Toto uspořádání je pravděpodobně důležité pro zahájení replikace plastidu, stabilizaci genomu a konzervaci genů (Turmel et al., 2017). V IR úseku se nalézá ribozomální operon, který je tak v genomu přítomen dvakrát. Ribozomální operon obvykle obsahuje 5 genů pro rRNA a tRNA v pořadí 16S, *trnI*, *trnA*, 23S, 5S. Invertované repetice se velice liší svou délkou, mohou obsahovat jen 5 zmíněných genů (*Pseudoclonium akinetum*) nebo expandovat až na celkem 51 genů (*Pelargonium X hortorum*) (Chumley et al., 2006). Délka IR se liší i mezi blízkými příbuznými druhy. I v rámci jednoho jedince byly popsány různé isomery plastomů, u kterých se liší orientace jednotlivých částí vůči sobě navzájem (Berriman et al., 2005; Wang et al., 2019). IR nemusí být zcela identické (časté u *Ulvothyceae*), mohou být v přímé orientaci (*Porphyra*) a mnohokrát byla jedna kopie IR nezávisle ztracena (Turmel et al., 2017).

Genomy primárních plastidů obsahují různý počet genů ve škále od několika desítek genů u organismů postrádajících fotosyntézu po 250 genů u červených řas. Představuje to jen několik málo procent z genů nacházejících se v současných sinicových genomech. Je tedy zřejmé, že velká část původních genů sinice byla buď ztracena, nebo přenesena endosymbiotickým genovým přenosem do jádra hostitelské buňky (Archibald, 2015; Smith & Keeling, 2015). Plastidové geny zajišťují především fotosyntézu, transkripci a translaci těchto genů, ale i u plastidů lze v proteomu najít proteiny, zajišťující další funkce. Jedná se především o fotorespiraci, syntézu chlorofylu, škrobů ( $\alpha$  glukany), mastných kyselin, lipidů, aminokyselin a vitamínů (Wise & Hooper, 2006). Stejně jako u mitochondrií existuje i u chloroplastů kontinuum funkčních typů, z nichž některé ztratily schopnost fotosyntézy, ale zachovaly si některé z dalších plastidových funkcí a jsou tak pro svého nositele nepostrádané (např. syntéza prekursoru hemu, mastných kyselina a některých aminokyselin nebo skladování škrobu) (Gray & Archibald, 2012; Hadariová et al., 2018; Janouškovec et al., 2015). Stejně jako u mitochondrií i u primárních plastidů lze najít příklady horizontálního genového přenosu; např. u ruduchy *Gracilaria tenuistipitata* se vyskytuje klastr genů pro syntézu leucinu, pocházející z proteobakterií (Janouškovec et al., 2013); plastomy některých eustigmatofytních řas zase získaly klastr šesti genů od bakterií z kmene *Bacteroidetes* (Yurchenko et al., 2016).

Jednotlivé linie archaeplastidů vykazují mezi sebou několik zásadních odlišností. Málo početná skupina sladkovodních řas Glaucophyta, zastoupená osmi rody, je nejpodobnější dnešním sinicím svou stavbou. Nese např. mezi svými dvěma membránami relikty peptidoglykanové stěny a obsahuje chlorofyl *a* a modrý fykocyanin. Plastidové genomy obsahují 185 – 206 unikátních genů. Skupina se oddělila od ostatních archaeplastidů jako první (Jackson et al., 2015).

Mnohem početnější a rozmanitější skupinou jsou Rhodophyta, která jsou zastoupena 5000 – 6000 druhy od nepatrných kulovitých extrémofilních buněk po makroskopické mořské i sladkovodní řasy. Obsahují chlorofyl *a* a červený fykoerythrin, a na rozdíl od zelených řas jsou u nich obě podjednotky proteinu RuBisCO, stejně jako u glaucofytů, kódované v plastidu. Dosud charakterizované plastomy obsahují 220 – 250 velice kompaktně organizovaných unikátních genů, což je nejvyšší počet u všech známých plastidových genomů a nejvíce se tak blíží k původnímu vzhledu genomu primárního plastidu (Janouškovec et al., 2013).

Zdaleka nejpočetnější a nejvariabilnější skupinou jsou Chloroplastida (Viridiplantae) s mnoha desítkami tisíc druhů, zahrnující Chlorophyta - zelené řasy a Streptophyta – vyšší rostliny a některé jim příbuzné sladkovodní zelené řasy (Adl et al., 2019). Základním znakem celé skupiny je přítomnost škrobu a chlorofylů *a* a *b* v plastidech, v jejichž genomu není kódována malá subjednotka RuBisCO, neboť tento gen byl přenesen v rámci EGT do jaderného genomu. Plastidové genomy obsahují obvykle 95 – 125 unikátních genů, nejméně (54 unikátních genů) je v plastomu parazitické zelené řasy *Helicosporidium* (Sun & Pombert, 2014; Turmel et al., 2017). Skupina se vyznačuje častou ztrátou IR v celých liniích (Turmel et al., 2017).



### 1.3 Druhá primární endosymbióza u rodu *Paulinella*

K primární endosymbióze došlo v evoluční historii nezávisle ještě jednou. Améba *Paulinella* (Rhizaria; Cercozoa) nese jedno nebo dvě modrozelená tělíska nazývaná chromatofory nebo cyanely, které připomínají sinice a umožnily paulinelám přechod od fagotrofní k fototrofní výživě. U tohoto organismu již lze hovořit o změně endosymbionta v organelu, vztah paulinel a jejich cyanel je obligatorní a buňky samostatně nepřežívají. Cyanely mají skutečně sinicový původ, ale na rozdíl od primárních plastidů archaeplastid jsou monofyletické se skupinou sinic *Prochlorococcus/Synechococcus* (Nowack et al., 2008).

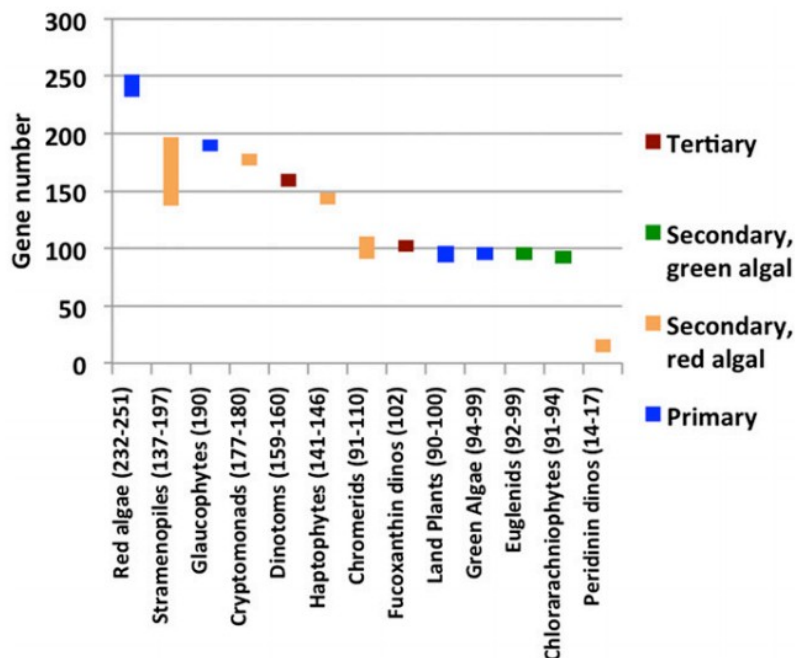
Dosud byly osekvenovány genomy cyanel tří druhů rodu *Paulinella*, dvou sladkovodních a jednoho mořského. Genomy cyanel obsahují cca 860 – 880 genů a jsou tak pouze třetinové až čtvrtinové v porovnání s genomem sinice rodu *Synechococcus*. K pohlcení sinice došlo pravděpodobně před cca 120 miliony let a následovala velice rychlá ztráta genů. K rozrůznění jednotlivých druhů pak docházelo postupně v posledních 60 milionech let, kdy se ovšem rychlost ztrácení genů velice zpomalila (Lhee et al., 2019).

### 1.4 Endosymbiózy sekundární, terciární a endosymbiózy vyššího stupně

Kromě výše zmíněných archaeplastid s primárním plastidem existuje ještě celá řada dalších fotosyntetizujících eukaryotních organismů, které jsou tradičně také nazývány řasami. Tyto organismy náležejí k různým fylogenetickým skupinám. Získaly svůj plastid pohlcením jiného eukaryota majícího primární plastid, takový proces se nazývá sekundární endosymbióza, či pohlcením eukaryota se sekundárním plastidem, pak se jedná o terciární endosymbiózu. V některých liniích zřejmě došlo k několikanásobnému postupnému převzetí plastidu a pak hovoříme o endosymbiózách vyššího řádu. Výsledkem je celé spektrum buněčných typů, které nesou plastidy obalené většinou třemi nebo čtyřmi membránami (výjimečně dvěma nebo pěti), některé mají zachovaný i zbytek jádra, tzv. nukleomorf, výjimečně je zachována i mitochondrie pohlceného eukaryota (Curtis et al., 2012; Dorrell & Howe, 2015; Imanian et al., 2010; Minge et al., 2010; Rogers et al., 2007; Wetherbee et al., 2019). Zvýšení počtu plastidových membrán přineslo další inovaci, proteiny musely být na začátku vybaveny další naváděcí sekvencí. Ta je obdobou signálního peptidu (SP), který mívají sekretované proteiny a který je rozpoznáván např. na membráně endoplasmatického retikula (Sheiner & Striepen, 2013). Tato signální sekvence převádí preproteiny přes vnější membránu hostitelského původu a je následně odštěpena. Přes druhou membránu hostitelského původu, pokud je přítomna, proteiny přenáší u řas s odvozeným červeným plastidem tzv. SELMA komplex (viz dále), u řas s odvozeným zeleným plastidem je jejich přenos nedostatečně prozkoumán. Odhalená sekvence TP směřuje preprotein přes dvě vnitřní membrány odvozené od původního primárního plastidu pomocí částečně redukováných TOC resp. TIC translokonů (Bolte et al., 2009; Füssy & Oborník, 2018).

Stejně jako při primární endosymbióze byla velká část genů přesunuta do jádra nového hostitele, a to jak z jádra endosymbionta, tak z jeho plastidu. Genomy plastidů vyšších stupňů

endosymbiózy tak mají přirozeně méně genů, než genomy primárních plastidů (obr. 1) (Dorrell & Howe, 2015). Vějíř organismů s komplexními plastidy je značně široký, ale u mnoha znaků je popsána konvergentní evoluce (Hirakawa, 2017).



Obr. 1. Obsah genů v plastomech různých linií fotosyntetických organismů (převzato z Dorrell & Howe, 2015).

#### 1.4.1 Komplexní plastidy odvozené od skupiny Chloroplastida

Sekundární plastidy u tří nezávislých linií eukaryot (Euglenophyta, Chlorarachniophyta a některé obrněnky) obsahují chlorofyl *a* a *b* a tedy pocházejí ze zelených řas. Fylogenetická srovnání ukázala, že ani tyto plastidy nejsou odvozeny od příbuzných linií zelených řas (Jackson et al., 2018; Keeling, 2010).

**Euglenophyta** (Excavata) jsou sladkovodní i mořští bičíkovci. Jejich chloroplast získaný před 650 – 540 miliony let má tři membrány a má původ v zelené řase podobné dnešní řase *Pyramimonas parkeae* (Jackson et al., 2018; Turmel et al., 2009). Všechny tři membrány k sobě těsně přiléhají, není mezi nimi periplastidiální prostor ani nukleomorf. Preproteiny se přes vnější membránu dostávají fúzí Golgiho váčku s membránou. Přenos přes prostřední membránu není objasněn, ale mohl by se díť za účasti derlin-like proteinů, ale bez účasti TOC přenašečů, což by svědčilo pro ztrátu vnější membrány pocházející z primárního plastidu. Přes vnitřní membránu se pravděpodobně dostávají velice redukováným komplexem TIC přenašečů (Novák Vanclová et al., 2019). Bližšímu popisu skupiny je věnována samostatná kapitola.

**Chlorarachniophyta** (Rhizaria; Cercozoa) jsou mořské mixotrofní améby. Jejich chloroplast pochází z předka řas se sifonální mnohjadernou stélkou *Bryopsidales*. K endosymbióze došlo

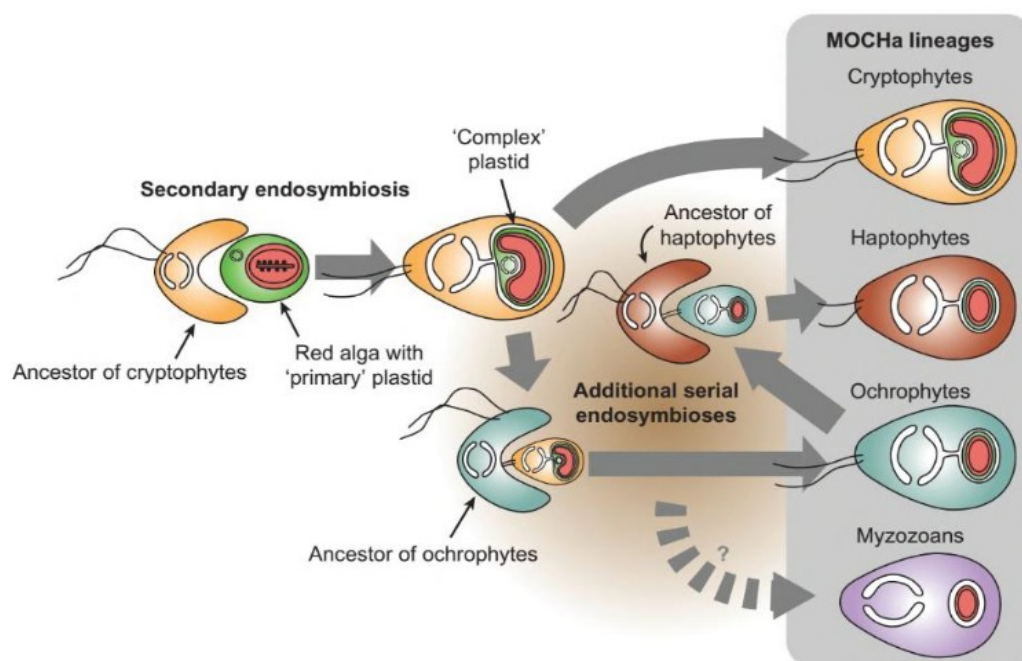
před 578 – 318 miliony let a není jasné, zda předkem plastidu byl jednobuněčný předek trubicovek, nebo zda plastid vznikl pohlcením části poškozené stélky již sifonální řasy (Jackson et al., 2018). Chloroplast je obalen čtyřmi membránami. Dvě vnitřní obalují samotný plastid, který je obklopen periplastidiálním prostorem, v němž se nachází ještě nukleomorfy s vlastním genomem. Plastidový genom obsahuje dvě invertované repetice a 87 unikátních genů bez intronů kódujících plastidové proteiny a RNA (Rogers et al., 2007). Genom nukleomorfy druhu *Bigeloviella natans* byl osekvenován a obsahuje 331 genů, z nichž 17 kóduje plastidové proteiny. Mezi nimi lze najít také geny pro některé translokační proteiny TIC/TOC, díky kterým se proteiny nesoucí N-terminální targetovací peptid, které jsou kódovány v nukleomorfu a také v jádře sekundárního hostitele, dostávají přes dvě vnitřní membrány do plastidového stroma. Proteiny kódované v jádře sekundárního hostitele jsou na začátku navíc vybaveny signálním peptidem (Gilson et al., 2006).

Třetí nezávislá sekundární endosymbióza zelené řasy byla popsána u **obrněnky rodu *Lepidodinium*** (Minge et al., 2010), která je zastoupena dvěma druhy a několika izoláty. Obrněnky ovšem obvykle mají plastid odvozený od červených řas (viz dále). Zelený plastid v tomto případě nahradil červený plastid a hovoříme zde o tzv. seriální sekundární endosymbióze. Tento zelený plastid je odvozený od velmi malé jednobuněčné řasy *Pedinomonas minor* (Jackson et al., 2018). Je také obalen čtyřmi membránami a obsahuje strukturu podobnou nukleomorfu (Waller & Kořený, 2017). O tom, zda tato struktura nese i DNA, není více známo. Plastidový genom neobsahuje na rozdíl od *P. minor* invertované repetice a kóduje 87 unikátních genů kódujících plastidové proteiny a RNA, což je o 18 méně než u *P. minor* (Kamikawa et al., 2015). Proteiny určené do plastidu mají bipartitní presekvenci, přičemž signální peptid (SP) ani target peptid (TP) nejsou podobné typickým SP a TP zelených řas ani ostatních obrněnek (Minge et al., 2010; Patron et al., 2005). V roce 2020 byly popsány další dva navzájem příbuzné kmeny obrněnek se zeleným plastidem nazvané MGD a TGD, které ale nejsou blízce příbuzné rodu *Lepidodinium* (Sarai et al., 2020). Jejich plastid je také odvozen od *P. minor*, ale byl pravděpodobně získán později a endosymbióza ještě nemusí být dokončena (Nakayama et al., 2020; Sarai et al., 2020)

#### 1.4.2 Komplexní plastidy odvozené od skupiny Rhodophyta

K sekundární endosymbióze plastidu odvozeného od červených řas došlo také u několika skupin organismů, ale situace zde není tak přehledná, jako u skupin se zeleným plastidem. Není zřejmé, zda a u kterých organismů došlo k získání plastidu nezávisle a u kterých mohlo dojít k seriálnímu předání plastidu, přičemž zjevné je pouze to, že červený plastid od rhodofyt nebyl získán pouze jednou společným předkem všech dotčených linií, tak jak to předpokládala dnes odmítnutá chromalveolární hypotéza (Baurain et al., 2010; Burki et al., 2020; Keeling, 2010). Situaci znepřehledňují četné nefotosyntetizující organismy, příbuzné známým fotosyntetizujícím liniím, které někdy mají, ale často nemají ani zbytek plastidu s redukováným genomem a není jisté, zda plastid mohly mít a ztratily jej sekundárně, nebo ho nikdy neměly.

Komplexní červený plastid, nebo jeho zbytek, je přítomen u skupin Cryptophyta, Ochrophyta, Haptophyta a alveolárních skupin Dinoflagellata, Apicomplexa, Chromerida a Colpodelida (tzv. Myzozoa) (Burki, 2017). U všech zmíněných linií byla v plastidu zjištěna přítomnost nového typu chlorofylu *c*. Současný převládající názor předpokládá, že k sekundární endosymbióze došlo u skrytěk, které obsahují nukleomorf jako pozůstatek jádra rhodofyt (Douglas et al., 2001; Griffin & Aken, 1990; R. B. Moore et al., 2008). Skrytěnky pak byly donorem pro ostatní skupiny přímo, nebo nepřímo přes ochrofyta (Burki, 2017; Dorrell & Smith, 2011; Janouškovec et al., 2010; Ševčíková et al., 2015; Stiller et al., 2014). Jeden z možných scénářů je znázorněn na obrázku 2.



Obr. 2. Jeden z možných scénářů původu a evoluce komplexního červeného plastidu seriální endosymbiózou z eukaryota do eukaryota (převzato z Burki, 2017).

Vnější membrány jsou u skrytěk, ochrofyt a haptofyt přímo spojeny s endomembránovým systémem hostitele, s nímž tvoří tzv. chloroplastové endoplasmatické retikulum (Flori et al., 2016). Tím je vyřešen první krok přenosu preproteinu, neboť je syntetizován přímo do jeho lumen. U alveolát jsou preproteiny přepravovány do chloroplastu pravděpodobně ve váčcích pocházejících z Golgiho systému (Dinoflagellata) nebo ER (Apicomplexa, Chromerida), které s vnější membránou chloroplastu fúzíjí (Bolte et al., 2009; Füssy & Oborník, 2018) Přes druhou vnější membránu jsou pak transportovány tzv. SELMA komplexem (symbiont-specific ERAD-like machinery), který tvoří kanál z několika tzv. derlin proteinů. SELMA komplex vznikl duplikací komplexu přenašečů původně zajišťujících transport špatně sbalených proteinů určených k degradaci (ERAD systém) (Hempel et al., 2009). Systém SELMA je typický pro komplexní červené plastidy s výjimkou dinoflagelát, které postrádají příslušnou membránu. Dvě vnitřní membrány obsahují TOC/TIC proteiny v různém stupni odvození (Sheiner & Striepen, 2013).

**Cryptophyta** je monofyletická skupina fotosyntetizujících bičíkovců náležejících společně s dalšími heterotrofními bičíkovci, kteří nikdy plastid neměli (např. *Goniomonas avonlea*), do superskupiny Cryptista (Burki et al., 2020; Cenci et al., 2018). Mají plastidy obalené čtyřmi membránami s nukleomorfem a obsahují chlorofyl *a* a *c2* a fykobiliproteiny (Adl et al., 2019). Plastidový genom většinou obsahuje IR, které jsou někdy neidentické. Počet genů v plastomu je okolo 180 u fotosyntetizujících skrytěnek, nefotosyntetizující druh *Cryptomonas paramecium* má jen 114 genů. Stejně jako u červených řas jsou velká i malá podjednotka Rubisco kódované v plastidu, což platí i pro *C. paramecium*, který jinak další geny pro fotosyntetické geny ztratil (Kim, Yoon, et al., 2015). Genom nukleomorfu osekvenovaný doposud u čtyř druhů obsahuje cca 470 – 500 genů, které kódují proteiny, z nichž 20 - 30 kóduje plastidové proteiny (Gilson et al., 2006; C. E. Moore et al., 2012; Tanifuji et al., 2011; Zauner et al., 2019).

**Ochrophyta** je monofyletická skupina fotosyntetizujících velice různorodých jedno- i mnohobuněčných organismů ze superskupiny Stramenopiles, která se vyznačuje dvěma nestejnými tzv. heterokontními bičíky přítomnými alespoň v některé části životního cyklu (Riisberg et al., 2009). Nesou červené plastidy obalené čtyřmi membránami a obsahují chlorofyly *a* a většinou *c1*, některé také *c2* a *c3* a velice různorodou směs dalších plastidových pigmentů (Adl et al., 2019). Plastidy kódují cca 140 – 170 genů a mají IR. Jednou z nejlépe prozkoumaných větví ochrofyt jsou rozsivky (Bacillariophyceae). Výjimečný je nedávno popsán stramenopilní rod *Chrysoparadoxa*, který má plastid pouze se dvěma membránami (Wetherbee et al., 2019).

**Haptophyta** je monofyletická skupina fotosyntetizujících bičíkatých organismů náležejících společně s heterotrofními centrohelidy do superskupiny Haptista (Burki et al., 2020). Mají plastidy se čtyřmi membránami, které obsahují chlorofyl *a*, *c1*, *c2* a *c3*, a jsou bez nukleomorfu (Adl et al., 2019). Plastidové genomy obsahují kolem 140 genů, ribozomální operon je buď samostatný nebo v repetici, která je buď invertovaná nebo přímá (Méndez-Leyva et al., 2019).

Myzozoa zahrnují ty alveolární skupiny, které se živí nebo živily myzocytózou, při níž není kořist pohlcena celá, ale buněčný obsah je vysátý. Zahrnuje skupiny Apicomplexa se sesterskou parafyletickou skupinou Chrompodellida (Chromerida a Colpodellida) a Dinoflagellata a jim blízce příbuzný parazitický rod *Perkinsus*. Předpokládá se, že tyto skupiny mají nebo měly komplexní plastid (Janouškovec et al., 2015).

Fotosyntetický plastid si zachovala **Chromerida**. Tato linie zahrnuje jednobuněčné volně žijící fotosyntetizující organismy asociované s korály, *Chromera* a *Vitrella*, které mají červený plastid obalený čtyřmi membránami. Jejich plastomy obsahují přes 140 genů, z nich ty, zapojené do transkripce a translace, nalézáme i u aplicomplex (Janouškovec et al., 2010; Oborník & Lukeš, 2015). Velice blízce příbuznou skupinou chromeridů jsou **Colpodellida**, dravé jednobuněčné organismy, u nichž plastid nebyl identifikován, ale v jejichž transkriptomu jsou přítomné geny, které by mohly být směřovány do kryptického nefotosyntetického plastidu (Gile & Slamovits, 2014). V transkriptomu byly také potvrzeny geny pro plastidovou syntézu tetraprolů a izoprenoidů (Janouškovec et al., 2015).

**Apicomplexa** představují jednu z hlavních větví superskupiny Alveolata. Jsou to obligátní parazité, kteří nesou bezbarvý nefotosyntetizující plastid (apikoplast), obalený čtyřmi membránami, který je ovšem nepostradatelný kvůli zapojení do syntézy isoprenoidů, hemu, lipidů a železo-sírných klastrů (Lim & McFadden, 2010). Genom apikoplastu je redukován a obsahuje jen okolo 65 genů, které nejsou spojeny s fotosynézou. Na rozdíl od mnoha nefotosyntetizujících organismů se zbytkovým plastidem nemají ve svém genomu geny pro podjednotky Rubisco.

**Dinoflagellata** tvoří další hlavní větev alveolát. Jsou to jednobuněčné mixotrofní organismy se dvěma bičíky, které mají původně sekundární červené plastidy obalené třemi membránami obsahující chlorofyl *a* a *c2* a často i specifický xantofylní pigment peridin (Keeling, 2010). Obrněnky s tímto tzv. peridinovým plastidem jsou považovány za původní typ (např. rod *Gymnodinium*) a ostatní typy jsou od něj odvozené.

Vzhledem k tomu, že Dinoflagellata tvoří sesterskou skupinu k dvojici Apicomplexa a Chrompodellida, lze usuzovat, že stejný plastid byl přítomen u společného předka všech těchto linií (Janouškovec et al., 2010; Woo et al., 2015). Plastidový genom obrněnek je oproti plastomu chromery velice redukován, zůstaly v něm však naopak pouze geny pro rRNA a geny pro proteiny fotosystému I a II, cytochromu a ATP syntáz. Plastidový genom netvoří jedno kruhové vlákno, ale tzv. minikroužky, které nesou kromě téměř shodné kratší nekódující oblasti vždy jeden nebo několik málo genů, celkově se v genomu nachází okolo 15 genů (Dorrell & Howe, 2015). U obou linií Chromerida a Dinoflagellata jádro kóduje také velkou podjednotku RuBisco typu II, která byla získaná druhotně horizontálním genovým přenosem z červených sírných bakterií *Rhodospirillum rubrum* (Dorrell & Howe, 2015; Tabita et al., 2008).

Mnohé druhy obrněnek nejsou fotosyntetické. Jsou mezi nimi jednak volně žijící zástupci (např. rod *Oxyrrhis*, *Noctiluca*), jednak parazité (např. rod *Hematodinium* a také obrněnkám příbuzný rod *Perkinsus*) a jednak obrněnky, které si po nějaký čas drží přechodný fotosyntetický plastid – "kleptoplastid" (*Dinophysis*). U většiny z nich je předpoklad, že mají kryptický plastid. Mají plastidové dráhy pro syntézu izoprenoidů a tetrapyrolů, u druhu *Perkinsus marinus* jen isoprenoidů. Výjimku tvoří *Hematodinium*, které ztratilo plastid docela a v jejím genomu se našlo jen několik málo původem plastidových genů, jejichž funkce byla přesměrována do cytosolu (Janouškovec et al., 2017; Waller & Kořený 2017).

U obrněnek kromě druhotné ztráty plastidu došlo také několikrát nezávisle k jeho výměně, kdy peridinový plastid byl nahrazen jiným plastidem a před nebo během této endosymbiózy byl původní plastid ztracen. Takové endosymbiózy nazýváme terciární či kvartérní a tyto komplexní plastidy jsou obaleny čtyřmi membránami u endosymbiontů pocházejících ze skupiny Haptophyta (*Kareniaceae*), nebo pěti membránami v případě pohlčení celých rozsivek (*Kryptoperidinium*, *Durinskia*). Zástupci Dinofhsiales i dalších tříd obrněnek si drží tzv. kleptoplastid získaný z kořisti, který si musí čas od času znovu opatřit. Častým zdrojem jsou skrytěnky, ale i další řasy s primárním nebo komplexním plastidem (Waller & Kořený, 2017).

## 2 Cíle práce

1. Charakterizovat plastidový genom *Eutreptiella gymnastica*, euglenofyta velmi vzdáleně příbuzného modelového druhu *E. gracilis*.

2. Charakterizovat mitochondriální genom *Pyramimonas parkeae*, žijícího příbuzného předka plastidů euglenofyt.

3. Na základě genomických a transkriptomických dat charakterizovat vybrané vlastnosti heterotrofního euglenida *Rhabdomonas costata* a porovnat je s vlastnostmi *Euglena gracilis*.

## 3 Přehled publikací a podíl autorky

Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012). The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. PLoS ONE, 7(3), e33746. <https://doi.org/10.1371/journal.pone.0033746>

Podíl autorky: anotace genomu a příprava rukopisu.

Hrdá, Š., Hroudová, M., Vlček, Č., & Hampl, V. (2017). Mitochondrial genome of prasinophyte alga *Pyramimonas parkeae*. Journal of Eukaryotic Microbiology, 64(3), 360-369. <https://doi.org/10.1111/jeu.12371>

Podíl autorky: anotace genomu a příprava rukopisu.

Soukal, P., Hrdá, Š., Karnkowska, A., Milanowski, R., Szabová, J., Hroudová, M., Strnad, H., Vlček, Č., Čepička, I., & Hamp, V. (2020). Genome and transcriptome draft of the heterotrophic euglenid *Rhabdomonas costata*. Rukopis v přípravě.

Podíl autorky: příprava transkriptomu, anotace mitochondriálního proteomu získaného *in silico* z celkového transkriptomu, rekonstrukce metabolických drah a podíl na přípravě rukopisu.

Vanclová, A. M. G., Hadariová, L., Hrdá, Š., & Hampl, V. (2017). Secondary plastids of Euglenophytes. In Y.Hirakawa (Eds.), *Secondary endosymbioses. Advances in Botanical Research*, 84, 321–358. <https://doi.org/10.1016/bs.abr.2017.06.008>

Podíl autorky: sepsání části kapitoly věnované genomům euglenofyt.

## 4 *Eutreptiella gymnastica* – zástupce opomíjené čeledi euglenofyt

Euglenofyta je společné pojmenování pro tři autotrofní řády z třídy Euglenida, které mají zelený sekundární plastid. Třída Euglenida zahrnuje jak tyto fotosyntetizující organismy s plastidem, tak jejich sesterské heterotrofní skupiny. Porovnání genomů těchto dvou skupin, stejně jako porovnání genomů v rámci euglenofyt může přinést detailnější vhled do procesu sekundární endosymbiózy. U zástupce fototrofních euglenidů *Eutreptiella gymnastica* jsme anotovali plastidový genom a porovnali ho se známým plastidovým genomem *Euglena gracilis*.

### 4.1 Charakteristika třídy Euglenida

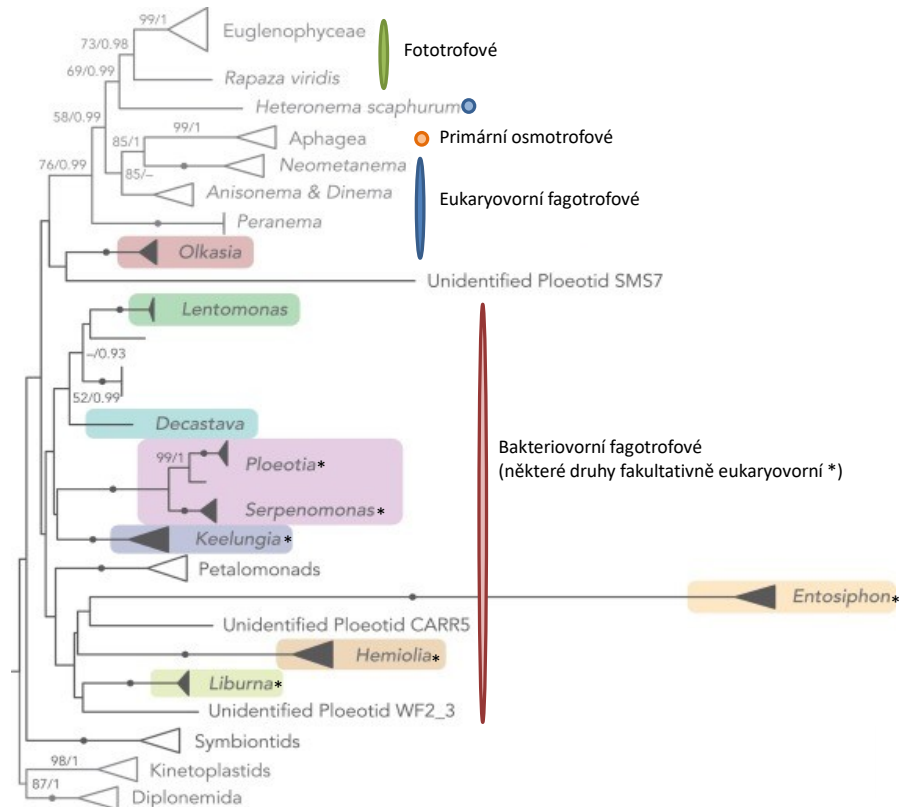
Třída Euglenida dnes patří spolu s dalšími třemi třídami Symbiontida, Diplonemea a Kinetoplastea do kmene Euglenozoa superskupiny Excavata (Adl et al., 2019; Cavalier-Smith, 1981). Euglenozoa jsou jednobuněčné organismy s typicky dvěma, někdy jen jedním bičíkem, které vycházejí z periflagelární kapsy, a jeden nebo oba jsou vyztuženy paraxonemální tyčí; Euglenozoa mívají komplexní potravní aparát zesílený mikrotubuly lišící se podle typu výživy, trojdílný kořenový systém bičíku a pelikulu – plasmatickou membránu podloženou vrstvami specializovaných proteinů, mikrotubuly a ER (Adl et al., 2019; Leander, 2004). Euglenozoa zahrnují především volně žijící, ale i parazitické organismy (Kinetoplastea), jsou to původně aerobové, výjimečně anaerobové (Symbiontida), a většinou jsou to heterotrofové s výjimkou zmíněných fototrofních euglenofyt. Vzájemná poloha větví jednotlivých tříd na fylogenetickém stromě se liší v různých studiích, na obrázku 3 je znázorněn nejnovější z nich (Bicudo & Menezes, 2016; Breglia et al., 2013; Thomas Cavalier-Smith, 2016; Lax et al., 2019).

Původní Euglenozoa mohly být buď eukaryovorní organismy s velice komplexním potravním aparátem (Thomas Cavalier-Smith, 2016), nebo to byly naopak drobné bakteriovorní organismy, jež dnes nalezneme na bázi fylogenetických stromů tříd Kinetoplastida (Bodonida) i Euglenida (Petalomonadida a "ploetiiida") (Leander et al., 2007). Oba autoři dokládají svá tvrzení fylogenetickými stromy, které se ovšem neustále mění v závislosti na zastoupení zkoumaných druhů a znaků. Nicméně se zdá, že fylogenetická a morfologická data dobře odrážejí způsob výživy jednotlivých skupin třídy Euglenida (Leander et al., 2007).

**Bakteriovorní euglenidi** mají potravní aparát ve tvaru kapsy vyztužené mikrotubuly stejně jako bodonidi, ale zároveň se u nich vyskytuje nový znak, synapomorfní pro třídu Euglenida, kterým je pelikula s podélnými pásy, což je plasmatická membrána podložená nejen mikrotubuly, ale také navzájem propojenými proteinovými pásy navazujícími na cisterny endoplasmatického retikula (Thomas Cavalier-Smith, 2017; Leander, 2004). U bakteriovorních euglenidů je pásů jen 10 – 12, jsou uspořádány podélně a pelikula je tak poměrně tuhá. Rigiditu buněk také podporuje komplexní systém podpírající potravní aparát; jednak tyč složená z velkého množství mikrotubulů, která může procházet i celou buňkou (např. rod *Ploetia*) a jednak další podpůrné prvky, jako jsou lopatky a amorfni hmota (Lax et al., 2019; Leander et al., 2001). U části



euglenidů se vyvinula schopnost pozřít i organismy větší než bakterie a vznikla skupina **eukaryovorních euglenidů** (např. "peranemida" a "anisonemida"). Potravní aparát je podložen kratšími mikrotubuly a pelikula je složená z několika desítek pásů uspořádaných helikálně. Buňky jsou proto mnohem plastičtější. Jedna monofyletická větev eukaryovorních euglenidů ztratila schopnost predace a stali se z nich **primární osmotrofové** (Aphagea). S touto změnou souvisí druhotná ztráta potravního aparátu, snížení počtu pelikulárních pásů a s tím spojené snížení plasticity buněk (Leander et al., 2007, 2017).



Obr. 3. Schematický fylogenetický strom kmene Euglenozoa (upraveno podle Lax et al., 2019).

Z eukaryovorních euglenidů je také odvozena monofyletická fototrofní linie Euglenophyceae. Původní kořist, zelenou řasu, si **fototrofové** ponechali a procesem sekundární endosymbiózy došlo k její přeměně na sekundární plastid, který je obdařen třemi membránami, s tylakoidy ve sloupcích po třech a chlorofylem *a + b* (Leander et al., 2017). Druhotně u některých rodů z této skupiny došlo k částečné redukci plastidu a fotosyntetického aparátu a objevili se tzv. **sekundární osmotrofové** (např. *Euglena longa*, pův. *Astasia longa*) (Leander et al., 2017). Jak fototrofové, tak sekundární osmotrofové, mají redukovaný potravní aparát. Došlo také ke změně v pohybu buňky; změny v rytmu a pohybu bičíku umožnily opuštění klouzavého pohybu po povrchu, který byl nahrazen volným plaváním umožňujícím pohyb ve vodním sloupci za světlem (Leander et al., 2007). Pelikula je u těchto euglenidů tvořena několika desítkami pásů, které velmi zvýšily plasticitu buněk. Pásky jsou helikálně stočené a umožňují

tzv. metabolii – pohyb typický pro euglenofyty, kdy se velice rychle může měnit tvar buňky přeléváním buněčného obsahu, přirovnávaným k peristaltickému pohybu (Leander et al., 2007; Triemer et al., 2006). Část druhů je druhotně rigidních a pohyb si zajišťují především pohybem bičíku, některé mají dokonce mineralizované schránky (Leander et al., 2017; Triemer & Farmer, 1991). Také co se týče tvaru, počtu a umístění plastidů a vůbec celkové morfologie buňky je tato skupina nesmírně rozmanitá. Pro celou skupinu je dále typická zvětšená flagelární kapsa, nazývaná rezervoár, u jejíž základny se volně v cytoplasmě nachází také světločivná skvrna "stigma" a fotosenzory, které obojí zajišťují orientaci za světlem (Leander et al., 2017).

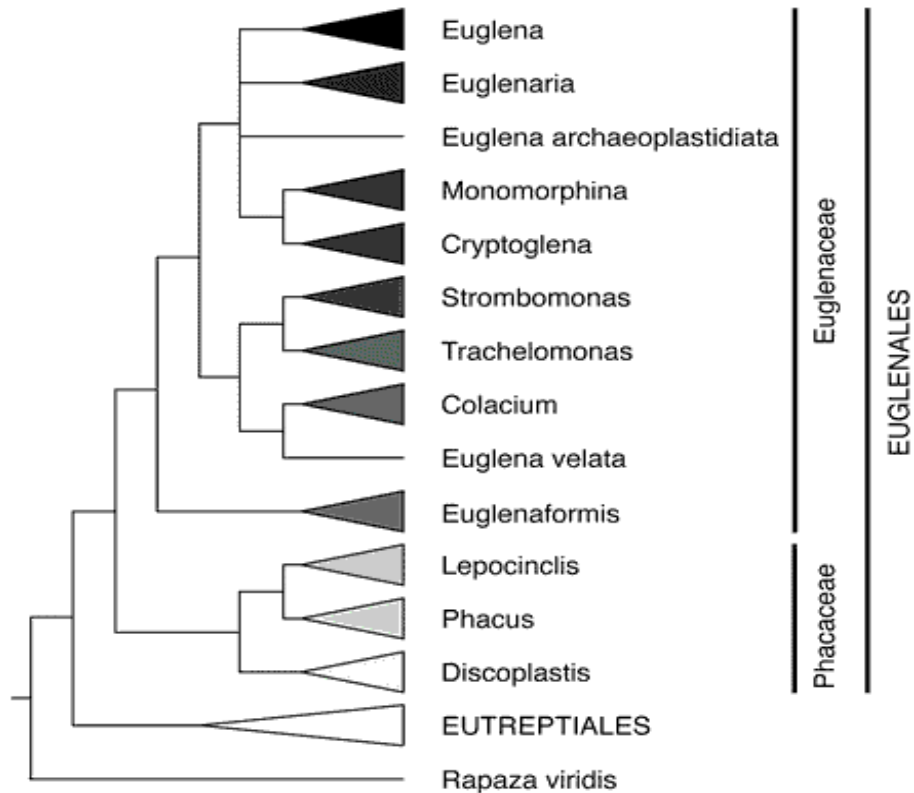
Představený scénář vzniku sekundárního plastidu pohlcením zelené řasy předkem euglenofyt je nazýván hypotéza "pozdního získání plastidu" a je dnes široce akceptován na základě nynějších fylogenetických studií (Hrdá et al., 2012; Jackson et al., 2018; Leander, 2004). Je třeba zde zmínit i alternativní, dnes podstatně méně přijímanou hypotézu "časného získání plastidu", která předpokládala, že euglenidi získali plastid již mnohem dříve, možná dokonce ještě před rozdělením sesterských skupin Euglenida a Kinetoplastea, a že se tento plastid udržel pouze v linii euglenofyt (Hannaert et al., 2003; Martin & Borst, 2003). Tato hypotéza se opírala o to, že genom trypanosom obsahuje minimálně 16 enzymů příbuzných zeleným rostlinám, které se nachází v cytoplasmě nebo v glykozómu a účastní se základního metabolismu cukrů.

## 4.2 Charakteristika Euglenophyceae

Euglenophyta (Euglenophyceae) jsou monofyletická skupina, kterou lze rozdělit na dvě podskupiny, z nichž jedna zahrnuje pouze jeden druh *Rapaza viridis* popsáný v roce 2012 (Yamaguchi et al., 2012) se sekundárním plastidem, nebo spíše kleptoplastidem nejasného původu, a druhá všechny ostatní fototrofy a jejich heterotrofní potomky, jejichž sekundární plastid má společný původ v zelené řase podobné dnešní řase *Pyramimonas parkeae* (Turmel et al., 2009).

Druh *R. viridis* je dnes zařazen do samostatného řádu na bázi euglenofyt. Je to volně žijící mořský organismus pohybující se pomocí dvou nestejných bičíků, u kterého byla pozorována i metabolie díky přítomnosti 16 helikálních pelikulárních pásů. Potravní aparát se skládá z potravní kapsy vypodložené čtyřmi řadami mikrotubulárních pásů bez dalších vyztužujících prvků. Jedná se o mixotrofní organismus, protože přestože obsahuje zelený plastid, nedokáže déle než 35 dní žít bez pravidelné predace určitého kmene zelené řasy *Tetraselmis* sp. (PRA-361). Tuto kořist pohlcuje celou, přičemž právě zmíněná metabolie napomáhá jednak při požití buňky a jednak při odstranění její buněčné stěny (theky). Řasa je posléze strávena celá, včetně svého plastidu. Po nějakou dobu však tyto plastidy přežívají (kleptoplastidy) a jsou obklopené dvěma membránami. Morfologicky jsou dobře odlišitelné od vlastních chloroplastů *R. viridis*, které mají tři membrány a jsou ultrastrukturně podobné chloroplastům euglenofyt. Při absenci světla *R. viridis* přežívá jen jeden týden, i když je přítomná řasa *Tetraselmis* jako kořist. Je tedy zřejmé, že oba typy výživy jsou nezbytné pro přežití tohoto euglenida, který tak tvoří přechodný článek mezi fagotrofními a fototrofními druhy (Yamaguchi et al., 2012).

"Tradiční" euglenofyta tvoří dva řády – Eutreptiales a Euglenales (obr. 4), který je podle většiny fylogenetických studií považován za sesterský k Eutreptiales (Lax et al., 2019), podle jiných se jedná o řád vnořený do Eutreptiales (Cavalier-Smith, 2016; Karnkowska et al., 2018).



Obr. 4. Schematický fylogenetický strom fototrofních euglenidů podle Karnkowska et al., 2015 a Kim et al., 2015 (převzato ze Zakryš et al., 2017).

První řád, Eutreptiales, zahrnuje dva rody *Eutreptia* a *Eutreptiella* v celkem 17 druzích podle AlgaeBase, které jsou všechny mořské nebo brakické a mají dva stejné či nestejně bičíky, což jsou znaky považované za původní (Adl et al., 2019). V ultrastruktuře psychrofilní výjimečně čtyřbičíkaté eugleny *Eutreptiella pomquetensis* (orig. *Tetreutreptia pomquetensis* (McLachlan et al., 1994) lze nalézt několikvrstevné svazky mikrotubulů, které by mohly představovat zbytky potravního aparátu fagotrofních euglenidů (Triemer & Lewandowski, 1994), byť u ní fagocytóza pozorována nebyla. Zato u druhu *Eutreptiella eupharyngea*, který je častým původcem vodního květu, bylo pozorováno, že se tyto organismy kromě fotosyntézy živily také eubakteriemi a sinicemi rodu *Synechococcus*, avšak ne předkládanou eukaryotickou kořistí, kterou tvořily zelené řasy např. *Tetraselmis* nebo *Isochrysis* (Walne et al., 1986; Yoo et al., 2018). Bakterie obklopené membránou v cytoplasmě byly pozorovány i u *Eutreptiella* sp., ale v tomto případě jsou autory považovány za endobionty, kteří dlouhodobě přežívají uvnitř buňky

a synchronně se s nimi množí (Kuo et al., 2013). Nicméně lze říci, že minimálně u některých zástupců Eutreptiales dochází také k mixotrofii, podobně jako u *R. viridis*.

Druhý řád, Euglenales, zahrnuje 11 rodů převážně sladkovodních bičíkoveců s jedním bičíkem vystupujícím a druhým skrytým v periflagelární kapse. Mohou být mixotrofní, v prostředí bohatém na organické látky nebo za anaerobních podmínek využívají alternativní zdroje energie a přechází na osmotrofní výživu (Mahapatra et al., 2013). Podle závěrů fylogenetických analýz na základě sekvence jaderných i plastidových ribosomálních podjednotek (ncLSU, ncSSU, ptLSU a ptSSU) je členěn na dvě samostatné monofyletické čeledi Euglenaceae a Phacaceae (Kim et al., 2010). Taxonomie euglenofyt prošla v posledních dvaceti letech několika změnami a upřesněními, které odrážejí nejen morfologické, ale také molekulárně biologické vlastnosti jednotlivých druhů (obr. 4). Mnoho druhů původně patřících do rodu *Euglena* doslalo nové rodové názvy, avšak dosud zbývají některé druhy a izolované kmeny, kterým by slušelo vlastní rodové jméno. Odhaduje se (Zakrýs et al., 2017), že z 560 druhů a poddruhů rodu *Euglena* popsáných v AlgaeBase bylo jen u jedné třetiny (asi 150 druhů) ověřeno a upřesněno jejich taxonomické postavení.

Čeď Euglenaceae zahrnuje osm rodů: *Colacium*, *Cryptoglana*, *Euglena*, *Euglenaformis*, *Euglenaria*, *Monomorphina*, *Strombomonas* a *Trachelomonas*. Kromě rodu *Euglena* jsou tyto rody s vysokou podporou monofyletické. Sesterské rody *Cryptoglana* a *Monomorphina* mají oba jeden velký parietální chloroplast a dále rigidní buňku, pročež postrádají metabolii. Rody *Trachelomonas* a z něj vyčleněný rod *Strombomonas* (Triemer et al., 2006) tvoří také monofyletickou jednotku a jsou navzájem sesterské, za jejich společný znak lze považovat mineralizovanou schránku zvanou lorika. Rod *Colacium* tvoří sesterskou větev k předchozím dvěma a vyznačuje se vejčitou buňkou přisedlou slizovitou stopkou k povrchu. Tvorba slizu je typická pro všechny tři naposledy zmíněné rody (Zakrýs et al., 2017).

Několik kryptických druhů rodu *Euglena* bylo na základě fylogenetických analýz přeřazeno do nového rodu *Euglenaria* (Linton et al., 2010), který je podle typu analýzy sesterský buď k rodu *Euglena* (Kim et al., 2015) nebo dvojici *Cryptoglana/Monomorphina* (Karnkowska et al., 2015). Podobně další rod *Euglenaformis* zastoupený jedním druhem *Euglenaformis proxima*, původně popsáný jako *Euglena proxima*, byl také vyřazen z rodu *Euglena* jako kryptický druh a tvoří linii sesterskou všem ostatním liniím čeledi Euglenaceae (Bennett et al., 2014).

Přes tato upřesnění v taxonomii je stále rod *Euglena* para/polyfyletický, rozdělený (zatím) do třech subkládů. První skupinu tvoří dobře podpořená monofyletická skupina zahrnující minimálně 23 druhů včetně modelového druhu *E. gracilis* či nefotosyntetizujícího druhu *E. longa*. Druhá skupina zahrnuje jeden druh *E. archaeoplastidiata*, která má některé morfologické znaky podobné sesterským rodům *Cryptoglana/Monomorphina* a některé dvojici *Euglena/Euglenaria*. Ve fylogenetických stromech se odvětjuje před rodem *Euglenaria* (Kim et al., 2015) nebo kládem *Euglenaria/Monomorphina/Cryptoglana* (Karnkowska et al., 2015). Třetí skupinu tvoří *E. velata* komplex, která je sesterská rodu *Colacium* (Kim et al., 2015).

Čeď Phacaceae zavedená v roce 2010 (Kim et al., 2010) zahrnuje tři rody: *Phacus*, *Lepocinclis* a *Discoplastis*. Všechny tři se vyznačují množstvím malých diskovitých chloroplastů. Rody *Phacus* a *Lepocinclis* jsou sesterské, mají rigidní nebo semirigidní buňky, které nemetabolují, ale pohybují se krouživým pohybem (Linton et al., 2010). Fylogenetické analýzy i zde vedly k velkým taxonomickým přesunům a přejmenováním mnoha druhů. Postupně byla část druhů rodu *Phacus* přesunuta do rodu *Monomorpha* (Marin et al., 2003) a dále druhy *Euglena limnophila* a *E. salina* byly přesunuty do rodu *Phacus* (Linton et al., 2010). K rodu *Lepocinclis* bylo postupně přezazeno 10 druhů z rodu *Euglena* (Kosmala et al., 2005; Marin et al., 2003) a následně i *P. horridus* jako *L. spinosa* (M. S. Bennett et al., 2012). Molekulární data ovšem nekorrespondují úplně s morfologickými, a tak hlavní rozlišovací znak, plochá buňka u rodu *Phacus*, není platný. Také v některých studiích vychází rod *Lepocinclis* jako vnitřní klád rodu *Phacus* (Karnkowska et al., 2015), ale zvýšení počtu taxonů ukazuje, že oba taxony jsou oprávněně monofyletické a nemá smysl je slučovat (Kim et al., 2015). Bazálně od těchto dvou rodů se odvíjí třetí rod *Discoplastis*, který se na rozdíl od nich pohybuje metabolii. Druhy *D. spathirhyncha* a *D. adunca* byly také přesunuty z rodu *Euglena* (Triemer et al., 2006) a jejich poloha na fylogenetickém stromě byla také několikrát aktualizována (Karnkowska et al., 2015; Kim et al., 2010; Kim et al., 2015; Linton et al., 2010).

### 4.3 Sekundárního plastid euglenofyt

Na přelomu sedmdesátých a osmdesátých let byl koncept vzniku sekundárního plastidu aplikován také na euglenofyta. Na základě buněčné ultrastruktury plastidu se třemi membránami byl u *E. gracilis* navržen vznik jejího plastidu z pohlcené zelené řasy, a nadbytečná membrána by pak představovala pozůstatek plasmalemy endosymbioticky žijící řasy (Gibbs, 1978). Pátrání po původu plastidu bylo nasměrováno pracemi hledajícími původ sekundárního plastidu chlorarachniofyt (Ishida et al., 1997; Rogers et al., 2007) a osekvenováním plastidu *Nephroselmis olivaceae* (Turmel et al., 1999) do skupiny zelených prasynofytních řas. K upřesnění došlo po osekvenování dalších tří plastidových genomů z této skupiny řas a dnes je za nejbližšího příbuzného původního endosymbionta považována mořská řasa *Pyramimonas parkeae* (Hrdá et al., 2012; Turmel et al., 2009), byť nejnovější fylogenetická srovnání nevyklučují ani možnost, že nejbližším příbuzným by mohl být jeden z rodů *Cymbomonas* nebo *Pterosperma* (Jackson et al., 2018). Vzhledem k tomu, že linie Eutreptiales zahrnuje také mořské druhy, předpokládá se, že k endosymbiotické události došlo v mořském prostředí (Marin, 2004), přestože většina euglenofyt je sladkovodních. Nově sice byly nalezeny stopy nových mořských druhů skupiny Euglenales v mořských enviromentálních vzorcích DNA, ale obecně nejsou euglenofyta v těchto vzorcích vůbec hojná (Lukešová et al., 2019).

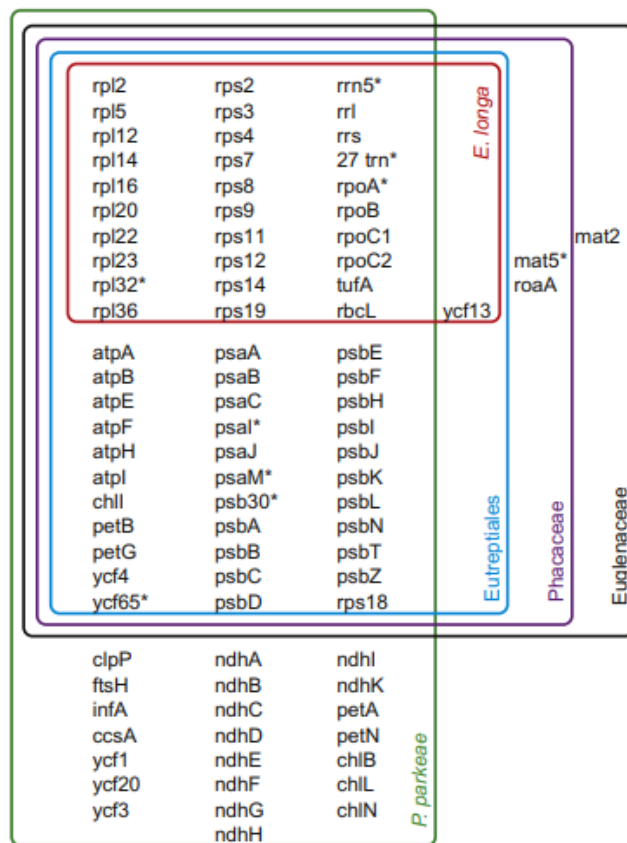
Alternativní hypotéza hovoří o vzniku skupiny ve sladkovodním prostředí. Protože do fylogenetických analýz byly prozatím zahrnuty pouze mořské druhy prasynofytních řas, je možné, že některý sladkovodní druh, např. *P. tetrahynechus*, by mohl být blíže příbuzný stávajícímu plastidu. Pro to by svědčil i fakt, že nefotosyntetický, euglenofytům nejbližší příbuzný rod *Heteronema* zahrnuje také množství sladkovodních druhů (Lax et al., 2019).

## 4.4 Obsah genů v plastidovém genomu euglenofyt

Genom chloroplastu *P. parkeae* obsahuje 110 genů (Satjarak & Graham, 2017b; Turmel et al., 2009). Během sekundární endosymbiózy byl tento počet redukován na 89 genů, které najdeme s jednotlivými výjimkami u plastidových genomů euglen (včetně genu pro 5S rRNA, který nebyl v plastomu *P. parkeae* identifikován). Jinými slovy, k redukcii plastidových genů došlo jedinkrát u společného předka řádů *Eutreptiales* a *Euglenales*. Soubor 22 plastidových genů, které u euglen nejsou přítomny, zahrnuje jednak geny, které byly pravděpodobně zcela ztraceny (např. geny pro subjednotky NADH-plastoquinon oxidoreduktázy plastidového respiračního řetězce *ndhA-K*), a jednak geny, které byly přeneseny do jádra společného předka a byly identifikovány v transkriptomu *E. gracilis* a *Etl. gymnastica* (např. cytochrom *c* biogenesis protein *ccsA*, metaloproteáza *ftsH* nebo fotosystém I assembly protein *YCF3*) (Hrdá et al., 2012).

Zmíněných 89 zachovaných plastidových genů euglen zahrnuje 32 genů pro fotosyntetické proteiny, 5 genů pro transkripční a translační proteiny, 22 genů pro ribozomální proteiny, 3 typy rRNA a 27 různých tRNA (obr. 5). V některých genomech chybí jeden až tři z těchto genů, ale k těmto ztrátám došlo u jednotlivých druhů víceméně náhodně. Jediný osekvenovaný nefotosyntetizující plastid *E. longa* postrádá dalších 33 genů, které kódují proteiny zajišťující fotosyntézu, zachovány jsou však geny pro rRNA a tRNA, transkripční a translační faktory, a geny pro ribozomální proteiny. Výjimku tvoří gen pro velkou podjednotku komplexního proteinu Rubisco, který u *E. longa* zůstal (Gockel et al., 2000). Zda je jeho přítomnost důvodem přetrvávání plastidu v buňce, není zřejmé, zvláště proto, že jeho exprese je velmi nízká (Záhonová et al., 2016). Jednotlivé genomy euglen dále obsahují více či méně otevřených čtecích rámců (ORF), jejichž funkce je většinou neznámá nebo pouze hypotetická, a jeden či více genů s funkcí maturázy/reverzní transkriptázy.

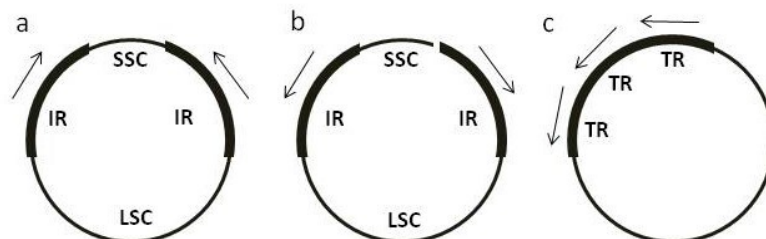
První osekvenovaný plastidový genom euglenofyt byl genom *Euglena gracilis* (Hallick et al., 1993), následoval genom nefotosyntetizujícího druhu *E. longa* (Gockel et al., 2000). S nástupem sekvenování nové generace bylo od roku 2012 publikováno dalších 28 plastidových genomů. Mezi prvními byly publikovány dva plastomy z řádu *Eutreptiales*, a to plastom *Eutreptiella gymnastica*, který byl charakterizován v naší laboratoři (Hrdá et al., 2012), a dále plastom *Eutreptia viridis* (Wiegert et al., 2012). Třetí plastom z tohoto řádu, *Eutreptiella pomquetensis*, přibyl v roce 2017 (Dabbagh et al., 2017). U řádu *Euglenales* je plastidových genomů osekvenováno mnohem více, 9 z čeledi *Phacaceae* – *Phacus orbicularis* (Kasiborski et al., 2016), *P. inflexus*, *P. pleuronectes*, *Lepocinclis tripteris* MI 101, *L. tripteris* UTEX 1311, *L. playfairiana*, *L. steinii*, *L. ovum* a *Discoplastis spathirhyncha* (Karnkowska et al., 2018), a ke dvěma původně charakterizovaným euglenám přibyl 16 dalších druhů, včetně poddruhů, z čeledi *Euglenaceae* (Bennett et al., 2012, 2014; Bennett & Triemer, 2015; Dabbagh & Preisfeld, 2017, 2018; Hewadikaramge & Linton, 2018; Pombert et al., 2012; Wiegert et al., 2013) (viz tabulka č. 1).



Obr. 5. Obsah genů v plastidovém genomu *P. parkeae* a euglenofyt. Hvězdička značí geny, které u některých druhů chybí (převzato z Vanclová et al., 2017).

## 4.5 Struktura genomu

Chloroplastový genom *P. parkeae* má typické kvadripartitní uspořádání, kde IR nesou kromě ribozomálního operonu ještě dalších 7 genů a jsou orientovány proti sobě na plus a minus vlákně směrem k SSC úseku, jak je obvyklé u plastidových genomů. V SSC úseku je většina genů pro podjednotky *ndh* a několik dalších genů. Většina ostatních genů je na LSC úseku. Takovéto uspořádání bylo pravděpodobně i u předka plastomu Euglen (Turmel et al., 2009) (obr. 6).



Obr. 6. Uspořádání plastidových genomů *P. parkeae* (a), *Etl. pomquetensis* (b) a *E. gracilis* (c), IR – invertovaná repetice, TR – tandemová repetice, LSC – Large single copy region, SSC – Small single copy region. Šipky značí směr ribozomálního operonu 16S, *trnI*, *trnA*, 23S, 5S.

Tab. 1. Základní charakteristiky chloroplastových genomů euglenofyt a *P. parkeae*. Uspořádání rRNA operonu: TR – tandemová repetice, IR – invertovaná repetice, ? – není známa orientace, č – částečný rRNA operon. Hvězdička \* označuje genomy s neuzavřenou kružnicí. Počet plastidových genů zahrnuje jen geny přenesené od společného předka plastidu včetně genů pro rRNA a tRNA, kopie nejsou započítány. V tabulce není uveden počet dalších otevřených čtecích rámců (ORF) s neznámou funkcí. Černě Euglenaceae (včetně nefotosyntetické *E. longa*, která je označena červeně), fialově Phacaceae, modře Eutreptiales a zeleně *P. parkeae*. Pořadí druhů odpovídá větvení schematického fylogenetického stromu z obr. 4. *mat1* je synonymem *ycf13* z obr. 5.

Druh	Velikost genomu	Počet plastidových genů	Počet intronů	Počet kopií rRNA operonu/ uspořádání	Maturázy/ rev. transkriptázy (RT)
<i>Euglena longa</i>	73,3 kbp	56	60	3+č / TR	mat1, RoaA
<i>Euglena gracilis</i> var. <i>bacillaris</i>	132 kbp	87	134	3+č / TR	mat1, mat2, RoaA
<i>Euglena gracilis</i> Z	143,2 kbp	87	145	3+č / TR	mat1, mat2, RoaA
<i>Euglena viridis</i> epitype	76,2 kbp	89	77	1	mat1, mat2, RoaA
<i>Euglena viridis</i> SAG 1224-17d	91,6 kbp	89	77	1	mat1, mat2, RoaA
<i>Euglena hiemalis</i>	128,3 kbp	88	100	2 / TR	mat1, mat2, RoaA
<i>Euglena clara</i>	98,4 kbp	89	138	2 / TR	mat1, mat2, RoaA
<i>Euglena mutabilis</i>	87 kbp	89	76	1	mat1, mat2, RoaA
<i>Euglenaria anabaena</i>	88,5 kbp	89	82	1	mat1, mat2, mat5, RoaA
<i>Euglena archaeoplastidiata</i>	97,2 kbp	89	92	1	mat1, mat2, mat5, RoaA
<i>Monomorphina aenigmatica</i>	74,7 kbp	89	53	1	mat1, mat2, RoaA
<i>Monomorphina parapyrum</i>	80,1 kbp	89	80	1	mat1, mat2, mat5, RoaA
<i>Cryptoglena skujai</i>	106,8 kbp	89	84	1	mat1, mat2, RoaA
<i>Strombomonas acuminata</i> *	144,2 kbp	89	110	1+č / IR *	mat1, mat2, mat5, RoaA
<i>Trachelomonas volvocina</i>	85,4 kbp	89	94	1	mat1, mat2, mat5, RoaA
<i>Trachelomonas grandis</i>	113,3 kbp	89	68	1	mat1, mat2, mat5, RoaA
<i>Colacium vesiculosum</i> *	128,9 kbp	88	130	1+č / TR *	mat1, mat2, mat5, RoaA
<i>Euglenaformis proxima</i>	94,2 kbp	87	113	1	mat1, mat2, mat5, RoaA
<i>Lepocinclis ovum</i> *	69,6 kbp	88	50	2 / IR *	mat1, mat2, RoaA
<i>Lepocinclis playfairiana</i> *	77,4 kbp	88	79	2 / IR *	mat1, mat2, RoaA
<i>Lepocinclis steinii</i> *	81,8 kbp	88	81	2 / IR *	mat1, mat2, RoaA
<i>Lepocinclis tripteris</i> MI 101 *	81 kbp	89	94	2 / IR *	mat1, mat5, RoaA
<i>Lepocinclis tripteris</i> UTEX 1311 *	83 kbp	89	95	2 / IR *	mat1, mat5, RoaA
<i>Phacus inflexus</i>	58,1 kbp	89	29	1	mat1
<i>Phacus orbicularis</i> *	66,4 kbp	88	67	1 *	mat1, Roa
<i>Phacus pleuronectes</i>	90,9 kbp	89	104	1	mat1, mat2, mat5, RoaA
<i>Discoplastis spathirhyncha</i>	84 kbp	89	59	2 / IR	mat1, mat2
<i>Eutreptia viridis</i> *	65,5 kbp	86	27	2 / ? *	mat1
<i>Eutreptiella pomquetensis</i> *	130,6 kbp	88	51	2 / IR *	mat1
<i>Eutreptiella gymnastica</i>	67,6 kbp	86	8	2 / IR	mat1, (mat5?)
<i>Pyramimonas parkeae</i> CCMP726	101,6 kbp	110	1	2 / IR	Put. maturaza / RT
<i>Pyramimonas parkeae</i> NIES254	104,8 kbp	110	1	2 / IR	Put. maturaza / RT



V průběhu procesu usídlování plastidu v hostitelské buňce však nedošlo jen ke ztrátě asi čtvrtiny genů, ale zároveň také k velkým genomickým přestavbám. U původnějších druhů (Eutreptiales) se jeví zachovalá kvadripartitní struktura, ale při bližším pohledu je jasné, že neodpovídá původní struktuře získaného plastomu. Většinu z genů nacházejících se na SSC úseku *P. parkeae* u euglen vůbec nenajdeme, kromě genu *rpl32* a tří genů pro tRNA, které se přesunuly na LSC. U plastomu *Etl. pomquetensis*, který je anotován jako lineární molekula, neboť se nepodařilo kružnici uzavřít, bylo kvadripartitní uspořádání nalezeno, ale jednotlivé části neodpovídají strukturou genů *P. parkeae*. Ribozomální operony na IR úsecích jsou orientovány směrem od SSC úseku, tedy opačně než u *P. parkeae* (Dabbagh et al., 2017). Ještě patrnější je přestavba u rodů *Discoplastis* a *Lepocinclis*, kde SSC úsek prakticky úplně chybí, neobsahuje žádný gen, ale pouze tandemové repetice (Karnkowska et al., 2018), a ribozomální operony jsou orientovány vždy od sebe. U *Etl. gymnastica* IR úseky směřují také od SSC úseku, ve kterém se nachází pouze jeden ORF neznámé funkce, a navíc je jeden IR úsek přerušen a do ribozomálního operonu je zde vložena část LSC úseku (Hrdá et al., 2012).

U dalších rodů nebyly IR úseky anotovány, a tak se může zdát, že celkem třikrát v evoluci euglenophyt byla invertovaná repetice ztracena a to u rodů *Eutreptia* a *Phacus* a dále u předka celého řádu Euglenales (Karnkowska et al., 2018). Ztráta jedné kopie repetice je sice poměrně vzácná, ale je zdokumentovaná umnoha zelených řas z většiny chlorofytických řádů (Lang & Nedelcu, 2012). Ztráta u rodu *Phacus* a řádu Euglenales je velice pravděpodobná, ovšem u rodu *Eutreptia*, kde je osekvenovaný pouze plastom jednoho druhu, o ztrátě IR nelze definitivně rozhodnout, neboť ani zde se nepodařilo anotaci cirkularizovat. Navíc, geny pro rRNA operon mají dvojnásobné pokrytí čtení oproti ostatním genům a je tak vysoce pravděpodobné, že druhá kopie ribozomálního operonu v genomu je přítomná (Wiegert et al., 2012). Vzhledem k neúplné anotaci ale není jasné, jestli je v invertovaném nebo v tandemovém uspořádání.

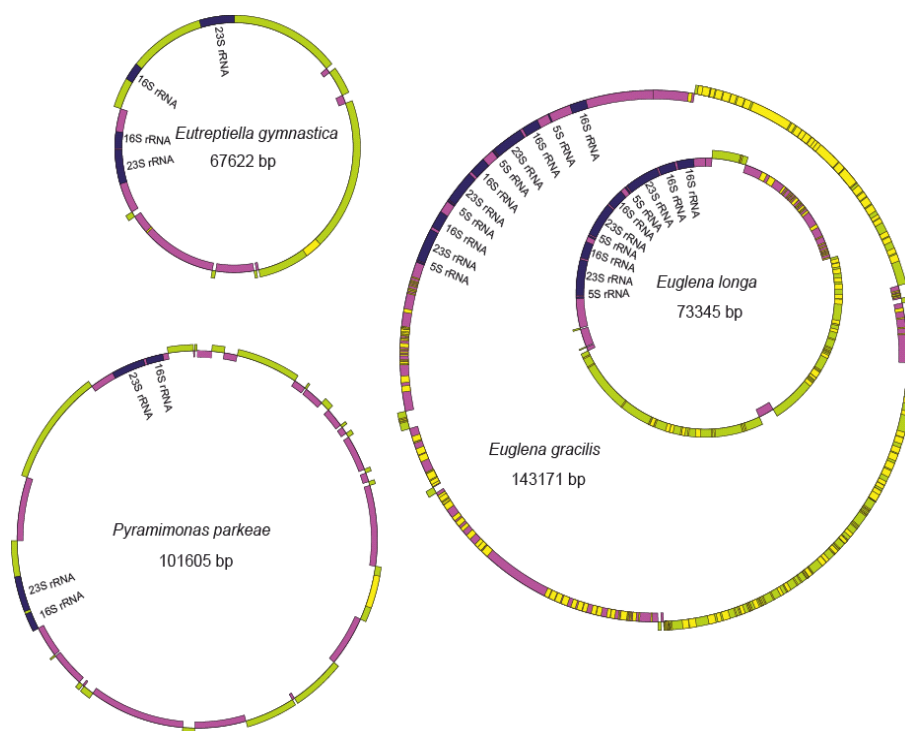
Zřejmě druhotně je pak ribozomální operon zmnožen a tvoří tandemovou repetici ve třech kopiích u druhů *E. gracilis* a *E. longa* (Gockel et al., 2000; Hallick et al., 1993), a ve dvou kopiích u *E. clara* a *E. hiemalis* (Hewadikaramge & Linton, 2018). Také u dvou dalších zástupců Euglenaceae (*Strombomonas* a *Colacium*) byly predikovány podle pokrytí ve čtení dva ribozomální operony, ale anotace nebyla cirkularizována a doplněna, takže konečné rozmístění operonů není známé. Navíc to podle překryvů vypadá tak, že *Colacium* by mohlo obsahovat tandemovou a *Strombomonas* spíše invertovanou repetici (Wiegert et al., 2013). Potíže při cirkularizaci genomů jsou časté, neboť v oblasti sousedící s ribozomálním operonem jsou u mnoha plastomů obvykle krátké tandemové repetitivní sekvence, které ztěžují či přímo znemožňují spojení konců osekvenovaných kontigů.

Původní pořadí genů plastomu společného předka nelze odvodit a obecně lze říci, že jednotlivé čeledi se pořadím genů velice liší. Lze vysledovat jednotlivé klastry genů, které jsou více či méně přítomné u všech euglenofytů a částečně i u *P. parkeae*. Těchto klastrů bylo definováno 15. Jsou seřazeny stejně nebo velmi podobně u blízké příbuzných druhů, např. u rodu *Euglena* jsou mezidruhové rozdíly jen v pořadí tří klastrů, jinak je pořadí klastrů a genů stejné. Úplně stejné mají pořadí klastrů rody *Monomorphina*, *Cryptoglana*, *Euglenaria* a *Strombomonas*

(Dabbagh & Preisfeld, 2017). Také mezi rody čeledi *Phacaceae* je pořadí klastrů velmi podobné s vysokou vnitrodruhovou podobností (Karnkowska et al., 2018). U bazální linie *Eutreptiales* je naopak pořadí klastrů i genů velice různorodé a velice se liší mezi sebou navzájem i v porovnání se zástupci *Euglenales*.

#### 4.6 Introny v plastomech euglenofyt

Plastidové genomy euglenofyt se přes téměř totožný obsah genů velice liší svou velikostí. Je to dáno počtem intronů, který je různý u jednotlivých linií. Lze zobecnit, že druhy, které se nacházejí na bázi fylogenetického stromu euglenofyt, mají intronů méně, kdežto odvozenější druhy jich mají více. Nejméně intronů, osm postupně anotovaných, má *Etl. gymnastica* (Hrdá et al., 2012; Pombert et al., 2012), nejvíce zatím bylo nalezeno u *E. gracilis*, a to 134 – 145 intronů (Doetsch et al., 1998; Hallick et al., 1993). Chloroplastový genom *Pyramimonas parkeae* nese jen jeden intron v genu *atpB*, který není homologní pozicí ani sekvencí s žádným intronem euglenofyt. Původním stavem tedy mohl být chloroplastový genom s minimem intronů nebo zcela bez intronů, ve kterém se postupně usazovali další a další introny (Pombert et al., 2012; Turmel et al., 2009) (obr. 7).

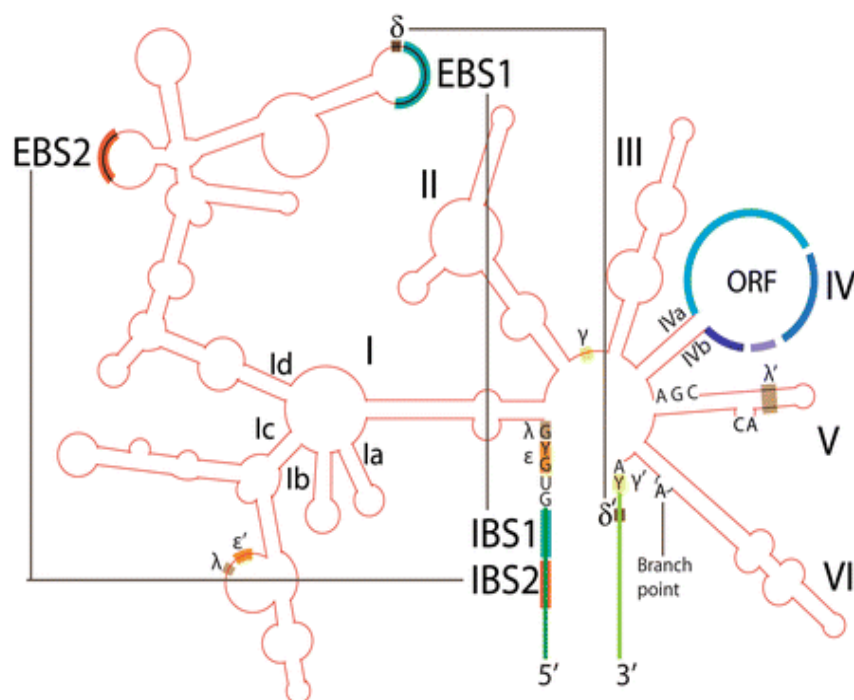


Obr. 7. Porovnání chloroplastového genomu *P. parkeae*, *Etl. gymnastica*, *E. gracilis* a *E. longa*. Velikosti kružnic jsou v poměru velikostí genomů. Zeleně geny na (+) vlákne, fialově na (-) vlákne, modře rRNA geny a žlutě introny (převzato z Hrdá et al., 2012).

U euglenofyt nalezneme několik typů chloroplastových intronů. Základním je intron typu II, který je schopen se sám vystřihnout a je typický pro plastidovou a mitochondriální DNA. Typicky obsahuje šest strukturních domén DI – DVI ve tvaru vlásenek se smyčkami, které jsou uspořádány do kruhu tak, že 5' a 3' konce intronu se dostávají k sobě. Introny zároveň tvoří

komplikovanou terciární strukturu, která je výsledkem párování několika různých konzervovaných sekvencí rozmístěných v definovaných místech intronu. Podle konkrétních sekvencí a jejich interakcí se introny typu II obecně dělí do minimálně tří subtypů (Bonen & Vogel, 2001). Konzervovaná je doména DV, sekvence GUGYG (kde Y je pyrimidinová база) na 5' konci, dinukleotid AY na 3' konci a doména DVI, což je vlásenka se smyčkou, která nese vyboulený nespárovaný nukleotid adenosin devět nukleotidů před 3' koncem intronu (obr. 8). Samotný sestřih probíhá autokatalyticky a byl prokázán *in vitro*. Při sestřihu intronů typu II dochází ke dvěma transesterifikačním reakcím. Při první, hydroxylová skupina nespárovaného adenosinu atakuje první nukleotid intronu za uvolnění 5' exonu, při druhé, nově uvolněná hydroxylová skupina 5' exonu atakuje první nukleotid 3' exonu, což vede ke spojení exonů a uvolnění intronu ve formě lariátové smyčky (Bonen & Vogel, 2001; Michel & Ferat, 1995; Zimmerly & Semper, 2015).

Některé introny typu II mají do domény DIV vložen otevřený čtecí rámec (ORF), který kóduje polypeptid, potřebný pro *in vivo* sestřih těchto intronů (Bonen & Vogel, 2001). Délka intronu typu II u *E. gracilis*, pokud nezahrneme délku vloženého ORF, je 277 – 671 bp (Hallick et al., 1993), v podobném rozsahu jsou i délky intronů typu II u ostatních Euglenofyt. To je v průměru o 100 bp méně, než u typického intronu typu II (Michel et al., 1989).



Obr. 8. RNA struktura plastidového intronu typu II, podle Zimmerly & Semper 2015. Strukturní domény intronu DI – DVI jsou červené, označené římskými čísly; exony jsou zelené. IBS, místo vázající intron (intron-binding site) a EBS, místo vázající exon (exon-binding site) spolu interagují a formují terciární strukturu umožňující sestřih. ORF, otevřený čtecí rámec modře, s doménami pro reverzní transkriptázu, maturázu, endonukleázu a doménu vázající DNA.

Dalším typem intronu je unikátní, jinde nepopsaný intron typu III (Christopher & Hallick, 1989; Copertino et al., 1994). Jedná se zřejmě o redukovanou formu intronu typu II, které chybí celé domény DII – DV. Má částečně zachovanou konsenzus sekvenci na 5' konci intronu (NTNNG) a částečně i doménu DI a dále vlásenku se smyčkou na 3' konci, která funkčně odpovídá doméně DVI s nespárovaným adenosinem většinou osm nukleotidů před 3' koncem intronu. Délka intronu typu III je pouze 91 – 120 nukleotidů a intron je AT bohatý (Copertino et al., 1994; Doetsch et al., 1998; Hong & Hallick, 1994). Kromě těchto redukovaných intronů typu III byla popsána ještě přechodná forma tzv. mini introny typu II (mini group intron II) (Dabbagh et al., 2017; Doetsch et al., 1998), u kterých chybí pouze některé z domén DII – DV a jejichž délka je kolem 200 pb. K samosestříhu dochází stejně, jako u intronu typu II (Copertino et al., 1994).

Intron typu III zatím nebyl zjištěn u zástupců Eutreptiales. Nejkratší introny u *Et. viridis* a *Etl. gymnastica* mají délku 156 resp. 179 párů bazí a lze u nich identifikovat nejen doménu DVI, ale i doménu DV. Lze je tedy považovat za mini introny typu II. U *Etl. pomquetensis* nebyl nalezen ani tento přechodný typ intronu (Dabbagh et al., 2017).

U euglen byly také poprvé popsány tzv. twintrony, introny uvnitř intronů (Doetsch et al., 1998; Hallick et al., 1993; Hong & Hallick, 1994). Twintrony mohou být sestaveny z jakékoli dvojice typů intronů – dvou intronů typu II, dvou intronů typu III, intronu typu III v intronu typu II a naopak. Dále byly nalezeny komplexní twintrony, vzniklé kombinací více než dvou intronů, kdy je několik intronů vložených do různých míst externího intronu, či interní intron obsahující další interní intron (Dabbagh et al., 2017; Hallick et al., 1993; Pombert et al., 2012; Thompson et al., 1995). V intronu i1 genu *petG* u *Etl. pomquetensis*, která má obecně velice konzervované sekvence intronů, byl dokonce nalezen interní twintron vložený do externího twintronu (Dabbagh et al., 2017). Předpokládá se, že pro správné vystřížení vnějšího intronu je nutné prvotní vystřížení vnitřních intronů (Copertino et al., 1991). Některé pokusy u *E. gracilis* však ukazují, že sestřih komplexních twintronů může probíhat alternativně a dávat vznik různým formám vystřížených intronů. Twintrony tak mohly vést ke vzniku intronů typu III, kdy po částečném sestřihu zůstaly mezi exony pouze koncové domény (DI a DVI) dvou původně různých intronů (Hong & Hallick, 1994). Také mini introny typu II jsou pravděpodobně výsledkem takového neúplného sestřihu (Dabbagh et al., 2017; Doetsch et al., 1998). Twintrony mohou být také velmi dobře použity jako fylogenetický marker, protože lze sledovat jejich přítomnost či nepřítomnost v různých liniích. *M. aenigmatica* má například introny typu II v místech, kde *E. gracilis* má popsané twintrony (Pombert et al., 2012).

Jak už bylo řečeno, součástí některých intronů je otevřený čtecí rámeček (ORF). U různých druhů se počet těchto genů liší, bývá jich méně (někdy mnohem méně), než je samotných intronů (např. *Etl. gymnastica* 4 ORF/8 intronů; *Etl. pomquetensis* 11 ORF/51 intronů; *E. gracilis* 4 ORF /145 intronů; *P. parkeae* 1 ORF/1 intron). U některých z těchto genů není funkce jasná, u velké většiny z nich lze vysledovat jejich homologii s geny pro reverzní transkriptázu s maturázou, které jsou nutné pro *in vivo* sestřih (maturáza) a pohyblivost v genomu (rev. transkriptáza). Introny typu II se totiž chovají jako mobilní elementy, které se šíří v rámci genomu nebo dokonce mezi genomy (Bonen & Vogel, 2001). To, že se ORF nenachází

v každém intronu, znamená, že produkt takového genu asistuje sestřihu i jiných intronů stejného typu (Dabbagh et al., 2017). V plastomu euglen bylo nalezeno několik takovýchto reverzních transkriptáz/maturáz, z nichž některé jsou již pojmenované a jejichž přítomnost je považována za důvod masivní přítomnosti intronů.

#### 4.7 Maturázy/reverzní transkriptázy u euglenofyt

U všech zástupců Euglenophyceae je v intronu genu *PsbC* přítomen gen pro maturázu *mat1* (syn. *yef13*). Má shodné umístění asi 140 – 154 bazí od začátku genu u všech druhů, přestože u Eutreptiales se nachází v prvním intronu a u Euglenales ve 2 – 4 intronu, podle toho, kolik intronů bylo před tento intron postupně vloženo (Dabbagh & Preisfeld, 2017). Tento gen *mat1* se nalézá také v plastidovém genomu *E. longa*, který ovšem ztratil příslušný fotosyntetický gen včetně jeho intronu a *mat1* zde zůstal jako samostatně stojící gen (Gockel et al., 2000). U Euglenales je gen *mat1* umístěn ve twintronu složeného ze dvou intronů typu III a byl proto považován za gen zodpovědný za sestřih intronů typu III (Copertino et al., 1994; Doetsch et al., 1998). U Eutreptiales je ovšem umístěn do intronu/twintronu typu II a zdá se tedy, že toto byl původní stav, ze kterého se twintron typu III vyvinul (Pombert et al., 2012).

Druhou maturázou je *mat2*. Byla identifikována pouze u Euglenales a to v intronu genu *PsbC*, který byl vložen před intron obsahující gen *mat1* do pozice 61 (Dabbagh & Preisfeld, 2017). Tento gen ovšem chybí u několika druhů Phacococcales (Karnkowska et al., 2018). Další maturáza *mat5* byla popsána také jen u Euglenales v intronu genu *PsbA*, její přítomnost je ještě více náhodná a gen byl ztracen několikrát u celých linií (Karnkowska et al., 2018). Zajímavé je, že po jeho identifikaci bylo možné najít slabou homologii *orf291* v druhém intronu *PsbC* *Etl. gymnastica* s tímto genem, což by znamenalo, že mohl být přítomen už u Eutreptiales a později ztracen, zanechav v *Etl. gymnastica* jen nefunkční pseudogen (Dabbagh & Preisfeld, 2017; Hrdá et al., 2012). Maturázy *mat3* v genu *PsbD* u *E. gracilis* (Hallick et al., 1993) a *mat4* v genu *PsbA* u *E. myxocylindracea* (Sheveleva & Hallick, 2004) byly popsány jen u těchto druhů. Druhý jmenovaný má prokazatelně sinicový původ a byl získán horizontální genovým přenosem (Sheveleva & Hallick, 2004).

Poslední reverzní transkriptázou/maturázou, která by měla být zmíněna, je gen *roaA* (ribosomal operon associated gene A). Tento gen se nalézá v mezigenové oblasti mezi geny *rps3* a *rpl16* v operonu ribosomálních proteinů, v rámci kterého je také přepisován a není umístěn v žádném intronu. Je opět přítomen pouze u Euglenales a na základě studia *E. gracilis* byl kandidátem na gen zodpovědný za sérii intronů typu III v tomto ribosomálním operonu (Jenkins et al., 1995). Analogicky ke genu *mat1* u *E. longa* i zde patrně došlo ke ztrátě původního intronu, ve kterém byl tento gen umístěn. To, že je tento intron v mezigenové oblasti také není výjimečné, např. v plastomech *E. mutabilis* a *E. gracilis* jsou další tři intercistronické introny (Dabbagh & Preisfeld, 2017).

Vzhledem k množství typů intronů a jejich struktuře je často velmi obtížné tyto správně identifikovat a anotovat. Také geny pro reverzní transkriptázy/maturázy, které už tak nemají

mezi sebou příliš velkou sekvenční podobnost, jsou často anotovány jen částečně s tím, že se někdy zřejmě jedná o pseudogeny (Dabbagh et al., 2017). S množstvím přibývajících plastidových genomů jsou doplňovány i anotace již publikovaných sekvencí o některé dosud přehlédnuté nebo nenalezené geny či jejich součásti, pomocí speciálně vytvořených skriptů jsou doplňovány i introny a twintrony (Dabbagh et al., 2017; Pombert et al., 2012).

#### **4.8 Plastidový genom *Eutreptiella gymnastica***

Prvním výstupem mé disertační práce je publikace plastidového genomu *Eutreptiella gymnastica* v roce 2012, ve které jsme popsali tehdy teprve třetí plastom Euglenophyt. Genom má délku 67,623 párů bazí a obsahuje 86 genů původně obsažených ve společném předkovi plastomu euglenofyt a dále pět otevřených čtecích rámců. Jeden je samostatně stojící a čtyři jsou součástí intronů, z nichž jeden odpovídá maturáze *mat1*. Gen pro *mat1* se stejně jako u ostatních euglen nachází v intronu v genu *PsbC*. Genom obsahuje pouze osm intronů, v práci byly původně popsány dva z nich, což je nejmenší nalezený počet u euglenophyt. Genom je velice kompaktní, mezigenové oblasti tvoří méně než 14 %. Přibližně dvě třetiny genů jsou přepisovány z jednoho řetězce dvoušroubovice DNA a jedna třetina z druhého. Plastom má kvadripartitní strukturu. V genomu však došlo k přestavbám, a tak část LSC oblasti byla přesunuta do jedné z invertovaných repetit, kterou takto rozdělila. V SSC oblasti se nachází pouze jeden otevřený čtecí rámec neznámé funkce. V pořadí genů lze vysledovat některé podobnosti jak s *E. gracilis* tak s *P. parkeae* a lze nalézt některé genové klastry společné pro všechny tři druhy. Tato práce popsala první plastidový genom z řádu Eutreptiales a předvedla, že ke ztrátě části genů z plastomu společného předka celé třídy Euglenophyta došlo ještě před jejich diverzifikací. Také potvrdila dřívější zjištění, že společným donorem plastidu euglenofyt byl blízký příbuzný prasinofytní řasy *Pyramimonas parkeae*.

## 4.9 Publikace I

Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012).

**The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids.**

PLoS ONE, 7(3), e33746.

<https://doi.org/10.1371/journal.pone.0033746>





# The Plastid Genome of *Eutreptiella* Provides a Window into the Process of Secondary Endosymbiosis of Plastid in Euglenids

Štěpánka Hrdá<sup>1</sup>, Jan Fousek<sup>2</sup>, Jana Szabová<sup>1</sup>, Vladimír Hampel V<sup>1\*</sup>, Čestmír Vlček<sup>2\*</sup>

**1** Charles University in Prague, Faculty of Science, Department of Parasitology, Prague, Czech Republic, **2** Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

## Abstract

Euglenids are a group of protists that comprises species with diverse feeding modes. One distinct and diversified clade of euglenids is photoautotrophic, and its members bear green secondary plastids. In this paper we present the plastid genome of the euglenid *Eutreptiella*, which we assembled from 454 sequencing of *Eutreptiella* gDNA. Comparison of this genome and the only other available plastid genomes of photosynthetic euglenid, *Euglena gracilis*, revealed that they contain a virtually identical set of 57 protein coding genes, 24 genes fewer than the genome of *Pyramimonas parkeae*, the closest extant algal relative of the euglenid plastid. Searching within the transcriptomes of *Euglena* and *Eutreptiella* showed that 6 of the missing genes were transferred to the nucleus of the euglenid host while 18 have been probably lost completely. *Euglena* and *Eutreptiella* represent the deepest bifurcation in the photosynthetic clade, and therefore all these gene transfers and losses must have happened before the last common ancestor of all known photosynthetic euglenids. After the split of *Euglena* and *Eutreptiella* only one additional gene loss took place. The conservation of gene content in the two lineages of euglenids is in contrast to the variability of gene order and intron counts, which diversified dramatically. Our results show that the early secondary plastid of euglenids was much more susceptible to gene losses and endosymbiotic gene transfers than the established plastid, which is surprisingly resistant to changes in gene content.

**Citation:** Hrdá Š, Fousek J, Szabová J, Hampel V V, Vlček Č (2012) The Plastid Genome of *Eutreptiella* Provides a Window into the Process of Secondary Endosymbiosis of Plastid in Euglenids. PLoS ONE 7(3): e33746. doi:10.1371/journal.pone.0033746

**Editor:** Jonathan H. Badger, J. Craig Venter Institute, United States of America

**Received:** November 23, 2011; **Accepted:** February 16, 2012; **Published:** March 20, 2012

**Copyright:** © 2012 Hrdá et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by Ministry of Education, Youth and Sport of the Czech Republic (project MSM0021620828), by Czech Science Foundation P506/11/1320 to VH, and by the Grant Agency of the Charles University 63409 to JS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: vlada@natur.cuni.cz (VH); vlcek@img.cas.cz(ČV)

## Introduction

Euglenids are a relatively large group of protists that contains species with different types of feeding strategies: some euglenid species (e.g. *Rhabdomonas*) are osmotrophic and feed by pinocytosis; others developed phagotrophic apparatuses for catching bacteria (e.g. *Entosiphon*) or even eukaryotes (e.g. *Peranema*) [1,2]. One large clade of euglenids is photoautotrophic and its members bear green secondary plastids (e.g. *Euglena gracilis*). The plastid has been subsequently and independently lost in several branches within this clade (*Euglena longa*, prev. *Astasia longa*, *Euglena quartana*, prev. *Khavokinea quartana*, *Euglena hyalina*, *Euglena viridis hyalina* and *Phacus ocellatus*, prev. *Hylophacus ocellatus*) [3–5]. The phototrophic euglenids and their secondary heterotrophic descendents are classified as class Euglenophyceae [4]. Complete plastid genome sequences are known so far for only two closely related euglenid species, *Euglena gracilis* [6] and *Euglena longa* [7].

The fact that plastids are present in a single clade of euglenids favors a hypothesis that the ancestor of this clade acquired the plastid by engulfing a green alga [6,8]. Our current knowledge on the phylogeny of euglenids implies that this endosymbiotic event happened after the split of *Peranema* but before the split of *Eutreptiella* and *Eutreptia*, the basal lineages of the phototrophic clade [9]. This “plastid late” hypothesis is further indirectly

supported by the fact that the autotrophic clade is derived from within the eukaryovorous euglenids; eukaryovory is regarded as the derived feeding mode in euglenids and at the same time it is a useful predisposition facilitating the engulfment of green algae [10]. The plastid of *Euglena gracilis* can be completely lost after bleaching with many environmental and chemical agents without effect on cell viability, and this fact is also used as an argument for a relatively recent acquisition of the plastid, which has not yet been recruited for cellular functions other than photosynthesis [11]. Recent study of introns in the plastid targeting presequences also agrees with the plastid-late hypothesis [12]. An alternative but currently less-accepted plastid-early hypothesis postulates that the euglenid plastid was acquired early in the evolution of euglenids, or even in the common ancestor of euglenids and kinetoplastids (e.g. *Trypanosoma*), their nearest sister group [13,14]. The presence of genes of red algal origin in the photosynthetic *Euglena* as well as in the heterotrophic *Peranema* suggests that the lineage of euglenids might have experienced a cryptic red algal plastid endosymbiosis before the current green algal plastid was established [15].

Analyses of 70 plastidial genes and conservation of gene order on the plastid genome has pointed to *Pyramimonas* (Pyramimonadales, Prasinophyceae) as the closest extant relative of the euglenid plastid [7]. *Pyramimonas* comprises marine flagellates, suggesting that the endosymbiotic event happened in the marine environ-

ment. Although the majority of euglenids live in freshwater, the basal lineage of the autotrophic clade contains the marine species *Eutreptia* and *Eutreptiella*, corroborating the hypothesis of a marine origin of photosynthetic euglenids [4,5]. The comparative analysis of the gene content between the plastid genome of *Pyramimonas parkeae*, which encodes 110 conserved genes (81 protein and 29 RNA species) [16], and *Euglena gracilis*, which comprises 88 conserved genes (58 protein and 30 RNA species) [6], has revealed a substantial loss of genes (for example all genes of NADH-plastoquinone oxidoreductase of the plastidial respiratory chain) happening from the common ancestor of *P. parkeae* and *E. gracilis* to extant *E. gracilis*. This reduction of gene repertoire is explained as a consequence of secondary endosymbiosis, although comparable gene losses took place in the prasinophyte lineages leading to *Pycnococcus* and to the coccoid microalgae *Ostreococcus* and *Monomastix* [16]. Further gene loss in euglenids accompanying the loss of photosynthetic activity has been observed in the closely related but non-photosynthetic *Euglena longa*, which has maintained 56 conserved genes (26 protein and 30 RNA species) [7]. Despite the reduction of coding capacity of the *Euglena* plastid in comparison to that of *P. parkeae*, the size of the *E. gracilis* genome increased (143.2 vs. 101.6 kb in *P. parkeae*). The increase in the genome size should mainly be ascribed to the expansion of self-splicing introns. While *P. parkeae* features a single group II intron, the genome of the *E. gracilis* plastid contains 160 group II and group III introns (15 of which formed twintrons), which is by far the most of all known organellar genomes [17,18]. There are indications that the expansion of introns may be a feature specific to *E. gracilis* and its relatives [17,18]; however, no other plastid genome of euglenids has been completely sequenced, which would be necessary to enable comprehensive comparisons.

Here we report the complete genome sequence of *Eutreptiella gymnastica*, a member of the basal lineage of the photosynthetic clade, and phylogenetically most distant from *Euglena gracilis* – the common ancestor of *E. gracilis* and *E. gymnastica* was the common ancestor of all currently known members of the photosynthetic lineage [1,19,20]. Comparative analysis of the gene content of euglenid plastids allows relatively precisely tracing the events of gene transfers and gene losses accompanying this particular case of secondary endosymbiosis. The vast differences in intron density suggest that the expansion of introns has happened specifically in the lineage leading to *E. gracilis*.

## Results and Discussion

The complete size of the circular chloroplast DNA of *Eutreptiella gymnastica* is 67 622 bp. An overview of the general features of this genome and its closest relatives is given in Table 1. The genome sequence is numbered from the first nucleotide after the second 23S RNA gene (see a physical map of chloroplast DNA – Figure 1). The organization of the genome resembles those of higher plants and algae (including *Pyramimonas parkeae*) with a large single copy region (LSC), a small single copy region (SSC) and two inverted repeats (IR). Simplified maps of plastid genomes of *Eutreptiella gymnastica*, *Euglena gracilis*, *Euglena longa*, and *Pyramimonas parkeae* are illustrated in Figure 2 for comparison.

As is apparent from the genome map (Figure 1), the SSC region is reduced (to 1055 bp), containing only one ORF of unknown function (orf248). This is not surprising, because *E. gymnastica* (like *E. gracilis*) has lost most genes usually found in the SSC region (NADH dehydrogenase complex and a few others). Two of them (rpl32 and psaC) are relocated to other sites. The large single copy region (47 528 bp) contains most genes for proteins and tRNAs.

Two regions resembling inverted repeats (IR, 6304 bp) contain one 16S rRNA gene (1463 bp), one 23S rRNA gene (2999 bp), and a 1726-bp-long sequence with unknown function that contains 2 tandem repeats – VNTR (3×11 bp and 3,4×33 bp). Between the IR copies, the 23S rRNA genes differ in three bases, while all other sequences are identical. The IR copy on the plus strand further contains an insertion of a block of genes (tRNA-Ala, tRNA-Cys, rps2, atpI, atpH, atpF, atpA, and orf372), and the IR copy on the minus strand contains the insertion of tRNA-Ile. The gene cluster of rps2, atpI, atpH, atpF, and atpA found within the IR is one of the ancestral gene clusters conserved in streptophyte and prasinophyte plastid genomes, but it is usually located in the LSC region [16]. The tRNA-Ala and tRNA-Ile genes are present also in the IR of *P. parkeae*.

The inverted repeats do not contain 5S RNA, and in fact *Eutreptiella* lacks it completely. Absence of 5S RNA was also recorded in the plastid genome of *Pyramimonas parkeae* and *Pycnococcus procasolii*, but the possibility exists that its sequence was unrecognized [16]. Interestingly, transcriptional analysis of the *E. gracilis* plastid chromosome showed that, although the genes for 5S, 23S and 16S RNA make one operon [6], the abundance of 5S RNA is much lower than the abundance of 23S and 16S RNA [21]. If 5S RNA is present but remains unrecognized in the plastid genome of *Eutreptiella*, it probably is not localized within the RNA operon, as the 16S RNA gene is very closely followed by neighboring genes. The symmetrical arrangement of tandem repeats in the non-coding part of the IRs suggests that this region may function as the origin of replication. According to the classical model [22], which has recently been challenged [23], the replication of plant and some green algal plastid genomes starts simultaneously from both IRs, and expands unidirectionally towards the SC region, forming two D-loop structures. After it passes the initiation site of the opposing D-loop, the two D-loops fuse to form Cairn-type bidirectional forks that move away from each other and meet approximately 180 degrees from the starting point. *E. gracilis* and *E. longa* plastid genomes lack IRs (Figure 2) and, so far, no model of their replication has been proposed. The origin of replication in the plastid genome of *E. gracilis* has been localized into the region of tandem repeats approximately 6 kb upstream from the extra 16S rRNA gene (Figure 2) [6,24,25]. From this site, the replication probably proceeds in both directions [6]. Because most genes are coded on the leading strand of replication, these genomes have a strikingly non-random distribution of genes. Starting from the ORI site, in one half of the circle, most genes are coded by the plus strand, and in the other half on the minus strand [6,26]. A similar situation is in *Eutreptiella*, but the switch of the coding strands is situated approximately 2/3 of the way through the circle (Figure 1 and 2).

The size of the *E. gymnastica* plastid genome is less than half of that of *E. gracilis*, though the number of conserved genes in both species is not very different (Table 1). The difference in the genome size is caused by different numbers of self-splicing introns. The genome of *E. gracilis* plastid contains 160 group II and group III introns, which is by far the most of all known organellar genomes [6,17]. The plastid genome of *Eutreptiella* apparently contains only two putative introns, and in this respect it resembles the plastid genome of *Pyramimonas parkeae*, which contains only one [16]. We have not found any sequential, structural or positional homology either between the introns of *Eutreptiella* and *Pyramimonas* or between the introns of *Eutreptiella* and *Euglena gracilis*. The first putative intron of *Eutreptiella* (1480 bp) is located in the *psaA* gene. This intron apparently contains an orf386 (1158 bp) that shows very weak homology to reverse transcriptases. The homology is so weak that it was revealed only after iteration in PSI-BLAST. As

**Table 1.** General features of euglenid and *Pyramimonas* cpDNA.

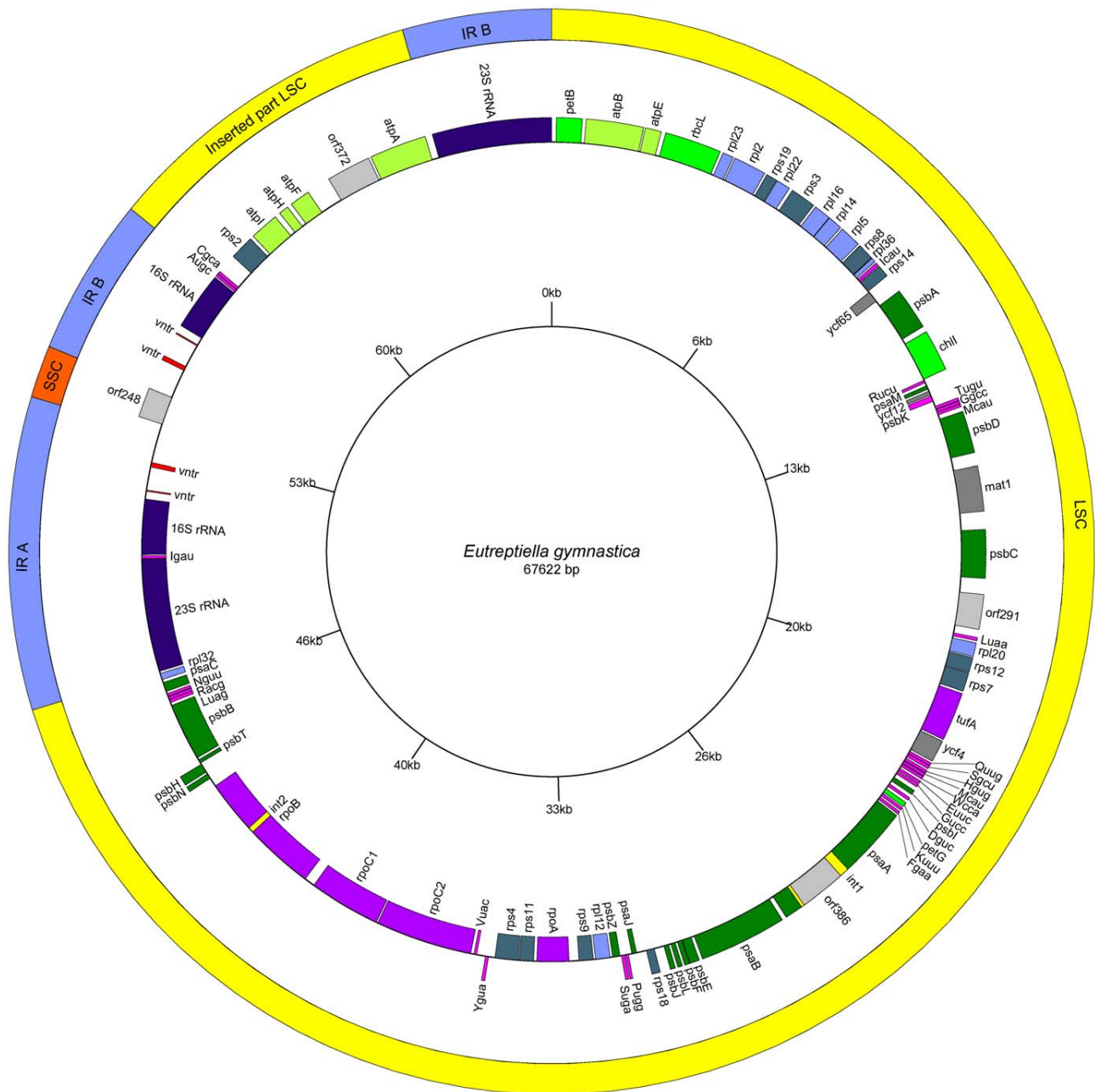
Feature	<i>Eutreptiella gymnastica</i>	<i>Pyramimonas parkae</i>	<i>Euglena gracilis</i>	<i>Euglena longa</i>
Genome size:	67 622	101 605	143 171	73 345
GC percentage:	34,32	34,7	26,13	22,41
Gene-unique loci:	91	123	96	76
Unique rRNA (count/bases):	2/8 924	2/9 086	3/15 057	3/15617
Unique tRNA (count/bases):	26/1 959	27/2 393	27/2 764	27/2122
CDS (conserved genes/all):	59/63	81/94	58/66	26/46
non-spliced (count/bases):	61/38 145	93/69 072	26/15 873	29/15 822
spliced (count/bases):	2/5 511	1/1 467	40/34 449	17/16 299
Introns (count/bases):	2/1 630	1/2 757	160/55 702	61/NA
Density (genes per kb):	0,932	0,925	0,468	0,627
Average length (excl. introns):	692	750	751	698
Coding percentage (excl. introns):	64,5	69,4	35,1	43,7
Intergenic sequences (excl.RNA):	12 614	18 720	25 535	NA
Overlapping sequences:	1 161	1 890	6 209	NA

doi:10.1371/journal.pone.0033746.t001

group II introns often encode for reverse transcriptases, which probably help with their splicing and retroposition [27,28], the homology should be taken seriously. The second intron is much shorter (152 bp), without an ORF, and is located in the *rpoB* gene. The sizes of both introns (excluding ORFs) are smaller than typical group II and longer than group III introns, and we have not been able to find any noticeable similarities in the secondary structure with self-splicing introns in *E. gracilis* or elsewhere. Therefore, their ability to self-splice as well as their exact boundaries should be considered only putative. Besides the *orf386* in intron 1, the *Eutreptiella* plastid genome encodes three other ORFs with homology to reverse transcriptases or intron maturases. Two of them (*orf291* and *orf372*) have no close homologues, and their evolutionary origin cannot be traced. The third (*mat1*) is clearly homologous to *mat1* (*ycf13*) of *E. gracilis* and other euglenids, and in the tree (Figure S1) it forms a well supported branch (98%) with these genes. *Mat1* was apparently present in the last common ancestor of euglenid plastids but interestingly this reverse transcriptase is unrelated to the single reverse transcriptase found in the plastid genome of *Pyramimonas* (*orf608*) (Figure S1). *Mat1* is also remarkably conservative regarding its position in the genome. In almost all investigated euglenids, including relatively closely related *Eutreptia*, it is situated in the internal group III intron of the 4<sup>th</sup> intron in the *psbC* gene [18]. In *Eutreptiella* it is located right next to the *psbC* gene, which in *Eutreptiella* does not contain any intron. *Mat1* was found also in the chloroplast of *E. longa*. As this organism has no *psbC* gene, the *mat1* gene is situated in different loci [7]. The RT and X domains of *E. gracilis* and *E. longa* *mat1* deviate from the consensus sequence of 34 group II intron-encoded proteins [29]. Sequence alignment of *mat1* in *Eutreptiella* and *E. gracilis* shows the loss of at least two conserved domains. Comparison between the genomes of *E. gracilis*, *Eutreptiella* and *Pyramimonas* suggests that the genome of the common ancestor of euglenid plastids was intron-poor but encoded at least one reverse transcriptase (*mat1*). Expansion of introns is apparently a feature specific to *E. gracilis* and its relatives, as already suggested by Thompson et al. [17]. On the other hand, the small number of introns, their unusual sizes and structures and the loss of the otherwise-conserved intron in *psbC* indicate the suppression of introns in *Eutreptiella*. The evidence for the recent horizontal

transfer of a group II intron from a cyanobacterial donor was found in the chloroplast genome of *Euglena myxocylindracea* [30]. This intron (in the *psbA* gene) includes ORF575, named *mat4*, which resembles cyanobacterial reverse transcriptases. *Mat4* is also homologous to the maturase of *Pycnococcus provasolii* and *Volvox carteri* (Figure S1), which is located in an intron of the *atpB* gene [16].

The content of the unique protein coding genes is surprisingly similar between *Euglena gracilis* and *Eutreptiella gymnastica* plastid genomes (Figure 3). The *Eutreptiella* plastid encodes for the same photosynthetic proteins (31), transcription/translation proteins (5), ribosomal proteins (21), and maturase *mat1* as *Euglena gracilis*. There are only 5 extra ORFs in *Eutreptiella* as compared with *E. gracilis* – four ORFs without strong similarity to known proteins (*orf291*, *orf386*, *orf248* and *orf372*) and one conserved protein with homology to *P. parkae* *ycf65* (putative ribosomal protein *rpl3*). Similarly, only eight genes (including intron maturases *mat2*, *roaA* and *orf506*) are specific to *E. gracilis*. Not surprisingly, many of the shared proteins have been lost in *Euglena longa*, whose plastid has lost photosynthetic activity. Given this almost exact match of protein coding capacity of two genomes, whose last common ancestor was at the same time the last common ancestor of all known euglenid plastid genomes, we can with reasonable confidence expect that the *Eutreptiella* plastid genome also matches the coding capacity of this last common ancestor. Using the *Pyramimonas parkae* plastid genome to represent the closest relative to the plastid endosymbiont, we may trace quite precisely the changes in the protein coding capacity of the plastid genome that took place right before and during the process of the secondary endosymbiogenesis. This coding capacity was reduced compared to *Pyramimonas* by the set of genes coding for: 10 proteins of NADH dehydrogenase complex, 2 proteins of cytochrome B<sub>6</sub>F (*petA*, *petN*), 3 proteins of chlorophyll metabolism (*ChlL*, *ChlN*, *ChlB*), heme binding protein *ccsA*, photosystem I subunit *psaI*, initiation factor *infA*, the protease subunit of *clp* protease *clpP*, chloroplast division protein *FtsH*, and several conserved and non-conserved ORFs with unknown function. A BLAST search of 23,372 transcriptome sequences of *E. gracilis* in GenBank and 268 530 transcriptome sequences of *Eutreptiella* produced by us (unpublished data) revealed that transcripts for some of these proteins,

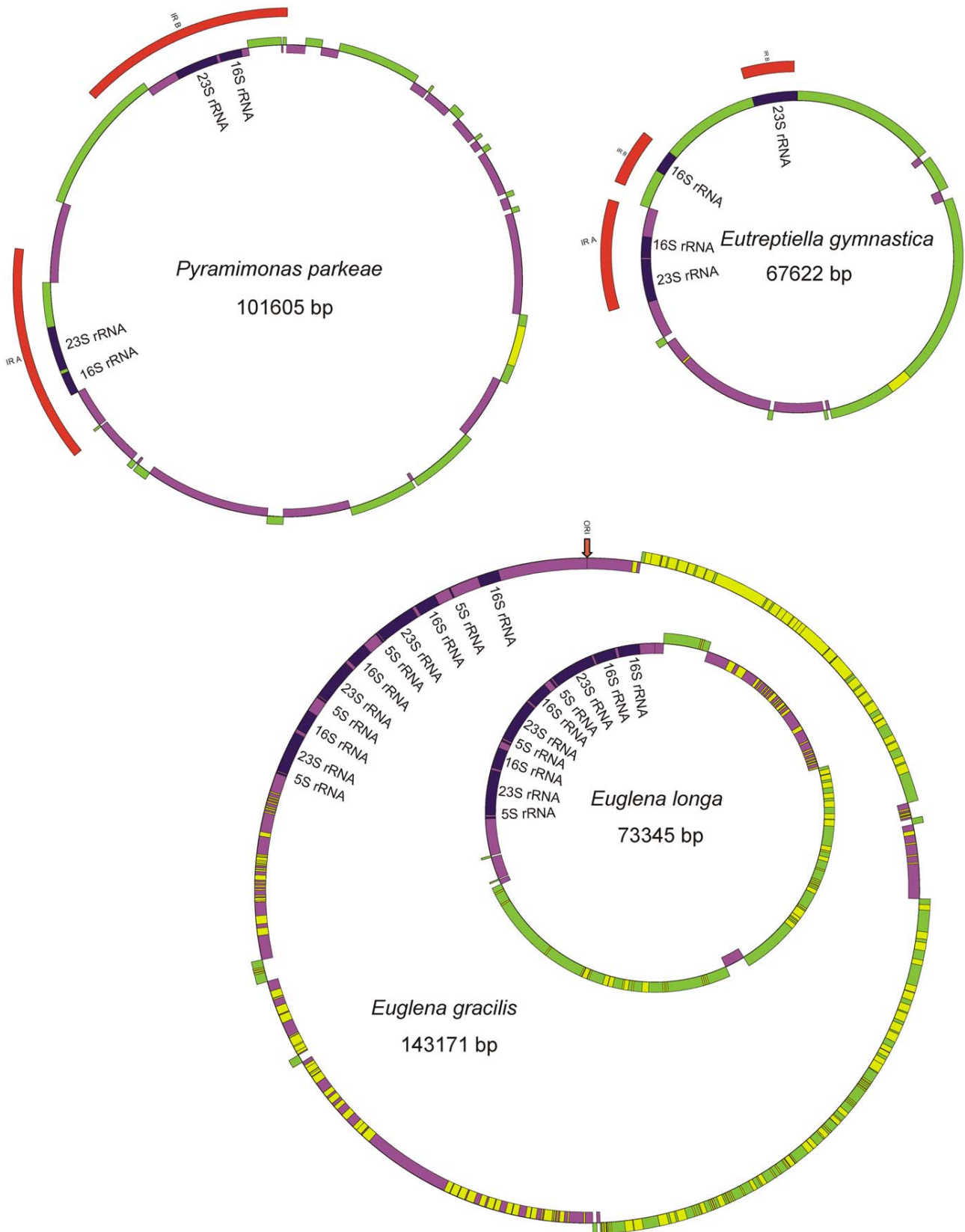


**Figure 1. Map of the plastid genome of *Eutreptiella gymnastica*.** Outer circle shows the large single copy region (LSC) (yellow), short single copy region (SSC) (red) and inverted repeats (IR) (blue). The inner circle shows genes and their division layout in respect to the DNA strands. The genes are color coded according to their function: photosynthesis (shades of green), translation (except maturases) (shades of blue), transcription (violet), tRNA (pink), maturases and unknown function (gray). doi:10.1371/journal.pone.0033746.g001

namely *petA*, *petN*, *ycf3*, *clpP*, and *ftsH*, are present in both transcriptomes, indicating that these genes were probably transferred into the nucleus of the common ancestor of photosynthetic euglenids during the endosymbiogenesis. The gene *ccsA* is present only in the transcriptome of *Euglena*, suggesting that it was transferred into the nucleus of the common ancestor of photosynthetic euglenids, but retained in *Euglena* while probably lost in *Eutreptiella*. The rest of these genes were not found in any transcriptome. Although we cannot rule out the possibility that their transcripts were missed by transcriptome sequencing (e.g. due

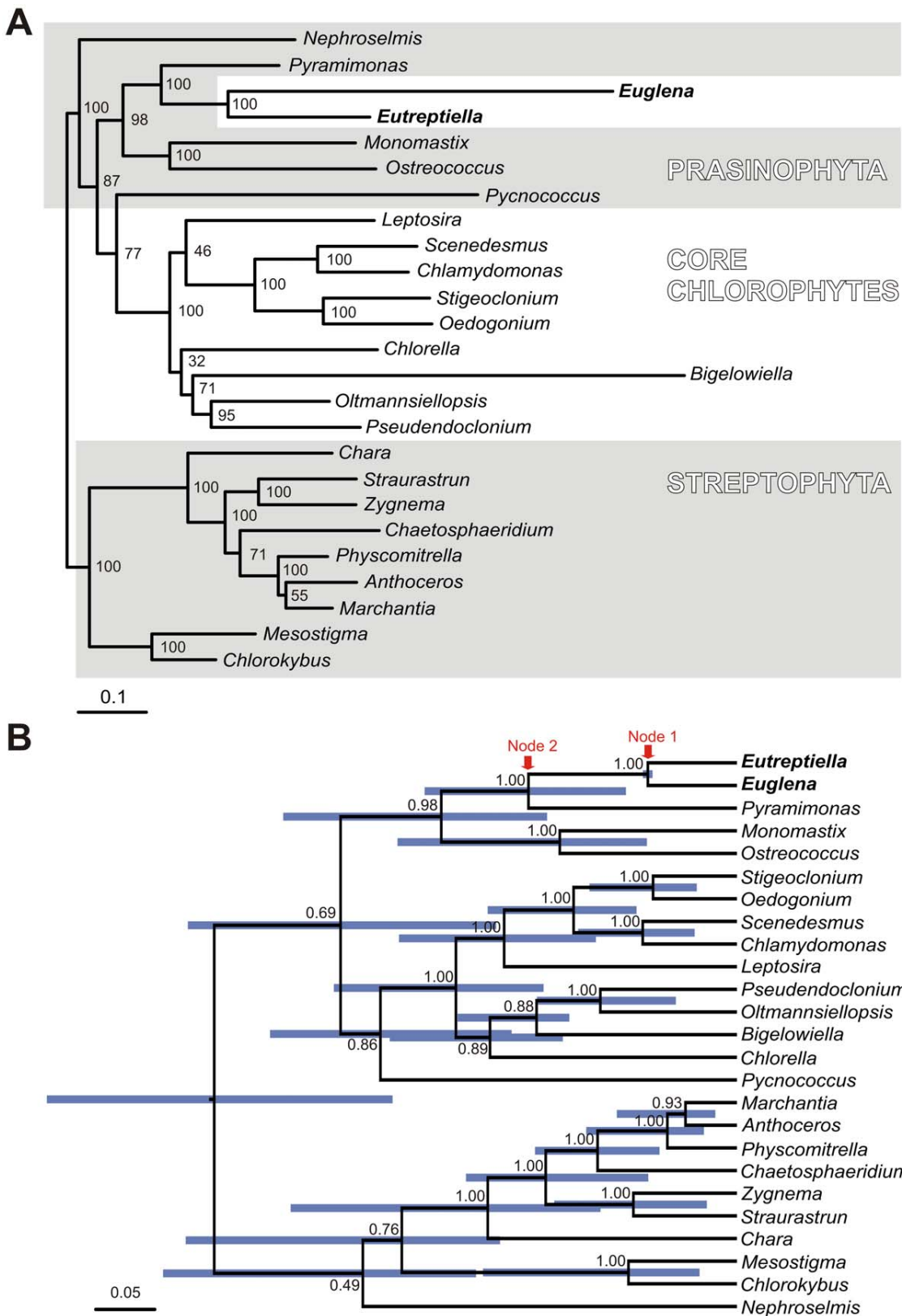
to the low abundance of transcripts), the observations here suggest that they might have been lost completely, either in the evolution of green algal ancestor of euglenid plastid after the split of the *Pyramimonas* branch, or later during endosymbiogenesis itself.

In contrast to the highly conserved gene content of *E. gracilis* and *Eutreptiella gymnastica* plastid genomes, the conservation of gene order is much lower between the two and also in comparison to *Pyramimonas*, indicating that many genome rearrangements have taken place. To get a rough estimate of the degree of gene conservation we counted the number of neighboring gene pairs



**Figure 2. Simplified maps of the plastid genomes of *Eutreptiella gymnastica*, *Euglena gracilis*, *Euglena longa* and *Pyramimonas parkeae*.** The maps are in scale to their sizes. The colors indicate the coding strands (plus-green and minus-violet), the ribosomal RNAs (blue) and introns (yellow). The inverted repeats IRA and IRB in *Pyramimonas* and *Eutreptiella* are marked in red. The ori site in *Euglena gracilis* is marked by an arrow. doi:10.1371/journal.pone.0033746.g002





**Figure 3. Venn diagrams showing overlaps in protein coding capacities between known euglenid plastid genomes and the plastid genome of *Pyramimonas parkeae*.** The schematic representation of genome relationships is indicated in the left. Arrows indicate the probable fate of the genes absent from euglenid genomes. The genes are colour coded in respect to the functional group of their products: housekeeping proteins

(black), proteins involved in photosynthesis (green), maturases of introns (red) and genes with unknown function (gray). Maturases of introns included in the phylogenetic tree of maturases (Figure S1) are marked by asterisks.  
doi:10.1371/journal.pone.0033746.g003

common for pairs of genomes. In this measure, the *E. gracilis* and *Eutreptiella* genomes are the closest as expected, sharing 61 adjacent gene couples; *Pyramimonas* shares with each of them 40 and 38 gene neighbors, respectively.

Phylogenomic analysis of 70 plastid protein coding genes confirmed with maximum bootstrap support the sister relationship of euglenid plastids and *Pyramimonas* (Figure 4) as reported by Turmel et al. [16]. The tip branch of *Euglena gracilis* is almost three times longer than the branch of *Eutreptiella*, probably a result of an accelerated substitution rate in the lineage leading to the genus *Euglena* (Figure 4A). The analyses with relaxed molecular clocks produced ultrametric trees (Figure 4B and Figure S2) that give estimates of relative ages of internal nodes. The branching order of these trees is virtually identical to the maximum likelihood tree. The relaxed clock analyses revealed that the common ancestor of *Euglena* and *Eutreptiella* (node 1 in Figure 4B and Figure S2) was not very recent, as it was approximately as old or older (depending on the clock model) as the common ancestor of vascular plants (common ancestor of *Marchantia*, *Anthoceros* and *Physcomitrella*). It also revealed that the age of the common ancestor of *Pyramimonas* and euglenid plastid (node 2 in Figure 4B and Figure S2), for the three clock models, was 1.2–2.3× older than the common ancestor of *E. gracilis* and *Eutreptiella* if considering the median of the age estimates and 1–5× older if considering the extreme values of the 95% confidence intervals of the age estimates (blue bars in Figures 4B and Figure S2). The time span from node 2 to node 1 was therefore similarly as long as or shorter than the time span from node 1 to the present time, but likely was not markedly longer. The period from node 2 to node 1 includes the green algal lineage that became the direct ancestor of the secondary euglenid plastid and then the stem branch of the secondary plastid before the split of genera *Euglena* and *Eutreptiella*. The exact point where the transition between alga and plastid happened is not known. During this period, 16 protein coding genes functioning in the plastid metabolism were possibly lost and six were transferred to the nucleus of the euglenid. This is in contrast to the at least comparable but very probably quite longer time of evolution that separates extant photosynthetic *E. gracilis* and *Eutreptiella* (twice the time from node 1 to present) during which only one gene (*yef65*) was lost and none was transferred to the nucleus. The rapid slow-down of gene loss could be explained by the fact that the gene set was relatively quickly reduced to an essential core that must be preserved if the photosynthetic function is to be retained. The complete halt of endosymbiotic gene transfer from plastid to the host nucleus is, however, unexpected, as such transfers are also reported in plastids that have been established for a long time in their hosts [31,32]. Unlike the gene content the gene order evolved relatively uniformly – 61 gene couples remained in neighboring positions after the period separating *E. gracilis* and *E. gymnastica*, and correspondingly fewer (40 or 38) gene couples remained positionally fixed to each other after approximately double the period separating *P. parkeae* and *E. gracilis* or *P. parkeae* and *E. gymnastica*.

In conclusion, the plastid genome of *Eutreptiella* turned out to be almost identical to *Euglena gracilis* in protein coding gene content that is reduced when compared to *Pyramimonas*. This indicates that virtually all gene losses and endosymbiotic transfers of genes to the host nucleus took place in the period before the last common ancestor of the euglenid plastid. In contrast to the frozen protein content, the genome organization (gene order, inverted repeats)

diversified significantly in the two sequenced lineages of euglenid plastids, and in the lineage leading to the genus *Euglena* it was furthermore accompanied by an accelerated substitutional rate in protein sequences and the expansion of self splicing introns. We have shown that the method of 454 sequencing could be widely applied to sequencing of organellar genomes.

## Materials and Methods

### Preparation of genomic DNA

A culture of *Eutreptiella gymnastica* strain SCCAP K-0333 was obtained from the Scandinavian Culture Collection of Algae and Protozoa and grown in TL30 medium in 12°C. 150 ml of well-grown culture (approx.  $25 \times 10^6$  cells) was used for DNA isolation. DNA was isolated using the Qiagen Blood and Tissue kit.

### Sequencing and assembly of the plastid genome

1 µg of whole genomic DNA was subjected to 454 sequencing according to GS FLX Rapid Library Preparation Method protocol (Roche). In total 548 056 reads of average size 370 bases were produced. Automatic assembly of reads in Newbler 2.5.3 (Roche) resulted in 19 417 contigs (N50 contig size was 791 bases) and 9.2 Mb of unique sequence. Using a BLASTn homology search it was determined that two contigs, by far the longest (26 365 bp and 20 813 bp), represented parts of the plastid genome. It is expected that contigs derived from the plastid genome should have approximately the same coverage, and so those contigs that had coverage similar to contigs 1 and 2 (35× for contig 1 and 30× for contig 2) were selected from the assembly and all subjected to BLASTn homology search. Five of them were found to represent parts of the plastid genome. All plastid derived contigs were then manually assembled into a 67,274 bp long linear supercontig. Because we expected that the plastid genome would be a circular molecule, a PCR from the ends of the linear supercontig was used to amplify and sequence the missing part (primer F: 5' - taacctgtgaacacgaag -3' and primer R: 5' - caaccagtaagttaggaa -3'). After adding 348 bases the genome was circularized.

### Annotation

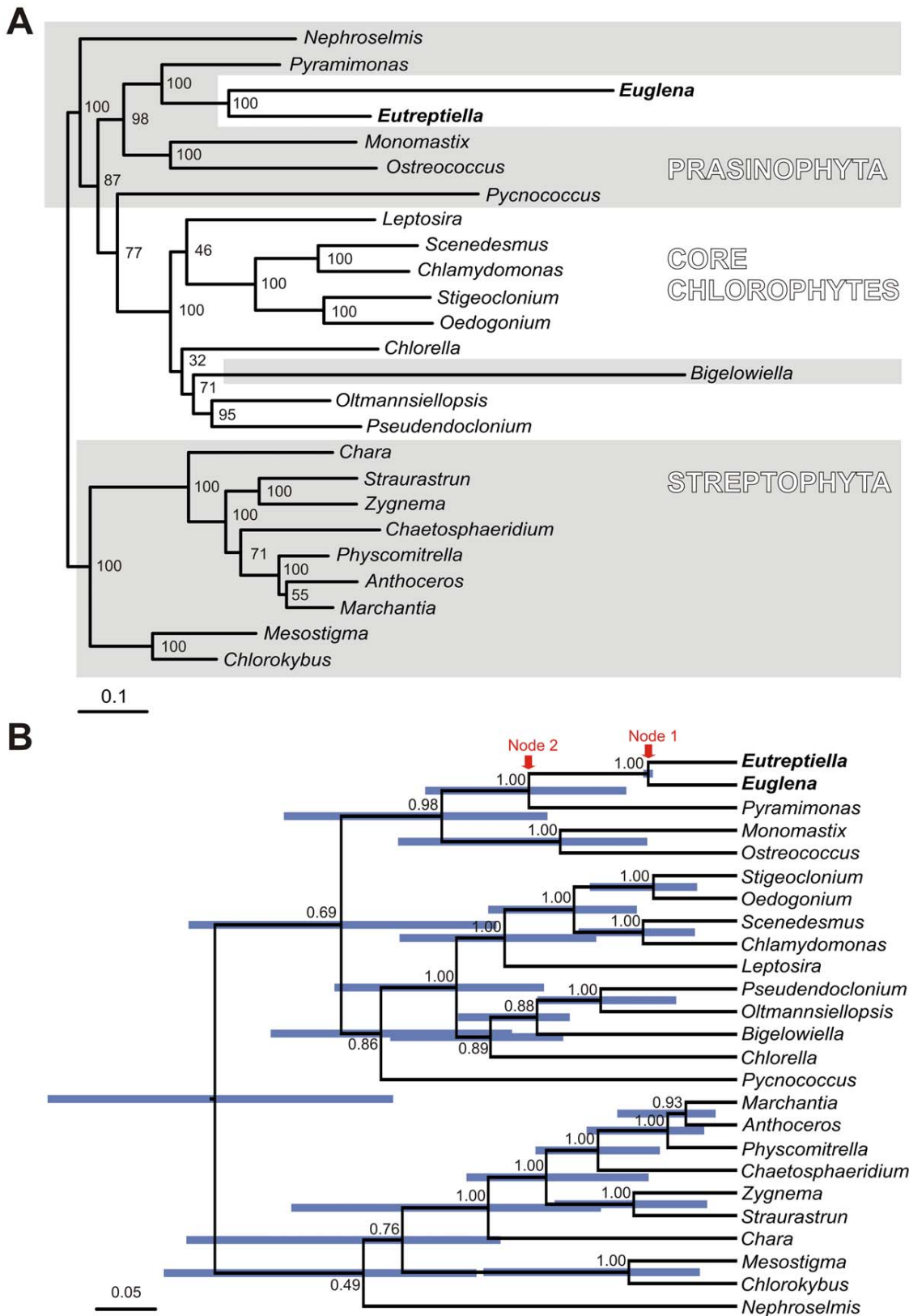
Annotation of ORFs was done using BLASTx homology search. tRNAs were found using tRNA Scan-SE [33], and rRNAs were annotated using a BLASTn homology search with their boundaries determined according to the alignment with rRNA from *Euglena gracilis* and *Pyramimonas parkeae*. The annotation was completed in Artemis 13.2.0 [34] and the annotated genome is deposited in the EMBL database under accession no. HE605038. The genome maps were plotted in GenomeV [35].

### Intron secondary structures

The secondary structures of intron candidates were predicted by mFOLD version 2.3 [36] (<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3>) using the default setting but with the temperature set to 12°C.

### Phylogenetic analyses

The set of maturases was assembled from *Eutreptiella* *mat1* and 121 homologues from GenBank representing both all available euglenid homologues and homologues from other taxa covering



**Figure 4. Phylogenies of plastid genomes of green algae, euglenids and *Bigelowiella* based on 70 genes.** **A.** This phylogenetic tree was constructed using the maximum likelihood method implemented in RAxML, using the LG+I+G model selected by ProtTest. The bootstraps were estimated in 500 replicates. **B.** This tree was constructed in Beast v 1.6.1 using the WAG+I+Γ model of substitution and an uncorrelated exponential



model of relaxed molecular clock. MCMCs were run for  $10 \times 10^6$  generations; trees from the first  $2 \times 10^6$  generations were discarded as the burn-in. Node labels represent posterior probabilities, node bars represent the 95% confidence interval of relative node ages. doi:10.1371/journal.pone.0033746.g004

the sequential diversity of this protein. The data set was aligned using ClustalX [37] and manually edited in Bioedit 7.0.5.3 [38]. The phylogenetic tree was constructed in RAxML v7.2.7 [39] using the PROTGAMMAILGF model. The bootstrap support was calculated using the same model and 500 permutations.

For the phylogenomic analysis we used the data set of 70 protein coding genes from 24 plastid genomes published by Turmel et al [17]. The *Eutreptiella* sequences were manually added to this set in Bioedit 7.0.5.3 [38], realigned using ClustalX [37], and the alignment was then manually edited in Bioedit 7.0.5.3 [38]. The phylogenetic tree was constructed in RAxML v7.2.7 [39] using a uniform PROTGAMMAILGF model for all gene partitions. The bootstrap support was calculated using the same model and 500 permutations. The analyses using relaxed molecular clocks were performed in Beast v 1.6.1 [40] using the WAG+I+ $\Gamma$  model of substitution and three models of relaxed molecular clock: an uncorrelated exponential model, an uncorrelated lognormal model and a random model. MCMC was run for  $10 \times 10^6$  generations; trees from first  $2 \times 10^6$ ,  $7 \times 10^6$  and  $3 \times 10^6$  generations were discarded as the burn-in, respectively.

## Supporting Information

**Figure S1 The phylogeny of intron maturases.** The phylogenetic tree was constructed using the maximum likelihood method implemented in RAxML, using the LG+I+G model selected by ProtTest. The bootstraps were estimated in 500

replicates. The eukaryotic maturases are marked by red, the cyanobacterial are marked by cyan and other bacterial maturases are marked by black.

(DOCX)

**Figure S2 Phylogenies of plastid genomes of green algae, euglenids and *Bigeloviella* based on 70 genes.**

These trees were constructed in Beast v 1.6.1 using the WAG+I+ $\Gamma$  model of substitution and an uncorrelated lognormal model of relaxed molecular clock (A) and random local model of relaxed molecular clock (B). MCMCs were run for  $10 \times 10^6$  generations; trees from the first  $7 \times 10^6$  and  $3 \times 10^6$  generations were discarded as the burn-in in A and B, respectively. Node labels represent posterior probabilities, node bars represent the 95% confidence interval of relative node ages.

(DOCX)

## Acknowledgments

The authors would like to thank to Dr. Steve Zimmerly for e-mail consultations regarding group II and group III introns and Aaron Heiss for language corrections.

## Author Contributions

Conceived and designed the experiments: VH CV. Performed the experiments: JF JS SH. Analyzed the data: SH VH CV. Contributed reagents/materials/analysis tools: CV VH. Wrote the paper: VH SH.

## References

1. Preisfeld A, Busse I, Klingberg M, Talke S, Ruppel HG (2001) Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *Int J Syst Evol Microbiol* 51: 751–8.
2. Leander BS, Esson HJ, Breglia SA (2007) Macroevolution of complex cytoskeletal systems in euglenids. *Bioessays* 29: 987–1000.
3. Linton E, Hittner D, Levandowski CF, Auld T, Triemer RE (1999) A molecular study of euglenoid phylogeny using small subunit rDNA. *J Eukaryot Microbiol* 46: 217–223.
4. Marin B, Palm A, Klingberg M, Melkonian M (2003) Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparison and synapomorphic signatures in the SSU rRNA secondary structure. *Protist* 154: 99–145.
5. Marin B (2004) Origin and fate of chloroplasts in the euglenoida. *Protist* 155: 13–14.
6. Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, et al. (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res* 21: 3537–44.
7. Gockel G, Hachtel W (2000) Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist* 151: 347–51.
8. Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Can J Bot* 56: 2883–2889.
9. Leander BS, Triemer RE, Farmer MA (2001) Character evolution in heterotrophic euglenids. *Eur J Protistol* 37: 337–356.
10. Leander BS (2004) Did trypanosomatid parasites have photosynthetic ancestors? *Trends Microbiol* 12: 251–8.
11. Krajčovič J, Ebringer L, Schwartzbach SD (2002) Reversion of endosymbiosis? In Seckbach J, ed. *Symbiosis: Mechanisms and Models. Cellular Origin in Extreme Habitats*, Vol. 4. Dordrecht: Kluwer Academic Publisher. pp 185–206.
12. Vesteg M, Vacula R, Steiner JM, Mateášiková B, Löffelhardt W, et al. (2010) A possible role for short introns in the acquisition of stroma-targeting peptides in the flagellate *Euglena gracilis*. *DNA Res* 17: 223–231.
13. Hannaert V, Saavedra E, Duffieux F, Szikora JP, Rigden DJ, et al. (2003) Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc Natl Acad Sci USA* 100: 1067–1071.
14. Bodyl A, Mackiewicz P, Milanowski R (2010) Did trypanosomatid parasites contain an eukaryotic alga-derived plastid in their evolutionary past? *J Parasitol* 96: 465–75.
15. Maruyama S, Suzuki T, Weber APM, Archibald JM, Nozaki H (2011) Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology* 11: 105.
16. Turmel M, Gagnon MC, O'Kelly CJ, Otis C, Lemieux C (2009) The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol Biol Evol* 26: 631–48.
17. Thompson MD, Copertino DW, Thompson E, Favreau MR, Hallick RB (1996) Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Res* 23(23): 4745–52. Errata: *Nucleic Acids Res* 24: 542, 24: 1792, 1996.
18. Doetsch NA, Thompson MD, Hallick RB (1998) A maturase-encoding group III twintin is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg? *Mol Biol Evol* 15(1): 76–86.
19. Müllner AN, Angeler DG, Samuel R, Linton EW, Triemer RE (2001) Phylogenetic analysis of phagotrophic, photomorph and osmotrophic euglenoids by using the nuclear 18S rDNA sequence. *Int J Syst Evol Microbiol* 51: 783–91.
20. Busse I, Preisfeld A (2003) Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa). *Int J Syst Evol Microbiol* 53: 617–24.
21. Geimer S, Belicova A, Legen J, Slavikova S, Herrmann RG, et al. (2009) Transcriptome analysis of the *Euglena gracilis* plastid chromosome. *Current Genetics* 55: 425–238.
22. Heinhorst S, Canon GC (1993) DNA replication in chloroplasts. *J Cell Sci* 104: 1–9.
23. Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16: 1661–6.
24. Ravel-Chapuis P, Heizmann P, Nigon V (1982) Electron microscopic localization of the replication origin of *Euglena gracilis* chloroplast DNA. *Nature* 300: 78–81.
25. Koller B, Delius H (1982) Origin of replication in chloroplast DNA of *Euglena gracilis* located close to the region of variable size. *EMBO J* 1: 995–8.
26. Morton BR (1999) Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*. *Proc Natl Acad Sci U S A* 96: 5123–8.
27. Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82: 5–30.
28. Michel F, Ferat JL (1995) Structure and activities of group II introns. *Annu Rev Biochem* 64: 435–61.
29. Mohr G, Perlman PS, Lambowitz AM (1993) Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res* 21: 4991–7.
30. Sheveleva EV, Hallick RB (2004) Recent horizontal transfer to a chloroplast genome. *Nucleic Acids Res* 32: 803–810.

31. Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422: 72–6.
32. Bock R, Timmis JN (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* 30: 556–66.
33. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucl Acids Res* 25: 955–964.
34. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–5.
35. Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24: 861–2.
36. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406–3415.
37. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
38. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.
39. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–90.
40. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.





## 5 *Pyramimonas parkeae* – donor plastidu

Neopominutelnou součástí zkoumání sekundární endosymbiózy u euglen je zmapování genomů pravděpodobného potomka blízkého příbuzného pohlčené řasy, za kterého je momentálně považována prasinofytní řasa *Pyramimonas parkeae*. Znalost jeho plastomu, mitogenomu a jaderného genomu a jejich porovnání s genomem resp. transkriptomem euglen nám může objasnit, jakým způsobem probíhalo formování plastidu včetně endosymbiotického přenosu genů.

### 5.1 Prasinofytní řasy

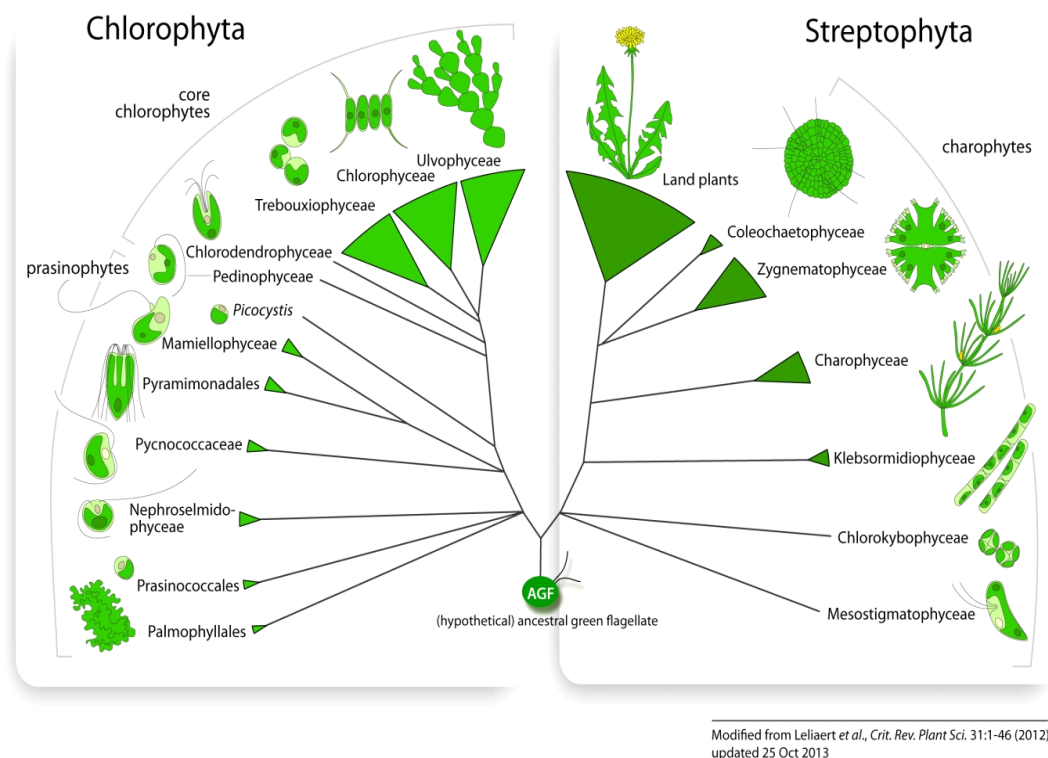
Prasinofytní řasy představují nezanedbatelnou součást fytoplanktonu a pikoplanktonu; většinou jsou to volně žijící jednobuněčné organismy, především mořské, s 1 – 16 bičíky nebo bez nich. Nemají buněčnou stěnu, ale některé linie mají povrch pokrytý polysacharidovými šupinami. Mají jeden chloroplast s fotosyntetickými pigmenty chlorofylem *a* a *b*, některé mají i specifický xantofyl prasinoxantin. *P. parkeae* je jednobuněčná mořská řasa se 4 bičíky mající některé typické prasinofytní rysy, jako je povrch buňky včetně bičíků pokrytý šupinami a jediný chloroplast s pyrenoidem (Adl et al., 2019).

Prasinophytes představují parafyletickou skupinu v kládu Chlorophyta, který společně s kládem Streptophyta tvoří Chloroplastida (Viridiplantae). Kromě prasinofytních tříd zahrnuje linie Chlorophyta ještě několik tříd označovaných jako "core chlorophytes" neboli "zelených řas" *sensu stricto* (zejména třídy Chlorophyceae, Ulvophyceae, Trebouxiophyceae a dále Pedinophyceae a Chlorodendrophyceae, dříve řazené jako klád IV mezi prasinofyta). Prasinophytes byly původně rozděleny do 9 subkládů, které více méně odpovídají nynějším řádům (Adl et al., 2019) a jejichž vzájemné postavení není zcela upřesněno. Jsou to sesterské Chloropicophyceae (klád VII a, b) a Picocystophyceae (klád VII c), sesterské Mamiellophyceae (klád II) a Pyramimonadales (klád I) větvcí společně s Pycnococcaceae (klád V), a dále Nephroselmidophyceae (klád III) a Palmophyllophyceae (Palmophyllales a Prasinococcales, klád VI) a dva klády VIII a IX, které zahrnují pouze environmentální vzorky (Adl et al., 2019; Guillou et al., 2004; Leliaert et al., 2012; Lopes Dos Santos et al., 2017; Viprey et al., 2008) (obr.9).

Podle nejnovějších fylogenetických analýz by měla být ustanovena třetí sesterská skupina ke skupinám Chlorophyta a Streptophyta, a to Prasinodermophyta, která zahrnuje stávající klád VI (Palmophyllales a Prasinococcales) (Li et al., 2020).

Fylogenetické analýzy vycházejí většinou ze sekvencí genů pro 18S rRNA (Guillou et al., 2004; Viprey et al., 2008), plastidových genů (Lemieux et al., 2014a; Turmel et al., 2009), mitochondriálních genů (Hrdá et al., 2017; Turmel, 1999) nebo jejich kombinací (Leliaert et al., 2012). Neúplné jaderné genomy jsou dostupné pouze u tří rodů Mamiellophyceae (*Ostroccoccus*, *Bathycoccus* a *Micromonas*), jednoho rodu Pyramimonadales (*Cymbomonas*), jednoho rodu Chloropicophyceae (*Chloropicon*) a jednoho rodu Picocystophyceae (*Picocystis*). Organelových

genomů prasinofyt je osekvenováno více, okolo 20 plastomů a 20 mitogenomů, které obojí pokrývají všech sedm výše jmenovaných řádů (dle NCBI 06/2020).



Obr. 9. Přehled fylogeneze skupiny Chloroplastida (převzato z Leliaert et al., 2012).

## 5.2 Plastidové genomy prasinofyt

Znamé plastidové genomy prasinofyt jsou spíše menší než plastomy ostatních chloroplastid. Tvoří je kruhová molekula mající délku 65 – 115 kbp (výjimku tvoří *Nephroselmis* s délkou 200 kbp) a obsahují 86 – 128 genů, což je spíše více ve větvi Chlorophyta, ale je to méně, než mají zástupci větve Streptophyta. Takový rozptyl v délce a genovém zastoupení je daný tím, že část organismů má ancestrální typ organizace genomu s kvadripartitní strukturou a velkým zastoupením genů v původních klastrech (např. *Nephroselmis*, *Pyramimonas*), ale část organismů má naopak odvozený redukováný a restrukturalizovaný genom (např. *Ostreococcus*), který začasťe postrádá IR a tedy i kvadripartitní strukturu (např. *Pycnococcus*). Obecně jsou ale všechny tyto genomy kompaktní a AT bohaté (55 – 65 %), mají krátké intergenové oblasti, spíše kratší IR (pokud jsou přítomné) a obsahují málo intronů a repetitivních sekvencí (Lang & Nedelcu, 2012; Leliaert et al., 2012; Turmel & Lemieux, 2018). Největší obsah genů byl zatím nalezen u rodu *Nephroselmis*. Porovnáním dvou druhů *N. olivacea* a *N. astigmatica* se zjistilo, že každý z těchto druhů nese několik genů, které nejsou přítomné u druhého, ale

celkově je obsah genů velice podobný genům nalezeným v plastidech streptofyt *Mesostigma* a *Chlorokybus* (Lemieux et al., 2014b; Turmel et al., 1999). Pořadí genů je velice variabilní, a to jak mezi jednotlivými liniemi, tak v rámci linií. Časté jsou také nezávislé ztráty IR nebo posun jejich hranic, což vedlo ke změnám v obsahu genů v IR. Právě přechody IR a single copy oblastí jsou zdrojem velké variability (Lemieux et al., 2014b; Turmel & Lemieux, 2018).

### 5.3 Plastidový genom *Pyramimonas parkeae*

První kompletní sekvence chloroplastového genomu *P. parkeae* kmene CCMP726 byla anotována v roce 2009 (Turmel et al., 2009), druhá sekvence *P. parkeae* kmene NIES254 pak v roce 2017 (Satjarak & Graham, 2017a). Oba genomy jsou si velice podobné a tvoří kruhovou molekulu se dvěma invertovanými repeticemi, mají tedy typickou kvadripartitní strukturu. Délky plastomů jsou 101,6 resp. 104,8 kbp. Rozdíl v délce je dán především rozdílnou délkou intergenových oblastí. Shodně obsahují 81 genů kódujících proteiny, 27 tRNA a malou (*rrs*) a velkou (*rrl*) podjednotku rRNA, nikoli však 5S rRNA. Genomy jsou kolineární, pořadí genů v obou si odpovídá. Přesto v nich jsou celkem čtyři oblasti – "hot spoty", kde je podobnost téměř nulová. Jedná se především o intergenovou část v LSC oblasti (mezi geny *psbA-trnS* a *ndhB*), obě rozhraní mezi SSC oblastí a invertovanými repeticemi a dále intron v genu *atpB*. Všechny tyto oblasti obsahují několik otevřených čtecích rámců, které jsou až na výjimku *orf91* a částečně *orf454* naprosto rozdílné (Satjarak & Graham, 2017a). Zajímavostí *orf454* je, že vykazuje sekvenční podobnost s řádovou DNA primasou, kterou řasa zřejmě získala horizontálním genovým přenosem od viru, který u ní byl také pozorován (Turmel et al., 2009).

Sekvenčně rozdílné jsou také otevřené čtecí rámce, které jsou součástí jediného intronu typu II v genu *atpB*. U obou kmenů sice kódují reverzní transkriptázu, jejich délka je však různá (608 resp. 520 aminokyselin) a sekvenční podobnost téměř žádná, stejně jako podobnost samotných intronů a místo vložení do genu (vlastní zjištění). Naše nepublikovaná data z neúplného plastomu třetího kmene *P. parkeae* SCCAP K-007 také vykazují ve výše uvedených "hot spotech" rozdíly v délce a nukleotidovém složení. Kmen SCCAP K-007 je ale patrně podobnější kmeni CCMP726, protože intron s ORF v genu *atpB* je až na několik málo nukleotidů stejný.

### 5.4 Plastidové geny *Pyramimonas parkeae* u euglenofyt

Jak již bylo popsáno výše, společný předek *P. parkeae* a plastidu euglenofyt měl pravděpodobně minimálně 111 (110 dnes přítomných u *P. parkeae* a gen pro 5S rRNA přítomný u většiny euglenofyt). Z toho 89 genů dnes zůstalo v plastomu euglenofyt a ze zbylých 22 jsme našli 5 v transkriptomu i plastidovém proteomu *E. gracilis* a transkriptomu *Etl. gymnastica*. Jedná se o geny *ccsA*, *fstH*, *infA*, *YCF3* a *petA*. Posledně jmenovaný kóduje cytochrom *f*, který je součástí chloroplastového *bf* komplexu, což je membránový komplex, přenášející elektrony mezi fotosystémem II a I. Přítomnost funkčního proteinu pro cytochrom *f* v chloroplastech i přítomnost mRNA včetně sekvence pro target peptid byla ověřena laboratorně (Santillán Torres et al., 2003).

## 5.5 Mitochondriální genomy prasinofyt

Anotované mitochondriální genomy prasinofyt mají podobné vlastnosti, jako genomy plastidové. Většinou jsou anotované jako kruhové molekuly. Jsou spíše menší než mitogenomy ostatních chloroplastid (s výjimkou malých genomů Chlorophyceae), mají délku 38 – 61 kbp a obsahují 55 – 65 genů (výjimku tvoří redukovaný *Pycnococcus* 24,3 kbp / 36 genů). Jsou AT bohaté (55 – 70 %) a mají výrazně kratší intergenové oblasti než ostatní chlorofyta i než streptofyta (Burger & Nedelcu, 2012; Leliaert et al., 2012; Martínez-Alberola et al., 2019). Nalezneme u nich některé typické starobylé klastry genů, především dobře zachovaný klastr pro geny kódující ribozomální proteiny (Turmel, 1999). U prasinofyt z kládu II a III byl nalezen i gen pro *RNasu P RNA*, který velice připomíná tento gen nalezený v mitogenomech heterotrofních jakobidů, které vykazují některé původní znaky (Burger et al., 2013; Burger & Nedelcu, 2012). U většiny prasinofytních genomů chybí gen pro tRNA (ucg) pro arginin. Výjimečně byl tento gen nahrazen jiným duplikovaným genem pro tRNA, jmenovitě byl tento nový gen nalezen u prasinofytního druhu *N. olivacea* a streptofyt *Mesostigma viride*, *Chlorokybus atmophyticus* a *Marchantia polymorpha* (Wang et al., 2009).

Mitogenomy prasinofyt neobsahují žádné nebo jen velice málo intronů typu I (*Monomastix*, *Nephroselmis*, *Prasinoderma*), výjimečně typu II (*Pyramimonas*, *Cymbomonas*) (Hrdá et al., 2017; Martínez-Alberola et al., 2019; Satjarak et al., 2017). To kontrastuje se situací u streptofyt, kde může být intronů velice mnoho, predominantním typem intronu je typ II, ale není výjimečné, že i v rámci jednoho genu byly nalezeny oba typy intronů (např. v genu *coxI* u *M. polymorpha*) (Lang et al., 2007; Ohta et al., 1993).

Prasinofyta jsou jednou z několika málo skupin, jejichž mitochondriální genomy obsahují invertované repetice. Některé druhy prasinofyt obsahují různě dlouhé IR obsahující 0 – 6 genů kódujících proteiny, 1 – 13 tRNA a geny pro velkou a malou podjednotku rRNA, z nichž jedna nebo obě mohou chybět, nebo jsou přítomny v několika částech (Hrdá et al., 2017; Satjarak et al., 2017; Turmel et al., 2019). Přítomnost či nepřítomnost IR nekopíruje fylogenetickou příbuznost skupin, ale nebyla zatím blíže studována.

## 5.6 Mitochondriální genom *Pyramimonas parkeae*

V roce 2017 byly anotovány dvě kompletní sekvence mitochondriálního genomu *P. parkeae*, jedna sekvence je mitogenom kmene NIES254 (Satjarak et al., 2017), druhá sekvence je mitogenom kmene SCCAP K-007 osekvenovaný v naší laboratoři, který je druhým výstupem mé disertační práce (Hrdá et al., 2017). Také tyto genomy jsou si podobné, tvoří je kruhová molekula se dvěma invertovanými repeticemi, mají tedy kvadripartitní strukturu. Délky genomů jsou 43,3 resp. 53,4 kbp, obsah AT je 68,8 % resp. 69,5 %, kódující oblasti tvoří 81,7 % resp. 63,3 %.

Genomy jsou zcela kolineární, obsahují 34 genů kódujících proteiny, 25 tRNA a malou (rrs) a velkou (rrl) podjednotku rRNA, ani mitogenomy neobsahují 5S rRNA. Jednotlivé geny mají



velice podobnou délku. Délka mezigenových oblastí mezi jednotlivými geny je také podobná a je velmi malá, okolo 80 bp. Genomy se ovšem výrazně liší v délce ve třech "hot spot" oblastech. Délka a obsah těchto oblastí je také důvodem rozdílů v celkové délce genomů i v procentuálním podílu kódujících oblastí. Zatímco genom kmene SCCAP K-007 je celkově velice kompaktně uspořádán, u kmene NIES254 jsou na třech místech vloženy delší nekódující oblasti. Jedna oblast je mezi geny *coxI* a *trnR/cob*, kde je vložen úsek dlouhý kolem 2,9 kbp. Další dva úseky dlouhé 2,4 kbp jsou vloženy do obou invertovaných repetitiv doprostřed, neobsahují ale žádný gen, pouze dva ORF neznámé funkce.

Invertované repetice obsahují šest genů kódujících proteiny (*nad1*, *nad2*, *nad6*, *rps2*, *rps4* a *rps12*) a 10 genů pro tRNA (*trn E*, *F*, *G*, *I*, *M*, *N*, *P*, *Q*, *R*, *W*). Nenachází se v nich však ani jedna rRNA podjednotka, které jsou jinak součástí IR, pokud jsou přítomné. U druhu *Cymbomonas tetramitiformis*, který patří společně s *P. parkeae* do prasinofytního kládu I, je obsah IR pro změnu omezen jen na geny pro malou podjednotku rRNA a tři různé tRNA (Satjara et al., 2017).

V porovnání s *Nephroselmis olivacea* a *Ostreococcus tauri*, které mají mezi prasinofytními mitogenomy nejbohatší genový repertoár, nebyly v mitochondriálních genomech *P. parkeae* nalezeny pouze čtyři geny (5S rRNA, *RNaza P RNA*, *rps8* a tRNA pro arginin). Některé geny se překrývají a další jsou ve velice blízkém sousedství, což napovídá, že by se mohly přepisovat jako jeden polycistronický operon, čímž by byla zachována proporcionalita produkce těchto genů. Jedná se především o operon pro ribozomální proteiny *rps12* – *rps11*, který je odvozen od bakteriálního operonu, je velice starobylý a zachovává pořadí genů napříč mnoha mitogenomy, od *Reclinomonas americana* přes *Acanthamoeba castelani* po prasinofyta (Burger et al., 1995; Burger et al., 2013; Hrdá et al., 2017; Satjara et al., 2017; Turmel et al., 1999).

Specifickým znakem mitogenu *P. parkeae* je přítomnost intronů typu II, které se nacházejí v genu *coxI*. Dva exony *coxI* jsou umístěny na vlákně DNA v opačném pořadí, exon 2 bezprostředně předchází exonu 1 (obr. 10). Ke vzniku maturované mRNA je třeba mechanismem trans-splicing spojit 3' konec exonu 1 s 5' koncem exonu 2. Na hranicích exon/intron byly identifikovány příslušné motivy, které odpovídají koncům typického intronu typu II. Většina domén intronu (viz obr. 8) nalezena nebyla. Mechanismus procesu není znám, ale k sestřihu zjevně dochází, jak dokazuje přítomnost správně sestřižené mRNA.

U kmene SCCAP K-007 jsme popsali pouze tento jeden intron. U kmene NIES254 byl jako gen *coxI* anotován pouze zmíněný exon 2 a exon 1 nebyl identifikován. Po důkladném prověření úseku mezi "exonem 2" genu *coxI* a následujícími geny *trnR/cob*, se ukázalo, že úsek odpovídající exonu 1 je zde také přítomen a odpovídá svou sekvencí, délkou i hranicemi exon/intron přesně exonu 1 u kmene SCCAP K-007 a i zde tedy zřejmě dochází k trans-splicing. Situace je ovšem o něco složitější, neboť tato nově identifikovaná část genu *coxI* je ještě rozdělena dalším vloženým intronem typu II. Tento intron je sestřiháván klasicky cis-splicingem, má standardní strukturu intronu typu II včetně vloženého ORF s funkcí reverzní transkriptázy/maturázy. Tento ORF má podobnost s reverzní transkriptázou/maturázou

nalezenou v genu *coxI* u některých zástupců chlorofyt a streptofyt, ale i bacilariofyt a metazoi. Intron je dlouhý 2534 bází. V genu *coxI* u *C. tetramitiformis*, který patří do stejného prasinofytního řádu, jsou dva introny typu II, které jsou ale vloženy do jiných míst genu a jejichž ORF jsou podobné reverzním transkriptázám různých druhů bacilariofyt (vlastní zjištění).



Obr. 10. Schéma rozložení exonů a intronů v genu *coxI* u dvou kmenů *P. parkeae*. Oranžově intron typu II sestřihávaný cis-splicingem, žlutě koncové části intron typu II sestřihávaného trans-splicingem (# a \* značí místa sestřihu).

## 5.7 Mitochondriální geny *Pyramimonas parkeae* u euglenofyt

Prasinofytní mitochondrie se u euglenofyt nedochovala. Ani v transkriptomu *E. gracilis* a *Etl. gymnastica* nebyly nalezeny žádné geny, které by byly přeneseny z mitochondrie donora plastidu (Hrdá et al., 2017). Druhá sada mitochondriálních genů je zřejmě pro organismy s komplexním plastidem zbytná, a proto se až na vzácné výjimky nedochovala. Tuto výjimku tvoří dinoflageláti řádu Kryptoperidiniaceae *Durinskia* a *Kryptoperidinium*, které získaly svůj komplexní plastid od rozsivek (Bacillariophyceae). Jejich endosymbiont si mimo jiných buněčných struktur (plastid, jádro, endoplasmatické retikulum, ribozómy) ponechal i mitochondrii včetně jejího genomu. Mitochondriální genomy nebyly zatím celé charakterizovány, až několik genů. Geny *coxI*, *coxII*, *coxIII*, *cob* a *rln* z endosymbionta a geny *coxI*, *cob* a *rln* hostitele jsou v genomech přítomny a také transkribovány. Zdá se tedy pravděpodobné, že se obě organely překrývají ve své funkci (Imanian et al., 2007).

## **5.8 Publikace II**

**Hrdá, Š., Hroudová, M., Vlček, Č., & Hampl, V. (2017).**

**Mitochondrial genome of prasinophyte alga *Pyramimonas parkeae*.**

**Journal of Eukaryotic Microbiology, 64(3), 360-369.**

**<https://doi.org/10.1111/jeu.12371>**



## ORIGINAL ARTICLE

**Mitochondrial Genome of Prasinophyte Alga *Pyramimonas parkeae***Štěpánka Hrdá<sup>a</sup>, Miluše Hroudová<sup>b</sup>, Čestmír Vlček<sup>b</sup> & Vladimír Hampel<sup>a</sup><sup>a</sup> Department of Parasitology, Faculty of Science, Charles University, Prague 128 43, Czech Republic<sup>b</sup> Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague 142 20, Czech Republic**Keywords***cox1*; intron; inverted repeats; phylogeny; trans-splicing.**Correspondence**V. Hampel, Department of Parasitology, Faculty of Science, Charles University, Vinicna 7, Prague 128 43, Czech Republic  
Telephone number: +420-221951074;  
FAX number: +420-221949704;  
e-mail: vladimir.hampel@natur.cuni.cz

Received: 23 June 2016; revised 7 September 2016; accepted September 14, 2016.

doi:10.1111/jeu.12371

**ABSTRACT**

Prasinophytes are a paraphyletic assemblage of nine heterogeneous lineages in the Chlorophyta clade of Archaeplastida. Until now, seven complete mitochondrial genomes have been sequenced from four prasinophyte lineages. Here, we report the mitochondrial genome of *Pyramimonas parkeae*, the first representative of the prasinophyte clade I. The circular-mapping molecule is 43,294 bp long, AT rich (68.8%), very compact and it comprises two 6,671 bp long inverted repeat regions. The gene content is slightly smaller than the gene-richest prasinophyte mitochondrial genomes. The single identified intron is located in the cytochrome c oxidase subunit 1 gene (*cox1*). Interestingly, two exons of *cox1* are encoded on the same strand of DNA in the reverse order and the mature mRNA is formed by trans-splicing. The phylogenetic analysis using the data set of 6,037 positions assembled from 34 mtDNA-encoded proteins of 48 green algae and plants is not in compliance with the branching order of prasinophyte clades revealed on the basis of 18S rRNA genes and cpDNA-encoded proteins. However, the phylogenetic analyses based on all three genomic elements support the sister position of prasinophyte clades Pyramimonadales and Mamiellales.

PRASINOPHYTES represent a paraphyletic assemblage in the clade of Chlorophyta (Viridiplantae) (Guillou et al. 2004; Turmel et al. 2009) and as such they comprise heterogeneous lineages of mainly marine unicellular green algae with coccoid morphology (e.g. genus *Ostreococcus*) or flagellates with 1–16 flagella (e.g. *Pyramimonas parkeae*). Cells of some genera (*Ostreococcus*, *Micromonas*) are very small (2–3 µm in length) and represent the smallest known eukaryotic phototrophs. In some cases (e.g. *Pyr. parkeae*), the surface of the cell including the surface of flagella is covered by scales. Based on the analysis of the nucleus-encoded 18S rRNA gene, prasinophytes were split into nine subclades: clade I—Pyramimonadales, clade II—Mamiellales, clade III—Nephroselmidophyceae, clade IV—Chlorodendrophyceae, clade V—Pycnococcaceae, clade VI—Prasinococcales, clade VII—*Picocystis* and its relatives (Guillou et al. 2004) and clades VIII and IX composed entirely of environmental sequences (Viprey et al. 2008). The monophyly of the Chlorophyta clade is well-supported in phylogenetic analyses based on 18S rRNA gene, however, the branching order of prasinophyte

lineages has not been resolved with confidence despite the rising number of sequenced species and sequences derived from environmental samples. The only robustly supported relationship in the 18S rRNA gene tree is the sister position of the clade IV—Chlorodendrophyceae and core chlorophytes (Guillou et al. 2004). The Pyramimonadales clade splits in two lineages, the first comprises genera *Cymbomonas* and *Halosphaera*, the second consists of genera *Pterosperma* and *Pyramimonas*. The latter genus was shown to be polyphyletic and was therefore subdivided into six subgenera (Suda et al. 2013). *Pyramimonas (Trichocystis) parkeae* is a marine species with cell size 20 × 15 µm and four anterior flagella.

The presence of plastid and mitochondrial genomes provides an opportunity to analyse phylogenetic relationships between the prasinophyte lineages using genes encoded in these genomes. The published plastid genomes are very divergent in size, gene content and gene order. All genomes map to a circle that may contain inverted repeats (IR) separated by large and small single copy regions (LSC, SSC). This arrangement was found in

cpDNA of *Nephroselmis olivacea* and *Pyr. parkeae* and it seems to be ancestral as it is present also in land plants (Jansen and Ruhlman 2012) and in ulvophyte alga *Oltmannsiellopsis viridis* (Pombert et al. 2006b). Some genomes are intronless (*N. olivacea*), but most of them contain one type II intron (*Pyr. parkeae*, *Pycnococcus provasolii*, *Monomastix* sp.) and rather exceptionally contain also several type I introns (*Monomastix* sp.) (Turmel et al. 1999, 2009). The plastid genome of *Pyr. parkeae* harbours one intron (type IIA) in the gene for ATP synthase subunit beta (*atpB*) with ORF608 containing domains for reverse transcriptase and maturase. Product of this ORF shares sequence similarity with the reverse transcriptases present in the genomes of firmicute bacteria and in an intron of the mitochondrial *cox1* gene of fungi (*Podospira anserina*, *Allomyces macrogynus*), rhodophytes (*Pyropia perforata*), cryptophyte *Rhodomonas salina* and the brown alga *Pylaiella littoralis* (Turmel et al. 2009).

Phylogenomic analyses of 70 plastid-encoded genes from 9 streptophytes, 13 chlorophytes, *Bigeloviella natans*, *Euglena gracilis* and *Eutreptiella gymnastica* identified *Pyr. parkeae* as the closest extant relative of the euglenid plastid (Hrdá et al. 2012; Turmel et al. 2009) and revealed the sister relationship between Pyramimonadales (clade I) and Mamiellales (clade II). The more recent phylogenetic analysis of cpDNA using the data set of 79 genes (15,549 positions) assembled from cpDNA-encoded proteins of 34 green algae and plants are essentially congruent with the branching order of prasinophyte clades revealed on the basis of 18S rRNA genes. Twelve sampled prasinophyte representatives from six clades (I–VII except IV) formed six independent branches of Chlorophyta. The clade of *Pedinomonas minor* and core chlorophytes (Chlorophyceae, Ulvophyceae and Trebouxiophyceae) branched within the prasinophyte clade VII sister to Prasinophyceae sp. CCMP 1205. Clades II, III and VI were well-supported as well as the sister position of clades I and II (Pyramimonadales and Mamiellales) (Lemieux et al. 2014).

The comparison of the third genetic element—mitochondrial genome—provides a chance to confirm and elaborate on the phylogeny of prasinophytes. Seven complete mitochondrial genomes that belong to four lineages of prasinophytes have been sequenced to date. Four of them represent clade II—*Ostreococcus tauri* (Robbens et al. 2007), *Micromonas commoda* (RCC299) (van Baren et al. 2016; Worden et al. 2009), *Bathycoccus prasinus* (Moreau et al. 2012) and *Monomastix* sp.OKE-1 (Turmel et al. 2013), clade III is represented by *N. olivacea* (Turmel 1999), clade V by *Pyc. provasolii* (Turmel et al. 2010) and clade VI by *Prasinoderma coloniale* (Pombert et al. 2013). The phylogenetic analysis based on the seven concatenated mitochondrial protein-coding genes strongly supported the sister position of *O. tauri* and *Micromonas* sp. in clade II but positions of other clades of prasinophytes and core chlorophytes are poorly resolved (Turmel et al. 2010). Similarly to chloroplast, also mitochondrial genomes of prasinophytes are very different in size, structure, gene

content and gene order. *Pyc. provasolii* harbours the smallest known mitochondrial genome of this group (24,321 bp and 36 genes) while *N. olivacea* the largest (45,223 bp and 65 genes). The quadripartite structure with two IRs, otherwise rare in mt genomes, was found in four species—*Pr. coloniale*, *O. tauri*, *Micromonas* sp. and *B. prasinus*. *B. prasinus* mitochondrial genome does not map to the circle and is probably linear. All prasinophyte mitochondrial genomes have high gene density, very short intergenic regions and contain almost no introns (Pombert et al. 2013). All so far identified introns are the type I and are localized in the gene for rRNA of the large ribosomal subunit (*rnl*) (*Pr. coloniale*, *N. olivacea*, *Monomastix* sp.OKE-1), in the cytochrome b (*cob*) gene (*N. olivacea*) and in the cytochrome c oxidase I (*cox1*) gene (*Monomastix* sp.OKE-1). The introns in the *rnl* genes of *Pr. coloniale*, *N. olivacea*, *Monomastix* sp.OKE-1 are inserted into homologous positions as introns in mitochondrial *rnl* genes of some other algae (e.g. *Mesostigma viride*) (Turmel et al. 2002), mitochondrial *rnl* genes of other eukaryotes (e.g. *Acanthamoeba castelani*) (Burger et al. 1995), and algal plastid *rnl* genes (e.g. *Oltmannsiellopsis viridis*) (Pombert et al. 2006a). The majority of mitochondrial introns are cis-spliced, however, two *Pr. coloniale* *rnl* introns are trans-spliced (Pombert et al. 2013).

Here, we report the complete mtDNA sequence of the member of prasinophyte clade I—*Pyr. parkeae*. This mitochondrial genome displays unusual quadripartite structure and trans-spliced *cox1* gene.

## MATERIALS AND METHODS

### Preparation of genomic DNA and cDNA

*Pyramimonas parkeae* strain SCCAP K-0007 was obtained from the Scandinavian Culture Collection of Algae and Protozoa and grown in TL30 medium (Larsen et al. 1994). Genomic DNA was extracted from culture (approx.  $20 \times 10^6$  cells) using DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). Total RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA) then mRNA was purified by Dynabeads mRNA Purification Kit (Invitrogen) and cDNA was prepared by SMARTer PCR cDNA Synthesis Kit (Clontech Laboratories, Mountain View, CA).

### Sequencing and assembly of the mitochondrial genome

1 µg of total genomic DNA was subjected to 454 sequencing according to GS FLX Rapid Library Preparation Method protocol (Roche Diagnostics, Mannheim, Germany). In total, 635,427 reads were produced (average length 688). An assembly of the reads was performed with the Roche GS De Novo Assembler software (Newbler, release 2.8, Roche Diagnostics) and resulted in 19,710 contigs (N50 contig size was 1,302 bases). Using a BLAST search (blast.ncbi.nlm.nih.gov/Blast.cgi) it was determined that the second longest contig (con2—27,036 bp) represented about half of the mitochondrial

genome. Further five mitochondrial contigs were identified by BLAST.

Five of the six mitochondrion-derived contigs were manually assembled into a 36,656 bp long linear supercontig that comprises one copy of contig 2 and two copies of other four contigs (con6875; con19412; con1769; con688) in the reverse order at both sides of contig 2. We were unable to join the sixth contig (con58) to the chromosome using bioinformatic approaches. Because we expected that the mitochondrial genome is a circular-mapping molecule, a PCR from the ends of the supercontig and con58 was used to amplify and sequence the missing parts. We also checked the overlaps of adjoining contigs using PCR. The primers used are given in Table S1.

### Annotation

Annotation of ORFs was done using BLASTx search. tRNAs were found using tRNA Scan-SE (Lowe and Eddy 1997), and rRNAs were annotated using BLASTn search. 5S rRNA subunit was searched by program RNAweasel (Lang et al. 2007). Protein coding genes and rRNA genes boundaries were confirmed by aligning with other known algal mitochondrial genes. The annotation was completed in Geneious version 9.0.4, www.geneious.com (Kearse et al. 2012). The boundaries of the single detected intron have been verified by RT-PCR using two pairs of primers—*cox1-dnaF* (5'-TACCTTTCTGACTTGGTGC-3')/*cox1-dnaR* (5'-TTCCTGGATGTGCTAATTCC-3') and *cox1-rnaF* (5'-TAATCTTGGTGCCTTCTCC-3')/*cox1-rnaR* (5'-GTGGATAAAGTCCATCCA-3'). Intron structure was inspected by program RNAweasel (Lang et al. 2007).

The annotated genome is deposited in the EMBL database under accession no. KX756655. The genome maps were plotted in GenomeVx (Conant and Wolfe 2008).

### Phylogenetic analyses

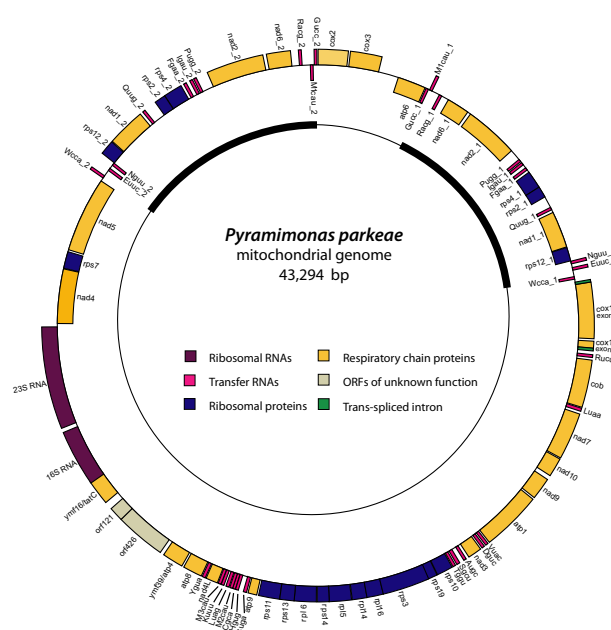
For the phylogenomic analysis we used the data set of 34 protein coding genes from 48 algal mitochondrial genomes available in NCBI database. Amino acid sequences were aligned using MAFFT G-ins-I algorithm (Kato and Standley 2013), trimmed in BMGE (Crisuolo and Gribaldo 2010), manually inspected and concatenated in Sequence-Matrix (Vaidya et al. 2011) into a supermatrix with 6,037 positions. The alignment is available upon request. The phylogenetic tree was constructed in RAxML v7.2.7 (Stamatakis 2006) implemented in CIPRES Science Gateway (www.phylo:sub\_sections/portal/) (Miller et al. 2010) using a uniform PROTGAMMALG4X model and using the partition specific models as determined in Prottest (Abascal et al. 2005). The bootstrap support was calculated using the same setting and 500 permutations. Bayesian analysis was performed using PhyloBayes-MPI v1.7. After the constant sites removal (-dc option), two MCMC chains were run for 3,150 generations using CAT + GTR model. The chain convergence was assessed and consensus tree calculated by bpcomp using burn-in of 500 generations.

## RESULTS

### Structure and gene content of the *Pyramimonas* mtDNA

The *Pyr. parkeae* mitochondrial genome (Fig. 1) assembles into a circle of 43,294 bp in size with overall A + T content of 68.8%, which is so far the highest AT content among prasinophyte mitochondrial genomes. An overview of the main features of *Pyr. parkeae* mitochondrial genome and its comparison with other prasinophyte mitochondrial genomes is given in Table 1. The genome is very compact and conserved genes represent almost 86% of the sequence: 34 protein-coding genes, two genes for rRNA, 25 tRNA genes, and two ORFs (*orf426* and *orf121*) with unknown function. The intergenic regions vary in size from 0 to 542 bp with an average length of 79 pb. There are four cases of overlapping coding DNA sequences (CDS)—*rps19/rps3*, *rps3/rpl16*, *rpl16/rpl14*, *rpl5/rps14* and several cases of closely adjoining genes. Genes are encoded on both DNA strands, and the coding strand switches several times, however, more than one half of the genome (about 22,720 bp) represents region that is continuously encoded by one strand.

The genome contains two inverted repeats, each 6,671 bp long. The quadripartite structure and the presence of IR are not unique in prasinophytes, but unusual is the gene content of the IR segment. All so far identified prasinophyte mitochondrial IRs contain the rRNA operon and the set of various 4–6 protein coding genes. The *Pyr. parkeae* IR contains six protein coding genes (three of them are found in other prasinophyte IRs) but it does



**Figure 1** Map of *Pyramimonas parkeae* mitochondrial genome. Genes are represented by rectangles colour-coded according to their functions. Genes shown on the outside of the circle are transcribed clockwise, whereas those on the inside are transcribed counter-clockwise. Inner ring indicates the position of the inverted repeats (IR).



**Table 1.** General features of prasinophyte and *Pyramimonas parkeae* mitochondrial genomes

Genome feature	<i>Pyramimonas parkeae</i>	<i>Monomastix</i> sp. OKE-1	<i>Micromonas commoda</i>	<i>Ostreococcus tauri</i>	<i>Bathycoccus prasinos</i>	<i>Nephroselmis olivacea</i>	<i>Pycnococcus provasolii</i>	<i>Prasinoderma coloniale</i>
Clade	I	II	II	II	II	III	V	VI
Size (bp)	43,294	60,883	47,425	44,237	43,614	45,223	24,321	54,546
A + T (%)	68.8	59.9	65.4	61.8	59.9	67.2	62.2	54.2
Gene content <sup>a</sup>	61	65	64	65	61	65	36	55
Protein coding genes	34	35	33	35	34	35	18	29
tRNA	25	26	27	26	23	26	16	24
rRNA <sup>b</sup>	2	3	3	3	3	3	2	2
RNase P RNA <sup>b</sup>	0	1	1	1	1	1	0	0
Coding seq. (%) <sup>c</sup>	85.6	71.2	82.5	92.1	88.2	76.6	87.7	71.0
Intergenic spacer (avg. size in bp)	79	159	105	42	73	125	83	194
Introns: number/type/splicing <sup>b</sup>	1/II <sup>d</sup> /trans	8/I/cis	0	0	0	4/I/cis	0	2/I/trans
Inverted repeat (bp)	6,671	–	12,234	9,771	9,731	–	–	14,364
Preferred stop codon	TAA (61%)	TAA (69%)	TAA (87%)	TAA (81%)	TAA (73%)	TAA (77%)	TGA (85%)	TGA (88%)

<sup>a</sup>Total number of conserved genes (does not include duplicates and ORFs).

<sup>b</sup>The presence of 5S RNA, RNase P RNA and introns was detected by RNAweasel (Lang et al. 2007).

<sup>c</sup>Genome fraction that includes conserved genes and ORFs.

<sup>d</sup>Derived intron, probably intron of type II.

not harbour the rRNA operon (Table 2). Contrary with some prasinophytes the rRNA genes do not contain any interruptions.

Thirty-four protein coding genes comprise genes with function in respiratory chain (complex I—*nad1*, 2, 3, 4, 4L, 5, 6, 7, 9, 10; complex III—*cob*; complex IV—*cox1*, 2, 3; complex V—*atp1*, 6, 8, 9 and *atp4* referred also as *yfm39*), genes for protein components of small and large ribosomal subunits (*rps2*, 3, 4, 7, 10, 11, 12, 13, 14, 19 and *rpl5*, 6, 14, 16) and one gene designated alternatively *TatC*, *yfm16* or *MttB* for Sec-independent protein translocase protein. Only one gene (*cox1*) contains an intron. All gene annotations were confirmed by alignments of their homologs. Only the *rps2* gene does not correspond in length with its homologs and it is either a pseudogene or a part of this gene remains unrecognized.

The best preserved gene cluster is the *Escherichia coli*-like *S10/spc/alpha* cluster that is shared also with most prasinophytes. In comparison with *N. olivacea*, only one

gene (*rps8*) is missing and the cluster consists of 10 genes (*rps10*, *rps19*, *rps3*, *rpl16*, *rpl14*, *rpl5*, *rps14*, *rpl6*, *rps13*, *rps11*). The region is very compact and contains all four cases of overlapping genes and three tightly adjacent genes. Other evolutionarily conserved gene clusters or gene pairs present in this genome are *nad4L/Y(gua)/atp8/atp4*, *cox2/cox3* and *nad10/nad9*.

The set of 25 identified tRNAs is almost identical to the set present in *N. olivacea* mt genome, with only one tRNA (arginine tRNA, codon ucg) absent in *Pyr. parkeae*. *Pyr. parkeae* mitochondrion uses the standard genetic code. Almost all protein coding genes begin with a start codon ATG, the exception is the *atp4* gene that uses alternative start codon TTG. The most common stop codon is TTA. TTA is preferred stop codon also in all known mitochondrial genomes of prasinophytes from clades II and III whilst in clades V and VI the codon TGA prevails. The most abundant amino acid is leucine (13% of used codons), a situation very similar to other prasinophytes.

We have identified about 230 repeat sequences longer than 15 bp (the longest one of 88 bp). Most of them are palindromic repeats often arranged in tandems. Around 80 of them are located in each IR part of mt genome. Two big clusters of tandems of palindromic repeats are found—in the 97 bp long cluster within IR and in the 114 bp long cluster before the *tatC* gene. Both clusters are able to form four stems with loops.

### Trans-spliced intron in COX1

The mitochondrial genome of *Pyr. parkeae* contains only one intron that is located in the *cox1* gene coding for

**Table 2.** Gene content of inverted repeats in prasinophyte mitochondrial genomes

Organism	IR gene content
<i>Pyramimonas parkeae</i>	<b>rps12-nad1-rps2-rps4-nad2-nad6</b>
<i>Micromonas</i> sp.	atp8-atp4-cox1-rns-rnl- <b>nad1</b> -atp6- <b>nad6</b>
<i>Ostreococcus tauri</i>	nad4L-atp8-atp4-cox1-rns-rnl-cob
<i>Bathycoccus prasinos</i>	atp8-atp4-cox1-rns-rnl-cob
<i>Prasinoderma coloniale</i>	<b>nad1</b> -atp6-rnl(1)-rns-rnl(2) -cob-rnl(3)- <b>rps12</b> -nad9-nad3

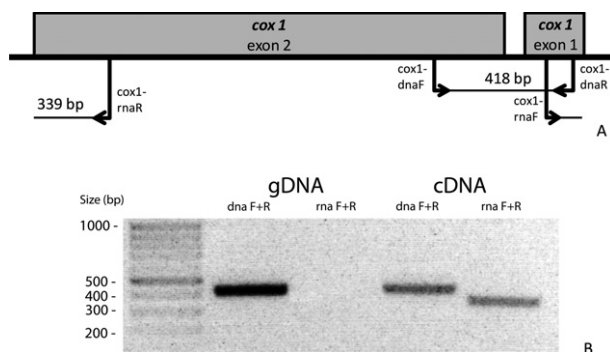
The genes that *Pyr. parkeae* shares with other prasinophytes are in bold.



subunit 1 of cytochrome oxidase. The two exons of *cox1* (exon 1—178 bp long and exon 2—1,421 bp long) lie on the same DNA strand but in the reverse order, i.e. exon 2 before exon 1. To produce the functional protein, the mRNA has to be trans-spliced. We have analysed the presence of a mature mRNA by performing RT-PCR with *Pyramimonas* total RNA as a template and primers (*cox1*-rna-F/R) specific for the neighbouring regions of exon 1 and 2 in the putative mature mRNA. The expected 339 bp long fragment was obtained, while the same primers with genomic DNA as a template did not produce any fragment. Amplification of DNA and cDNA with primers (*cox1*-dna-F/R) specific for the adjacent part of exon 2 and exon 1 on the DNA resulted in 418 bp fragments that correspond to the expected length of this genomic region and the region on the non-spliced pre-mRNA, respectively (Fig. 2). All these fragments were sequenced to precisely determine the intron boundaries.

The 5' end of the exon 1 bears sequences 5'-UCAUUUAU—3' and 5'-AUAUAG—3' which correspond to sequences of intron-binding site 2 and 1 (IBS2 and IBS1), respectively, in the *cox1* gene of *Marchantia polymorpha* (intron ai2) and *Saccharomyces cerevisiae* (intron ai1) (Ohta et al. 1993). The 5' exon-intron boundary bears sequence 5'-↓UUGUG-3' that partially resembles the splicing site consensus sequence 5'-↓GUGYG-3' of type II introns (Fig. 3A,B).

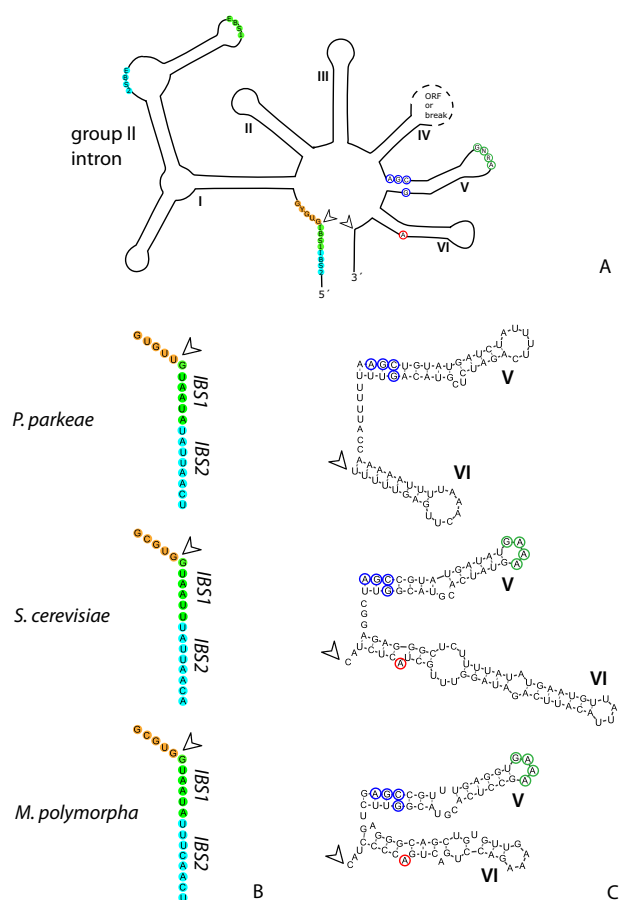
The 3' exon-intron boundary bears the sequence 5'-AGUUUUU↓-3' that only distantly resembles the 3' end of type II intron with adenosine in position 7 before splicing site but A seems not to be typically bulged. A 36 bp long sequence starting 66 nt upstream of the 3' exon-intron boundary was recognized as derived domain V of type II intron by the program RNAweasel (Lang et al. 2007).



**Figure 2** Analysis of trans-spliced intron in the *cox1* gene of *Pyramimonas parkeae*. **(A)** The diagram displays relative position of two exons of *cox1* and binding sites of two primer pairs designed to amplify fragments from gDNA and mature mRNA. **(B)** Confirmation of intron trans-splicing by RT-PCR analysis. Electrophoretic analysis of PCR fragments amplified from gDNA and cDNA, respectively, with both primer pairs. The amplicon in the first lane corresponds to the expected size (418 bp) on gDNA. The same size has the amplicon in the third lane that probably represents non-spliced pre-mRNA. The length of the amplicon in the fourth lane corresponds to the spliced mRNA (339 bp). The results were confirmed by sequencing.

Typical domain V forms a stem-loop that often carries a GNRA-tetraloop motif (N—any nucleotide, R—A or G), frequently GAAA or GUGA (Lang et al. 2007). The domain has no such motif but contains a catalytic triad AGC, which binds catalytically important  $Mg^{++}$  ions (Lambowitz and Zimmerly 2011) (Fig. 3B).

No other domains of type II introns were identified. Many type II introns encode open reading frames (ORFs) that allow their splicing and retrotransposing (Burger and Nedelcu 2012). For type I introns the typical ORF is the one encoding a homing endonuclease with LAGLIDADG motif. We did not find any such ORF located near the intron sequences or anywhere else in the mitochondrial genome. Based on these indices we propose that this trans-spliced intron is an extremely derived type II intron. We were not able to identify the complete structure of the intron nor its splicing machinery.



**Figure 3** Comparison of group II intron domains of *Pyramimonas parkeae*, *Saccharomyces cerevisiae* and *Marchantia polymorpha*. **(A)** Schematic representation of the secondary structure of mitochondrial group II introns. **(B)** Comparison of domain I consensus sequences: intron binding sites (IBS1, IBS2) and GUGYG motif, which is modified in the case of *Pyr. parkeae*. **(C)** Comparison of domains V and VI consensus sequences: *Pyr. parkeae* does not bear bulged A nor GNRA motif. Catalytic site AGC is present. Splicing sites are indicated by arrows.

### Phylogenetic analyses

We used 34 genes encoded in the mitochondrial genome of *Pyr. parkeae* for multigene alignment to reconstruct its phylogeny, in which we have included representatives of five prasinophyte clades (I–II, V and VI), Trebouxiophyceae, Ulvophyceae, Chlorophyceae, *Pedinomonas* and representatives of Streptophyceae as an outgroup (48 taxa in total). Phylogenetic tree is given in Fig. 4. In all tree reconstructions, *Pyramimonas* forms a sister lineage of prasinophyte clade II with bootstrap support (BS) 90% and posterior probability (PP) 0.98. In the maximum likelihood tree constructed by RAxML, the groups Trebouxiophyceae, Ulvophyceae, Chlorophyceae and *Pedinomonas* do not form a clade, as they are interrupted by prasinophytes *Prasinoderma* and *Pycnococcus* (BS < 50%). However, this clade was recovered by Bayesian analysis in Phylobayes (PP 0.91). The lineages of Chlorophyceae and *Pedinomonas* form a clade in all reconstructed trees (BS 99% and 89% and PP 0.95) and the branch lengths of this clade are always markedly increased, which may affect the tree topology.

### Tracing of mt genes in Euglenophytes

*Pyramimonas parkeae* is considered the closest relative to the plastid endosymbiont of Euglenophytes. We have searched for the homologues of *Pyr. parkeae* mitochondrial genes in the published and our in-house sequenced unpublished transcriptomes of *E. gymnastica* and *E. gracilis* in order to identify potential endosymbiotic gene transfers from the mitochondrial genome of the plastid donor into the Euglenophyceae host lineage. No close homologues were found, so we have no evidence of such a transfer.

### DISCUSSION

The mitochondrial genome of *Pyr. parkeae*, the first reported genome from group I prasinophyte clade, generally resembles the mitochondrial genomes of other prasinophytes, particularly those from clades II and III.

The *Pyramimonas* mt genome and mt genomes of other prasinophytes are very compact and their sizes are very similar (around 45 kbp and 55–65 genes) with the exception of the reduced mitochondrial genome of picoplanktonic *Pyc. provasolii* (24.3 kbp and 36 genes). Mitochondrial genomes of *Monomastix* sp. *OKE-1* and *Pr. coloniale* are slightly expanded due to longer intergenic regions and several introns. Unlike the most gene-rich genomes of *N. olivacea* and *O. tauri* (65 genes), the *Pyr. parkeae* genome lacks only four conserved genes coding for *rps8*, *tRNA(ucg)* for arginine, *RNase P RNA*, and *5S rRNA* (see table 6.3. in Burger and Nedelcu 2012). *Rps8* as a part of cluster of ribosomal genes is present in almost all prasinophytes studied so far but is frequently absent in other chlorophytes and basal streptophytes. It has not been found in the mitochondria of any angiosperm where its function is probably substituted by the cytosolic

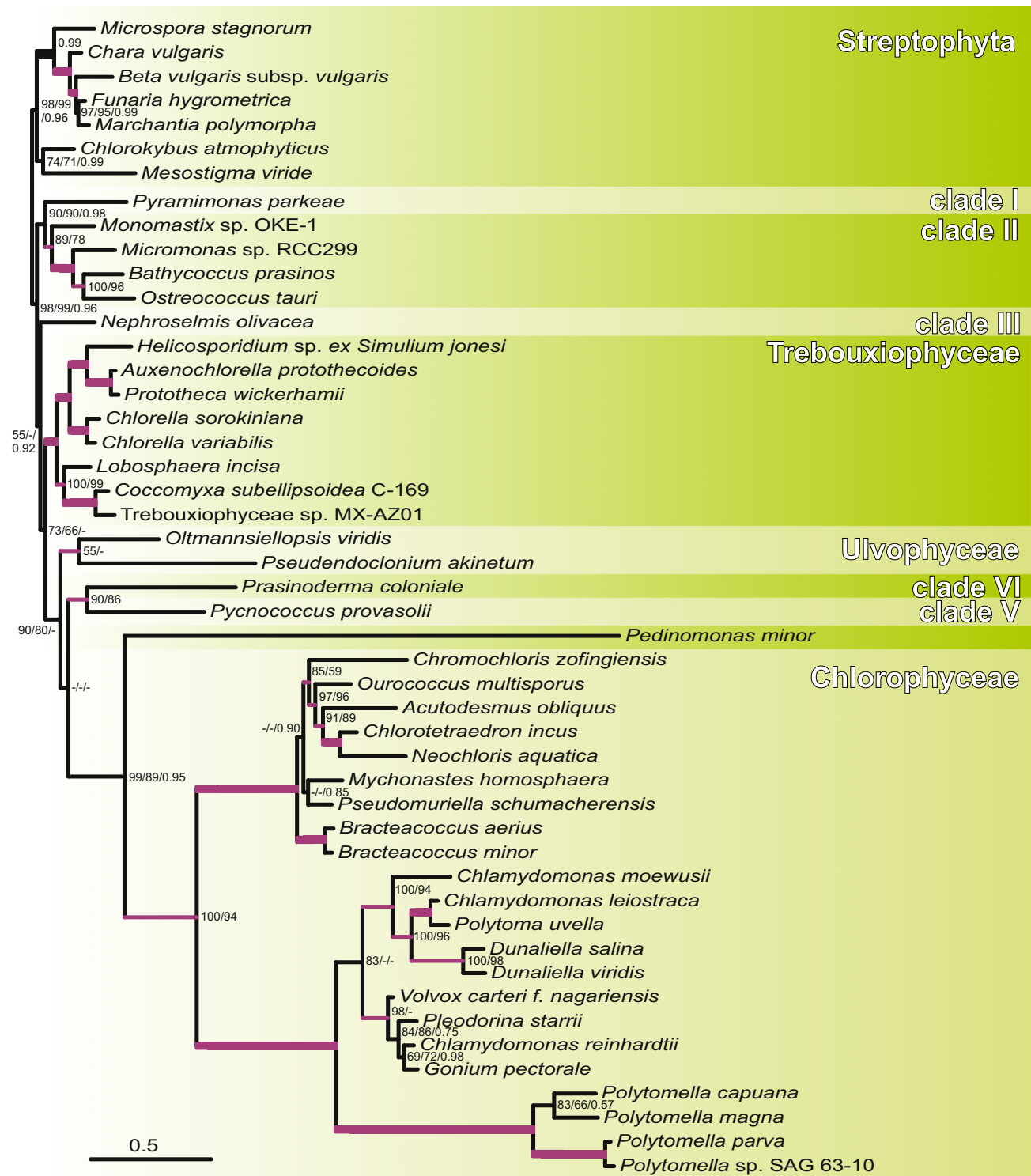
*rps15A* (Adams et al. 2002). The chance that the gene *rps8* remained unrecognized in the *Pyr. parkeae* genome is low due to the absence of any space for this gene in the cluster of ribosomal genes.

*Pyramimonas parkeae* also lacks *5S RNA* and *RNase P RNA* that has been found in mt genomes of prasinophyte clades II and III (Burger and Nedelcu 2012). The gene for *RNase P RNA* is considered to be ancestrally present in mitochondrial genomes as it closely resembles *RNase P RNA* of the heterotrophic jakobids, a group in which the mitochondrial genomes have many ancestral features (Burger et al. 2013). From this point of view, *Pyr. parkeae* seems to be more derived lineage than *Nephroselmis* and *Ostreococcus*, which still contain this gene. On the other hand, the absence of arginine *tRNA(ucg)* is not surprising because this tRNA was probably lost early after the acquisition of mitochondrion as it is not present in the mitochondrial genome of *Reclinomonas americana* and in many other mt DNAs. *Nephroselmis olivacea* and three streptophytes (*Mesostigma viride*, *Chlorokybus atmophyticus* and *Marchantia polymorpha*) independently gained this tRNA again by gene duplication and modification of other tRNAs (Wang et al. 2009).

The circular-mapping mitochondrial chromosome of *Pyr. parkeae* shows quadripartite structure with two inverted repeats (IR) and two single copy regions. This structure is present in some other mitochondrial genomes of prasinophytes but absent in higher plants (Burger and Nedelcu 2012). The presence of mitochondrial IR varies between prasinophyte species and does not correspond to their phylogenetic relationships nor to the presence of chloroplast IR. IRs in mitochondria are very variable in size (6.7–12.5 kbp) and have somewhat similar gene content across lineages, although the gene order varies. They typically encode the rRNA operon like IR regions in chloroplast genomes. The exception represents the partially sequenced mtDNA of *Tetraselmis subcordiformis*, which apparently contains 1.5 kbp IRs without any gene at all (Kessler and Zetsche 1995). The gene content of the *Pyr. parkeae* IR is to some extent similar to others but it completely lacks the rRNA operon.

This unusual organization of IRs is a result of rearrangements in the mt genome, which may be related to high number of short palindromic repeats that are assumed to facilitate such rearrangements (Nedelcu and Lee 1998). As rearrangement can affect also the individual exons it seems possible that short repeats could play a role in scrambling genes (Smith and Lee 2009). The presence of such repeats was recently revealed in the mt genome of a protist *Diphyllia rotans* that contains four genes with scattered exons. Trans-splicing was confirmed in three of them (Kamikawa et al. 2016).

Prasinophyte mt genomes harbour no or few introns that are all of the type I. The type II introns are the predominant type in the land plant mt genomes (Lang et al. 2007) but many mitochondrial genomes of land plants bear both types of introns even in one gene, e.g. in liverwort or yeast *cox1* gene (Ohta et al. 1993). Most of these genes are transcribed as one transcriptional unit and are



**Figure 4** Phylogenomic analysis based on 34 proteins encoded in the mitochondrial genome. The tree topology was constructed from the alignment containing 6,037 amino acid positions using maximum likelihood implemented in RAXML with uniform PROTGAMMALG4X model of substitution. The numbers on nodes represent bootstrap support calculated using the uniform model/partition specific model/PhyloBayes posterior probabilities. Thickened branches indicate nodes with 100% bootstrap support in both models, magenta branches indicate nodes with PhyloBayes posterior probabilities 1. Bootstraps < 50% and posterior probabilities < 0.5 are indicated by -. The tree is arbitrarily rooted by Streptophyta.

cis-spliced. When the mt genome undergoes rearrangement, the exons with flanking half-intron may become separated and transcribed independently. The mature mRNA is then formed by trans-splicing, during which the scrambled intron fragments fold based on secondary and tertiary structures (Bonen 2008).

*Pyramimonas parkeae* mt genome contains one trans-spliced intron in the gene for cytochrome oxidase I. We were not able to establish with confidence the intron type but based on the presence of several sequence motifs we hypothesize that it is a highly reduced type II intron. Group II introns are ribozymes spliced by two transesterification steps with the presence of lariat form and have a typical secondary structure with six conserved domains I–VI. Several conserved sequencing motifs were identified in the typical group II introns: splicing site 5'- $\downarrow$ GUGYG—3' (domain I), GNRA loop and catalytic site AGC (domain V), bulged A involved in the first transesterification step of splicing and splicing site 5'-AY $\downarrow$ —3' (domain VI) (Bonen and Vogel 2001; Lang et al. 2007). Pairing between exon binding sites (EBS1,2) presented in intron and intron binding sites (IBS1,2) on the 3' end of the foregoing exon forms the tertiary structure guiding exon ligation (Costa et al. 2000). An ORF with a reverse transcriptase activity is present in some type II introns and is located typically within domain IV. In this domain one may localize the break point of trans-spliced introns (Bonen and Vogel 2001) (Fig. 3). Rarely the trans-spliced intron is tripartite with the second break point within domain I (Goldschmidt-Clermont et al. 1991; Knoop et al. 1997).

We identified only short motif of domain I and derived domains V and VI in the intron of *cox1* gene *Pyr. parkeae*. Although the intron is theoretically too short to form the typical secondary structure with splicing activity and we were unable to detect the putative missing part of the intron elsewhere in the mitochondrial genome, the amplification of mRNA shows that the gene indeed undergoes trans-splicing as predicted.

The comparison of *cox1* genes carrying introns shows that there are several organisms harbouring group IIA intron at the same site as *Pyr. parkeae*, although these introns are not closely related to each other neither in primary sequences nor in their secondary structure (Ohta et al. 1993). One such intron is intron 1 in *cox1* of *S. cerevisiae*. This intron is closely related to other yeast group IIA introns and bears an ORF potentially encoding a protein related to reverse transcriptases (Michel and Ferat 1995). As introns are mobile genetic elements, their distribution between yeast lineages is not conserved and this particular intron is present in *cox1* of some strains of *S. cerevisiae* (e.g. S288c) but not in others (e.g. isolate NCYC3594). Other organisms with introns in homologous position to the one in *Pyr. parkeae* are the liverworts *Marchantia polymorpha*, *Aneura pinguis*, *Pellia epiphylla*, *Treubia lacunosa* and *Pleurozia purpurea*, belonging into four different subclasses of Marchantiophyta. In these organisms, the intron is always the second one in the *cox1* gene, it belongs also to group IIA and contains ORF potentially encoding a protein related to reverse

transcriptases, but unrelated to the ORF in *cox1* intron of *S. cerevisiae* (Ohta et al. 1993). We did not identify a corresponding ORF with reverse transcriptase activity in the mitochondrial genome of *Pyr. parkeae*.

The phylogeny of mitochondrial genomes is generally considered to correspond to the species tree. Our phylogenetic analyses based on 34 conserved genes encoded in mitochondrial genomes of green algae and plants is, indeed, in most respects concordant with the phylogenies based on 18S rRNA or chloroplast genes. *Pyr. parkeae*, the only representative of prasinophyte clade I, branches (with bootstrap support 90%) as sister to Mamiellales (prasinophyte clade II). Unlike the phylogeny based on chloroplast genes, Trebouxiophyceae, Ulvophyceae, Chlorophyceae and *Pedinominas* do not form a clade in the maximum likelihood tree, as they are interrupted by genera *Prasinoderma* and *Pycnococcus*. However, the node interrupting this monophyly is weakly supported and this topology was not recovered by Bayesian analysis. Unlike our tree, *Prasinoderma* branches robustly as the most basal prasinophyte lineage in the phylogenetic tree reconstructed on the basis of chloroplast encoded genes (Lemieux et al. 2014). The topologies of our trees are, however, very likely affected by the highly increased branch lengths of *Pedinomonas* and *Chlorophyceae*, which exceed the lengths of other branches approximately three times. This phenomenon has been observed already in the phylogenetic analysis of Turmel et al. (2010) on a smaller set of taxa. Reason for the increased substitution rates in these lineages is not known to us, but the presence of such uneven branch lengths may bias the topology and taxa with long branches (including *Prasinoderma* and *Pycnococcus*) may be grouped artificially as the result of long-branch attraction artefact.

## CONCLUSIONS

Mitochondrial genome of *Pyr. parkeae* displays some unique features among prasinophytes. It possesses IRs like some members of prasinophyte clades II and VI but the gene content of these IRs is quite different and unlike all others it does not contain an rRNA operon. The single intron in the *cox1* gene is probably a highly reduced type II intron and it is trans-spliced at the level of mRNA as confirmed by RT-PCR. Further study on the splicing mechanism of such miniature self-splicing intron may reveal interesting findings.

The phylogenetic analyses based on the mitochondrion-encoded genes support the relationship of *Pyr. parkeae* to Mamiellales. Despite the few peculiarities, *Pyr. parkeae* mitochondrial genome forms one of the shortest branches in the tree, has a quadripartite structure, and contains large repertoire of densely packed genes, which overall suggests that it may resemble the ancestral mt genome of green algae. As the same holds for the *Pyr. parkeae* chloroplast genome and 18S rRNA, Pyramimonadales likely represent genetically one of the least derived lineage of Chlorophyta and could be used as such in evolutionary considerations.



## ACKNOWLEDGMENTS

This work was supported by the Czech Science Foundation P506/11/1320 (to V.H.), the Grant Agency of the Charles University 604313 (to S.H.) and by the Ministry of Education, Youth and Sports of CR within the National Sustainability Program II (Project BIOCEV-FAR) LQ1604 and by the project "BIOCEV" (CZ.1.05/1.1.00/02.0109).

## LITERATURE CITED

- Abascal, F., Zardoya, R. & Posada, D. 2005. ProtTest manual. *Bioinformatics*. doi:10.1093/bioinformatics/bti263.
- Adams, K. L., Daley, D. O., Whelan, J. & Palmer, J. D. 2002. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell*, 14(April):931–943.
- van Baren, M. J., Bachy, C., Reistetter, E. N., Purvine, S. O., Grimwood, J., Sudek, S., Yu, H., Poirier, C., Deerinck, T. J., Kuo, A., Grigoriev, I. V., Wong, C.-H., Smith, R. D., Callister, S. J., Wei, C.-L., Schmutz, J. & Worden, A. Z. 2016. Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genom.*, 17(1):267. doi:10.1186/s12864-016-2585-6.
- Bonen, L. 2008. Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion*, 8(1):26–34.
- Bonen, L. & Vogel, J. 2001. The ins and outs of group II introns. *Trends Genet.*, 17(6):322–331.
- Burger, G., Gray, M. W., Forget, L. & Lang, B. F. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol. Evol.*, 5(2):418–438.
- Burger, G. & Nedelcu, A. M. 2012. Mitochondrial genomes of Algae. 2012. In: Bock, R. & Knoop, V. (ed.), *Genomics of Chloroplasts and Mitochondria*. Springer, Dordrecht, Netherlands. 6:127–157.
- Burger, G., Plante, I., Lonergan, K. M. & Gray, M. W. 1995. The mitochondrial DNA of the amoeboid protozoan, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. *J. Mol. Biol.*, 245:522–537.
- Conant, G. C. & Wolfe, K. H. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*, 24(6):861–862.
- Costa, M., Michel, F. & Westhof, E. 2000. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.*, 19(18):5007–5018.
- Criscuolo, A. & Gribaldo, S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.*, 10(1):210. doi:10.1186/1471-2148-10-210.
- Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M. & Rochaix, J. D. 1991. A small chloroplast RNA may be required for trans-splicing in *Chlamydomonas reinhardtii*. *Cell*, 65:135–143.
- Guillou, L., Eikrem, W., Chrétiennot-Dinet, M.-J., Le Gall, F., Masana, R., Romari, K., Pedrós-Alió, C. & Vaulot, D. 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist*, 155(June):193–214.
- Hrdá, Š., Fousek, J., Szabová, J., Hampl, V. & Vlček, Č. 2012. The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS ONE*, 7(3). doi:10.1371/journal.pone.0033746.
- Jansen, R. K. & Ruhlman, T. A. 2012. Plastid genomes of seed plants. In: Bock, R. & Knoop, V. (ed.), *Genomics of Chloroplasts and Mitochondria*. Springer, Dordrecht, Netherlands. 5:103–126.
- Kamikawa, R., Shiratori, T., Ishida, K. I., Miyashita, H. & Roger, A. J. 2016. Group II intron-mediated trans-splicing in the gene-rich mitochondrial genome of an enigmatic eukaryote, *Diphyllia rotans*. *Genome Biol. Evol.*, 8(2):458–466.
- Katoh, K. & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30(4):772–780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649.
- Kessler, U. & Zetsche, K. 1995. Physical map and gene organization of the mitochondrial genome from the unicellular green alga *Platymonas (Tetraselmis) subcordiformis* (Prasinophyceae). *Plant Mol. Biol.*, 29(5):1081–1086.
- Knoop, V., Altwasser, M. & Brennicke, A. 1997. A tripartite group II intron in mitochondria of an angiosperm plant. *Mol. Gen. Genet.*, 255(3):269–276.
- Lambowitz, A. M. & Zimmerly, S. 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.*, 3(8):1–19.
- Lang, B. F., Laforest, M. J. & Burger, G. 2007. Mitochondrial introns: a critical view. *Trends Genet.*, 23(3):119–125.
- Larsen, N. H., Moestrup, Ø. & Pedersen, P. M. 1994. Catalogue 1994. Scandinavian Culture Centre for Algae & Protozoa. Department of Phycology. Botanical Institute. University of Copenhagen. <http://www.sccap.dk/media/marine/1.asp>
- Lemieux, C., Otis, C. & Turmel, M. 2014. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genom.*, 15:857. doi:10.1186/1471-2164-15-857.
- Lowe, T. M. & Eddy, S. R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, 25(5):955–964.
- Michel, F. & Ferat, J. L. 1995. Structure and activities of group II introns. *Annu. Rev. Biochem.*, 64:435–461.
- Miller, M. A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop, GCE 2010. doi:10.1109/GCE.2010.5676129
- Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F., Piganeau, G., Rouzé, P., Da Silva, C., Wincker, P., Van de Peer, Y. & Vandepoele, K. 2012. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.*, 13(8), R74. doi:10.1186/gb-2012-13-8-r74.
- Nedelcu, A. M. & Lee, R. W. 1998. Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. *Mol. Biol. Evol.*, 15(6):690–701.
- Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohya, K. & Michel, F. 1993. Group-I introns in the liverwort mitochondrial genome – the gene coding for subunit-1 of cytochrome-oxidase shares 5 intron positions with its fungal counterparts. *Nucleic Acids Res.*, 21(5):1297–1305.
- Pombert, J. F., Beauchamp, P., Otis, C., Lemieux, C. & Turmel, M. 2006a. The complete mitochondrial DNA sequence of the

- green alga *Oltmannsiellopsis viridis*: evolutionary trends of the mitochondrial genome in the Ulvophyceae. *Curr. Genet.*, 50(2). doi:10.1007/s00294-006-0076-z.
- Pombert, J. F., Lemieux, C. & Turmel, M. 2006b. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol.*, 4 (3). doi:10.1186/1741-7007-4-3.
- Pombert, J. F., Otis, C., Turmel, M. & Lemieux, C. 2013. The mitochondrial genome of the prasinophyte *Prasinoderma coloniale* reveals two trans-spliced group I introns in the large subunit rRNA gene. *PLoS ONE*, 8(12). doi:10.1371/journal.pone.0084325.
- Robbens, S., Derelle, E., Ferraz, C., Wuyts, J., Moreau, H. & Van De Peer, Y. 2007. The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction. *Mol. Biol. Evol.*, 24(4):956–968.
- Smith, D. R. & Lee, R. W. 2009. The mitochondrial and plastid genomes of *Volvox carteri*: bloated molecules rich in repetitive DNA. *BMC Genom.*, 10:132. doi:10.1186/1471-2164-10-132.
- Stamatakis, A. 2006. Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690.
- Suda, S., Bhuiyan, M. A. H. & Faria, D. G. 2013. Genetic diversity of *Pyramimonas* from Ryukyu Archipelago, Japan (Chlorophyceae, Pyramimonadales). *J. Mar. Sci. Technol.*, 21 (Suppl.):285–296.
- Turmel, M. 1999. The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: two radically different evolutionary patterns within green algae. *Plant Cell*, 11 (9):1717–1730.
- Turmel, M., Gagnon, M. C., O’Kelly, C. J., Otis, C. & Lemieux, C. 2009. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.*, 26(3):631–648.
- Turmel, M., Otis, C. & Lemieux, C. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proc. Natl Acad. Sci. USA*, 96(18):10248–10253.
- Turmel, M., Otis, C. & Lemieux, C. 2002. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol. Biol. Evol.*, 19(1):24–38.
- Turmel, M., Otis, C. & Lemieux, C. 2010. A deviant genetic code in the reduced mitochondrial genome of the picoplanktonic green alga *Pycnococcus provasolii*. *J. Mol. Evol.*, 70(2):203–214.
- Turmel, M., Otis, C. & Lemieux, C. 2013. Tracing the evolution of streptophyte algae and their mitochondrial genome. *Genome Biol. Evol.*, 5(10):1817–1835.
- Vaidya, G., Lohman, D. J. & Meier, R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, 27(2):171–180.
- Viprey, M., Guillou, L., Ferréol, M. & Vaulot, D. 2008. Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ. Microbiol.*, 10(7):1804–1822.
- Wang, B., Xue, J., Li, L., Liu, Y. & Qiu, Y. L. 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Curr. Genet.*, 55(6):601–609.
- Worden, A. Z., Lee, J. H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., Allen, A. E., Cuvelier, M. L., Derelle, E., Everett, M. V., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S. M., Parker, M. S., Rombauts, S., Salamov, A., Von Dassow, P., Badger, J. H., Coutinho, P. M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J. E., John, U., Lanier, W., Lindquist, E., Lucas, S., Mayer, K. F. X., Moreau, H., Not, F., Otiillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E. V., Bhattacharya, D., Goodenough, U. W., Van de Peer, Y. & Grigoriev, I. V. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, 324(5924):268–272.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

**Table S1.** Primers used for confirmation of contigs joining.

## 6 Endosymbiotický genový přenos

Přesun DNA z jádra endosymbionta do jádra hostitele je významný způsob obohacení organismu o nové geny a je proto důležitým procesem eukaryotické evoluce. Část takto přesunutých genů kóduje proteiny, které se vrací zpět do příslušné organely, část ale kóduje proteiny, které jsou nově směřovány do jiných kompartmentů buňky, kde mohou nabývat nových funkcí. Fragменты organelové DNA vložené do jaderného genomu mohou také modifikovat expresi jiných genů. Má se za to, že největší podíl genů byl přenesen záhy po získání organel, ale v malé míře přesun probíhá neustále (Lloyd et al., 2012; Stegemann et al., 2003; Thorsness & Fox, 1990). Předpokládá se, že v průběhu přenosu se kopie genu dostává do jádra a po čase organelová kopie mizí (Jiroutová et al., 2010). Rozsah EGT, tedy rozlišení genů pocházejících od endosymbionta a hostitele, je obtížné stanovit především kvůli různorodým souběžným evolučním procesům, ale je velice důležitý pro určení fylogenetického původu hostitele i endosymbionta. Práce zabývající se rozsahem EGT zkoumají buď podíl endosymbiozou získaných genů v jaderném genomu a celkovém transkriptomu hostitele, nebo se zaměřují na zjišťování tohoto podílu pouze v transkriptomu nebo proteomu plastidu. Kromě celých funkčních genů se během EGT přesouvaly z organel do jádra i neúplné nefunkční fragmenty DNA ("Nuclear plastid DNA" (NUPT) a "Nuclear mitochondrial DNA" (NUMT)), kterých může být překvapivě mnoho (Makai et al., 2015).

### 6.1 EGT při primární endosymbióze

Problém odlišení genů pocházejících z případného archezoálního nebo archebakteriálního předka eukaryotů a genů pocházejících z eubakteriálního předka mitochondrie je především v tom, že dosud nebyl nalezen eukaryotní archezoální organismus bez mitochondrie nebo velmi blízká linie archebakterií, ke kterému by bylo možné vztahovat změny v zastoupení genů. Předpokládá se, že z původního genomu endosymbionta bylo cca 1000 – 3000 genů zcela ztraceno, malé procento genů zůstalo v mitochondriálním genomu, a minimálně 800 – 1700 genů, tedy 20 – 30 %, bylo přeneseno do jádra hostitelské buňky (Gray & Archibald, 2012). Ovšem pouze část těchto genů exprimuje proteiny, které se vrací zpět do mitochondrie a nalézáme je mitochondriálním proteomu. Nemalá část exprimovaných produktů je přesměrována a funguje v jiných buněčných oddílech (Gabaldón & Huynen, 2003; Karlberg et al., 2000). Odhaduje se, že proteom mitochondrie je velký několik set až několik málo tisíc proteinů (Gonczarowska-Jorge et al., 2017; Lee et al., 2013; Peikert et al., 2017; Perez et al., 2014), ovšem jen menšina těchto proteinů včetně těch kódovaných v mitochondriálním genomu má jednoznačně  $\alpha$ -proteobakteriální původ (okolo 10 – 15%) (Gabaldón & Huynen, 2003; Gray & Archibald, 2012). Až polovina proteinů má blíže neurčitelný prokaryotický původ, asi třetinový podíl mají proteiny neprokaryotického původu, tedy podobné eukaryotickým proteinům, a nezanedbatelná část genů je specifická jen pro určité linie organismů a nemá homologii ani u prokaryot, ani u eukaryot (Brueckner & Martin, 2020; Esser et al., 2004; Gray & Archibald, 2012; Karlberg et al., 2000; Marcotte et al., 2000).

Podobná situace jako u mitochondrií byla pozorována i u plastidů. Plastidový proteom také zahrnuje několik tisíc proteinů a tyto proteiny musí být ve své většině kódovány v jádře hostitele, protože genomy plastidů kódují jen několik desítek až několik málo stovek proteinů, což odpovídá jen asi 5 % genů volně žijících sinic (Green, 2011). Předpokládalo se, že geny přenesené z endosymbionta do jaderného genomu hostitele a směřující své produkty zpět do plastidu mají sinicový původ. I zde se ukázalo, že takových genů je menšina. V proteomu několika archaeplastid tvořily původem sinicové proteiny kódované v plastidu 5 – 25 % a původně sinicové proteiny kódované v jádře dalších 12 – 24 %, dohromady však pouhých 25 – 37 %. Ostatní plastidové proteiny kódované v jádře jsou odvozeny od původních proteinů hostitele, které převzaly funkci potřebnou pro fungování plastidu, přičemž odpovídající sinicové geny se ztratily. V proteomech bylo těchto genů dokonce více než polovina, 56 – 65 % (Qiu et al., 2013). Překvapivě je v proteomech zastoupeno nezanedbatelné množství proteinů majících homologii k jiným prokaryotům než sinicím, které zřejmě byly získány horizontálním genovým přenosem (HGT) a kterých je 7 – 15 % (Qiu et al., 2013). Další podobné práce zkoumající původ proteinů v plastidovém proteomu u vyšších rostlin nebo jiných řas došly k rozdílným procentům zastoupení těch původem sinicových, ale tento podíl se vždy pohybuje jen v několika desítkách procent (Martin et al., 2002; Reyes-Prieto et al., 2007).

Při zkoumání proteomu a transkriptomu *Paulinella chromatophora* bylo nalezeno více než 450 proteinů kódovaných jadernými geny hostitele, které jsou směřovány do cyanely. Překvapivě pouze 4 % těchto genů pochází z EGT, dalších 6 % genů pochází z HGT a 90 % genů je tak hostitelského nebo neznámého původu (Singer et al., 2017). Geny, které *Paulinella* získala díky HGT, pocházejí od různých jiných linií bakterií a předpokládá se, že jejich produkty doplňují mezery v biochemických drahách cyanely, které vznikly v důsledku ztráty vlastních genů při endosymbiotickém procesu (Nowack et al., 2016).

Celkově lze shrnout, že výsledkem EGT při primární endosymbióze je organelární proteom, který zahrnuje spíše méně proteinů odvozených z prokaryotického endosymbionta a že převažují proteiny odvozené od proteinů původního hostitele, který nad organelou převzal vládu. Dlouhá evoluční doba od získání mitochondrie a primárních plastidů (vyjma *Paulinella*) ovšem velice komplikuje správné zařazení proteinů do jednotlivých kategorií. Odlišení genů pocházejících z EGT je ale usnadněno tím, že lze poměrně dobře odlišit geny prokaryotického a eukaryotického původu (Ponce-Toledo et al., 2019).

## 6.2 EGT při endosymbióze komplexního plastidu

Při endosymbióze vyššího stupně dochází k pohlcení jiné eukaryotické buňky, která má ovšem mnohem větší a komplexnější genom, než primární endosymbiont. Geny ovládající plastidové funkce se nenacházejí pouze v plastidovém genomu, ale především v genomu jaderném. Podíl genů přenesených z plastidového genomu do nového jádra je obvykle zanedbatelný (Hrdá et al., 2012), přenos genů z mitochondrie endosymbionta nebyl dokumentován. Otázkou zůstává, kolik a jakých genů je přeneseno z jádra do jádra. Fylogenetické postavení těchto genů nám také napovídá, ze kterých skupin řas komplexní plastid pochází. Problém při klasifikaci plastidů



na sekundární, terciální a ev. plastidy vyššího řádu je ten, že mnoho genů v genomu hostitelské buňky zdaleka neodkazuje na jeden zdroj, tedy předpokládaného jediného endosymbionta, ale vykazuje evoluční původ napříč skupinami. Celkem dobře lze rozlišit geny pocházející ze zelené a červené větve archaeplastid, a proto již delší dobu překvapovaly nálezy "červených" genů u linií odvozující svůj plastid od zelených řas a naopak.

Fylogenomická analýza genomů rozsivek *Thalassiosira* a *Phaeodactylum* našla u každého druhu přibližně 2,5 tisíce proteinů odvozených od červených nebo zelených řas. Překvapivě až 70 % těchto genů bylo podle automatického třídění stromů odvozeno od genů zelených řas a ne řas červených, jak by odpovídalo původu jejich plastidu (Moustafa et al., 2009). Tento podíl byl tak zarážející, že byla data ručně přeanalyzována a jako výsledek EGT byl uznán jen omezený počet genů (286). Pouze 209 z nich bylo možné použít pro konstrukci stromů a z nich jen 28 bylo monofyletických se zelenými řasami, 139 s červenými řasami a u zbytku nebylo možno rozhodnout. Geny odvozené od červených řas tak mají výsledný podíl 68 %, nicméně "zelené" geny mají stále nezanedbatelný 14 % podíl (Deschamps & Moreira, 2012). Tento příklad také výborně ilustruje metodologické problémy při třídění genů. Jedním z nich je převaha určitého typu genů v databázích (v tomto případě "zelených"), což pokrývá fylogenomické analýzy, další otázka je výběr genů, které jsou skutečně důsledkem EGT (Deschamps & Moreira, 2012). Podobně analyzování příspěvku "zelených"/"červených" genů k EGT u *Chromera velia* vedlo nejprve k poměru 250/263 (Woehle et al., 2011), pozdější reanalýza určila poměr 23/9, přičemž u dalších 19 genů nebylo podle topologie stromu možno rozhodnout (Burki et al., 2012). V práci je kladen důraz na robustnost konstruovaných stromů. Zároveň upozorňuje na další problém, a to je kontaminace zkoumaných dat i dat umístěných ve veřejných databázích sekvencemi DNA, které jsou sekvenačními artefakty.

S většími či menšími metodologickými výhradami tak lze přijímat i další práce, z nichž několik uvádím dále. U chlorarachniofyta *Bigeloviella natans* bylo zjištěno 353 genů, které pochází z řas. Topologie fylogenetických stromů jednotlivých genů určily jejich podobnost ke známým genům v pořadí "zelené"/"červené"/nerozhodnuté (či jiné) jako 207/45/101, což odpovídá tomu, že plastid pochází ze skupiny Chloroplastida. U skrytěnky *Guillardia theta* bylo analyzováno 508 genů původem z řas s výsledkem 252/100/156, což naopak spíše neodpovídá předpokládanému původu plastidu ze skupiny Rhodofyta (Curtis et al., 2012). Obrněnka *Lepidodinium chlorophorum* se sekundárním zeleným plastidem, který nahradil původní červený plastid obrněnek, má mezi 22 rozpoznávanými geny s plastidovou funkcí část genů "zelených", ale zůstalo jí i mnoho genů "červených", které ale pochází z různých fylogenetických skupin (6/12/4) (Minge et al., 2010). Souhrnná studie genomu a proteomu několika desítek ochrofyt vedla k nalezení 770 proteinů targetovaných do plastidu, z nichž 263 bylo možné jednoznačně zařadit. Více než polovina z těchto genů (149) je původem z červených řas, ale celá čtvrtina (67) je původem z řas zelených (Dorrell et al., 2017).

### 6.3 Navrhovaná vysvětlení různorodého původu přenesených genů

Se vzrůstajícím počtem osekvenovaných genomů a transkriptomů v databázích (především ze skupiny červených řas) a s vylepšováním vyhodnocovacích algoritmů se určení podílu EGT stává přesnějším. Ukazuje se, že v jednotlivých genomech neustále zůstává nezanedbatelný podíl genů, které nepocházejí z předpokládaného endosymbionta, a že toto pozorování zřejmě není artefakt. Bylo vysloveno několik podobných hypotéz, které by mohly vysvětlit u dosud zkoumaných organismů s komplexním plastidem popsanou genetickou mozaiku.

Jedním z vysvětlení je, že u organismů s komplexním plastidem došlo k seriální kryptické endosymbióze; původní plastid pocházel z jiné větve archaoplastid než ten, který dnes pozorujeme, a že nalezené geny jsou pozůstatkem EGT prvotního endosymbionta. Seriální sekundární endosymbióza je dnes známa u obrněnek např. rodu *Lepidodinium*, kde primární zelený plastid nahradil původní peridinový plastid (Minge et al., 2010). V transkriptomu rozsivek *Thalassiosira* a *Phaeodactylum* s komplexním plastidem odvozeným z červené řasy byly nalezené "zelené" geny interpretovány jako pozůstatek hypotetického kryptického zeleného plastidu (Moustafa et al., 2009). Nahrazení jednoho plastidu plastidem jiným bývá ve většině případů úplné. Zajímavé je ztotožnění reliktu peridinového plastidu se světločivnou skvrnou ("eyespot") u obrněnek *Durinskia baltica*, *Glenodinium foliaceum* a *Kryptoperidinium foliaceum*. Eyespot má tři membrány stejně jako peridinový plastid. V transkriptomu obrněnek byly nalezeny geny, které odpovídají svým původem tomuto plastidu, a zdá se, že jsou skutečně targetovány do světločivné skvrny. Tato organela nemá zjevně fotosyntetickou funkci, ale zůstala jí funkce při tvorbě tetraprolů a isoprenoidů (Hehenberger et al., 2014).

Mnoho autorů se domnívá, že k tomu, aby se geny dostaly do jádra hostitele, nebylo potřeba, aby endosymbióza byla dokončena. Stačilo, aby nějakou dobu endosymbiont jako přechodný plastid přebýval uvnitř buňky. Tehdy mohlo dočasně docházet k přesunu genů do jádra. DNA mohla být také zainkorporována do jádra až poté, co přechodný endosymbiont zlyžoval. Těchto přechodných endosymbióz mohly některé organismy absolvovat i několik a v genomickém záznamu pak vidíme zastoupení genů z několika evolučních linií. Pro tento pohled se vžil název "nákupní košík" ("shopping bag" model), odkazující na způsob nakupování, kdy jsou v nákupním koší (jaderném genomu) položky (geny) z různých obchodů (organismů) (Larkum et al., 2007).

Dalším způsobem, jak mohlo dojít k horizontálnímu genovému přenosu, bylo získání genů nebo jejich částí z kořisti. Model "jste to, co jíte" byl původně koncipován pro přenos genů z bakteriální kořisti do genomu eukaryot (Doolittle, 1998), ale jistě může být rozšířen i na přenos z eukaryotní do eukaryotní buňky. Obrněnka *Dinophysis acuminata* požívá nálevníky *Myrionecta rubra*, který sám požívá skrytěnku *Geminigera cryophila*. *Myrionecta* si ale dočasně po několik týdnů drží kleptoplastid a odděleně i jádro a mitochondrii skrytěnky. Poté co se stane kořistí obrněnky, kleptoplastid přechází do nového hostitele už pouze se dvěma membránami a i zde několik týdnů přetrvává. V transkriptomu obrněnky bylo identifikováno

několik genů, které jsou směřovány do kleptoplastidu. Jeden z těchto genů *psbM* zcela nepochybně pochází přímo od kořisti *G. cryophila* (Wisecaver & Hackett, 2010).

Předchozí hypotézy nevylučují jedna druhou a navzájem se doplňují. Přesto zůstává jeden aspekt, který tyto hypotézy neřeší. Je to velký podíl "červených" genů v organismech se zeleným plastidem, který v opačném případě pozorujeme v menší míře. Byla proto navržena hypotéza "červeného koberec" ("red carpet hypothesis"), která předpokládá, že úspěšná endosymbióza komplexního zeleného plastidu nastala až poté, co bylo jádro hostitele jedním ze zmíněných způsobů obohaceno o geny z červených řas. Tyto geny poskytovaly důležité plastidové funkce a připravily tak takřikajíc červený koberec, který usnadnil následné přijetí zeleného endosymbionta (Ponce-Toledo et al., 2019).

## 6.4 EGT u skupiny Euglenida

Také u *Euglena gracilis* bylo publikováno již několik studií, týkajících se EGT. První z nich byla zaměřena právě na geny pocházející z "červené linie" řas. Nejprve bylo v transkriptomu vybráno 621 proteinů, které na fylogenetických stromech měly jako nejbližšího souseda zástupce řasy s primárním nebo komplexním červeným plastidem (oproti tomu jen 528 proteinů mělo jako nejbližšího souseda zástupce chloroplastid). Tato skupina stromů byla podrobena detailní analýze, která ukázala, že ve 14 z nich je sekvence *E. gracilis* monofyletická s vysokou podporou s "červenou linií". Pro doplnění byl zkoumán i transkriptom nefotosyntetického euglenida *Peranema trichophorum*, kde byly také identifikovány dva z těchto 14 "červených" genů. Autoři diskutují možnosti přenosu těchto genů do jádra a kromě HGT z kořisti původního eukaryovorního předka euglenofyt připouští i možnost EGT z hypotetického předchozího červeného plastidu (Maruyama et al., 2011).

Další studie se zaměřila na geny Calvinova cyklu u třinácti euglenofyt, pěti nefotosyntetizujících euglenidů a dvou diplomem. Byla provedena fylogenetická analýza celkem 11 hlavních enzymů cyklu, z nich u tří byla nalezena kromě plastidové i cytoplasmatická forma a některé další zahrnovaly více plastidových izoform. Bylo zjištěno, že proteiny nemají ani zdaleka jednotný původ. U čtyř nebylo možno rozhodnout, ale ze zbývajících sedmi se pouze u genu pro RuBisCO potvrdil jeho původ ze zelených řas. Enzym triosafosfát isomeráza má jako sesterský klád Rhodophyta a ostatní enzymy jsou příbuzné enzymům různých organismů s červeným komplexním plastidem (Markunas & Triemer, 2016).

Jak už bylo zmíněno, plastidy nemají jen fotosyntetickou funkci. V plastidech je mimo jiné také umístěna dráha pro syntézu tetrapyrrolů – hemu a chlorofylu. *E. gracilis* tak má pro tento účel dvě sady proteinů. Jedna sada je v mitochondrii a částečně v cytoplasmě a odpovídá enzymům primárních heterotrofních eukaryotů. Druhá sada se nalézá v plastidu, je sdílena většinou fotosyntetických eukaryotů a očekávalo se, že ji *E. gracilis* získala společně se svým zeleným plastidem (Kořený & Oborník, 2011). Fylogenetická analýza však ukázala, že jenom dva z deseti enzymů pocházejí ze zelených řas, jeden byl získán HGT z bakterií, jeden má neurčitý původ, ale šest enzymů pochází z řas s komplexním červeným plastidem (Kořený & Oborník,

2011; Lakey & Triemer, 2017). Bylo by možné spekulovat, že právě tato dráha může demonstrovat hypotézu "červeného koberce", a být onou prerekvizicí pro získání zeleného plastidu. Například také u řasy *Bigelowiella natans* s komplexním zeleným plastidem byly v uvedené dráze odhaleny geny původem z řas s červeným komplexním plastidem (Cihlář et al., 2016).

Jiný přístup byl zvolen v práci, kde u různých skupin řas s komplexním plastidem stopovali původ genů, které zcela nepochybně pochází od archaeplastid, jsou vysoce konzervované a jsou zapojené do esenciálních plastidových funkcí. V sestavených fylogenetických stromech bylo nalezeno 82 takových, které zahrnovaly alespoň tři fotosyntetizující organismy a zároveň klády Chloroplastida a Rhodophyta byly jednoznačně oddělené. U všech skupin řas s červeným komplexním plastidem byla většina genů (84 – 90 % podle skupiny) příbuzná rhodofytům. U euglenidů sice větší podíl tvoří geny pocházející z chloroplastid (56 %), ale "červené" geny mají podíl celých 36 %. U chlorarachniofyt, které mají také zelený komplexní plastid, vychází tento poměr ještě výrazně hůře, a to 44 % vs. 54 % ve prospěch "červených" genů (Ponce-Toledo et al., 2018).

Nejnovější genomická a transkriptomická data získaná pro *E. gracilis* byla také využita k odhadnutí podílu EGT na jaderném genomu. Z poměrně kompletního transkriptomu byly vytvořeny fylogenetické stromy pro jednotlivé transkripty, ze kterých bylo vyselektováno 4087 stromů, kde byla *E. gracilis* s vysokou podporou sesterským kládem pro některou z definovaných monofyletických linií. Mezi nimi bylo identifikováno 572 genů původem pocházejících z chloroplastid, 50 pocházejících z rhodophyt a 723 odvozených od řas s komplexním plastidem červené linie (skrytěnek, haptofyt a ochrofyty) (Ebenezer et al., 2019).

Obrázek doplňuje studie plastidového proteomu *E. gracilis*, získaného kombinací *in silico* predikce plastidových genů z celkového transkriptomu a proteomické analýzy plastidu. Přibližně polovina stromů ze získaných sekvencí splnila požadované kritérium, že *E. gracilis* je sesterským kládem pro některou z taxonomicky homogenních linií s vysokou podporou. Z těchto 416 vybraných proteinů bylo 95 příbuzných skupině Chloroplastida, 25 skupině Chlorarachniophyta a 34 těmto zeleným skupinám společně (dohromady tedy cca 37 %). Velký podíl genů tvoří geny odvozené od řas, kde nebyla příbuznost možná rozhodnout (107) a také geny odvozené od jiných než fotosyntetizujících eukaryot a prokaryot (70). Přesto genů odvozených od skupin s primárním nebo komplexním červeným plastidem bylo identifikováno 83, což tvoří podíl 20 % ze zkoumaných genů (Novák Vanclová et al., 2020).

Lze shrnout, že u *E. gracilis* je v genomu zastoupeno poměrně hodně genů, jejichž původ lze najít u řas s komplexním červeným plastidem. V genomu jsou také geny, které se tam dostaly pomocí HGT z bakterií, např. gen pro plastidovou porfobilinogen deaminázu a mitochondriální ferochalatázu (Kořený & Oborník, 2011; Lakey & Triemer, 2017) nebo bakteriální regulační faktor Rho, který je směřován do plastidu, kde pravděpodobně ukončuje transkripci některých genů, a který byl nalezen v transkriptomu *E. longa* i u dalších euglenofyt (Záhonová et al., 2018). Zdá se pravděpodobné, že tyto geny mohly eugleny získat z kořisti, vycházíme-li

z předpokladu, že jejich předek byl fagotrofní (Leander, 2004), a některé druhy euglenofyt jsou dokonce schopné lovit ještě dnes (Yamaguchi et al., 2012; Yoo et al., 2018).

Dovolím si zde uvést ještě jednu doměnkou. Výše uvedený případ obrněnky *Dinophysis* nás také navádí na možnost, že část genů, které *E. gracilis* získala přes svého endosymbionta, prasinofytní řasu příbuznou dnešnímu *P. parkeae*, nemusí být s touto řasou příbuzná. Například arktický příbuzný *Pyramimonas gelidicola*, stejně jako mnohé další prasinofytní druhy, je totiž mixotrofní a při nutričním nedostatku je bakteriovorní (Bell & Laybourn-Parry, 2003). Lze si tedy představit, že DNA z bakteriální a sinicové kořisti prasinofytní řasy se mohla přes její genom dostat až do euglenofyt.

## 7 Srovnání osmotrofního euglenida *Rhabdomonas costata* a fototrofního euglenida *Euglena gracilis*

Část genů, kterou zástupci euglenofyt nezdědily vertikálně, pochází z endosymbiotického genového přenosu spojeného se vznikem plastidu a pochází tedy z prasinofytní řasy *P. parkeae*. Není vyloučeno, že některé z nich mohly být do jejich jaderného genomu přeneseny už u jejich dávnějších předků, možná již u toho, který byl společný s dnes osmotrofním euglenidem *R. costata*. Osekvenování jeho transkriptomu a genomu může kromě doplnění dat pro tuto málo prozkoumanou fylogenetickou větev přispět i k lepšímu pochopení procesu endosymbiózy.

### 7.1 *Euglena gracilis* a *Rhabdomonas costata* – obecná charakterizace

*Euglena gracilis* patří do monofyletické skupiny fototrofních euglenidů s komplexním zeleným plastidem, která je odvozena od fagotrofních euglenidů (obr. 3) (Lax et al., 2019). *E. gracilis* je schopná žít i heterotrofně v podmínkách bez světla a/nebo bez kyslíku, kdy neprobíhá fotosyntéza. Je to sladkovodní druh s buňkou velkou 50 – 80  $\mu\text{m}$ , který má dva nestejně bičíky, z nichž pouze jeden vystupuje z periflagelární kapsy a slouží k pohybu. K pohybu také slouží helikálně uložené navzájem klouzající podélné pásy, kterých je několik desítek a které umožňují typický pohyb metabolii (Leander et al., 2017).

*E. gracilis* se velice dobře množí i v laboratorní monokultuře, což ji společně s velkou metabolickou komplexitou a flexibilitou předurčuje jako alternativní zdroj biopaliv, farmaceutik, kosmetických i výživových doplňků i jako součást čistírenských biokultur (Krajčovič et al., 2015; Suzuki, 2017). Zásobní látkou produkovanou za aerobních podmínek je  $\beta$ -1,3-glucan zvaný paramylon, který tvoří až 700 lineárně svázaných molekul glukánů. Paramylon je syntetizován v cytoplasmě jak u fototrofních tak i u heterotrofních euglenidů a jeho přítomnost tak není navázána na přítomnost plastidu, jako je tomu u zásobního škrobu rostlin. Za anaerobních podmínek se paramylon metabolizuje na estery vosků, což je spojeno s generováním ATP (Inui et al., 2017). Syntézu esterů vosků zajišťuje enzym wax syntáza, který má v cytosolu *E. gracilis* minimálně 2 izoformy příbuzné nejspíše proteinům chloroplastid, a dále také enzym wax ester syntáza/diacylglycerol acyltransferáza (WSD), který má anotovaných šest izoform, příbuzných bakteriálním genům, z nichž WSD2 má hlavní podíl na syntéze esterů vosků (Tomiyama et al., 2017). Podíl esterů vosků v metabolomu euglen rostoucích za anaerobních podmínek je přibližně 50 %. Za aerobních podmínek je naopak v metabolomu nejvíce zastoupen paramylon a to z cca 40 % ve světelné fázi a až 60 % v temnostní fázi (Matsuda et al., 2011).

Plastidový genom *E. gracilis* byl osekvenován již v roce 1993 (Hallick et al., 1993). Kromě plastidového genomu byl u *E. gracilis* publikován i mitochondriální genom (Dobáková et al., 2015), několik transkriptomických studií (Ebenezer et al., 2019; O'Neill et al., 2015; Yoshida et al., 2016), částečný jaderný genom (Ebenezer et al., 2019) a plastidový (Novák Vanclová et al., 2020) a mitochondriální proteom (Hammond et al., 2020).

*Rhbdomonas costata* patří do monofyletické skupiny primárně osmotrofních euglenidů společně s rody *Menoidium*, *Rhabdospira*, *Gyropaigne*, *Parmidium*, *Distigma* a *Astasia*. Také tato skupina se na fylogenetickém stromě větví mezi několika fagotrofními skupinami euglenidů (obr. 3) (Lax et al., 2019). Pohybuje se pomocí jednoho vystupujícího bičíku stejné délky jako je jeho buňka (20 – 30  $\mu\text{m}$ ), která má na povrchu 7 – 9 fúzovaných podélných pásů se širokými zářezy (Soukal, Hrdá et al., v přípravě). Buňka je tak poměrně rigidní a na rozdíl od euglenofyt u ní neprobíhá metabolie. V cytoplasmě je obsaženo velké množství paramylonových zrn (Zakryš et al., 2017; Soukal, Hrdá et al., v přípravě). *Rhbdomonas costata* je sladkovodní druh žijící v rašelinistích (Cabała, 2003), víceméně za anaerobních podmínek. V laboratorních podmínkách je pěstován ve směsné kultuře s bakteriemi a kultury se nedaří axenizovat. Kromě několika genů dosud genom a transkriptom *R. costata* nebyl sekvenován. Třetím výstupem mé disertace je práce, aktuálně ve stavu rukopisu, zaplňující tuto mezeru.

## 7.2 Velikost a struktura jaderného genomu *E. gracilis* a *R. costata*

Již první odhady velikosti genomu *E. gracilis* vedly ke zjištění, že tento genom je velmi velký, mnohokrát větší než genomy jiných zástupců kmene Euglenozoa – trypanosom. Odhady se v závislosti na způsobu měření (většinou zjišťování celkového množství DNA v buňce) a metod přepočítávání na počet nukleotidů lišily v rozsahu téměř jednoho řádu od 1 Mbp po 9 Mbp (pro přehled Ebenezer et al., 2017). Poslední genomická studie na rozdíl od předchozích studií velikost genomu odhaduje ze sekvenačních dat, která byla získána několikerým masivním sekvenováním na různých platformách. Genom se nepodařilo zatím celý složit, jeho současná pracovní verze je velice fragmentovaná a zcela určitě nekompletní. Podle odhadu má haploidní genom velikost okolo 500 Mbp, ale ploidie není známa a celková dosud asemblovaná velikost genomu je přes 1 400 Mbp. Porovnáním dobře složených částí genomu (kontigy delší než 10 kbp) s geny identifikovanými v souběžně osekvenovaném transkriptomu bylo zjištěno, že jen asi 1 % genomu jsou kódující sekvence (Ebenezer et al., 2019). Transkriptom má vyšší obsah GC párů než genom, z čehož lze usuzovat, že v genomu je velké množství AT bohatých a tedy zřejmě repetitivních sekvencí (Ebenezer et al., 2017).

Genomická DNA obsahuje také velké množství intronů, které jsou jednak konvenční s kanonickými hranicemi GT/AG, které jsou vystřihovány ribonukleoproteinovým komplexem spliceozomem, a jednak nekanonické, které nemají hranice s jednoznačnou sekvencí nukleotidů, ale typicky tvoří vlásenku, která je z DNA odstraňována zatím nejasným způsobem (Milanowski et al., 2016). Bližší analýza intronů byla provedena u několika euglenidů (včetně heterotrofních) pro geny fibrillarin (*nop1p*) (Russell et al., 2005),  $\alpha$ -tubulin (*tubA*),  $\beta$ -tubulin (*tubB*) a  $\gamma$ -tubulin (*tubG*) (Canaday et al., 2001; Milanowski et al., 2014), heat shock protein 90 (*hsp90*) a glycerinaldehyd-3-fosfát dehydrogenáza (*gapC*) (Henze et al., 1995; Milanowski et al., 2016). Ve všech těchto genech byly nalezeny jak konvenční, tak nekonvenční introny, ev. introny intermediální s přechodnými znaky. Množství intronů se liší v jednotlivých liniích euglenidů, některé jsou umístěny ve stejném místě, jiné jsou liniově specifické (Milanowski et al., 2016) a mohou tak být do budoucna dobrým znakem pro fylogenetické analýzy (viz obr. 5 v příloženém rukopise).

Zastoupení jednotlivých typů intronů u *E. gracilis* bylo v jaderném genomu zjišťováno u kontigů, jež měly délku větší než 10 kb. Zatímco konvenčních intronů bylo 80 % a přechodných typů 11 %, nekonvenčních intronů bylo necelých 9 % (Ebenezer et al., 2019). Konvenční introny byly nalezeny i u genů směřovaných do chloroplastu, některé jsou dokonce přítomny v signální sekvenci. Tyto introny jsou poměrně sekvenčně podobné a tak se jejich případnou rekombinací se mohla tato sekvence přenášet v genomu *E. gracilis* mezi jednotlivými geny (Vesteg et al., 2010).

Úsilí a prostředky věnované úplnému osekvenování genomu ikonické a i po ekonomické stránce významné *E. gracilis* nevedly ke konečnému cíli. Nepřekvapí tedy, že ani snaha o osekvenování genomu *R. costata* v podmínkách mnohem skromnějších, nebyla zakončena uspokojivým výsledkem. Celková délka asemblovaného genomu je jen 100 Mbp a získaná data jsou extrémně fragmentovaná, takže z nich nebylo možné získat relevantní predikci genů. Fragmentovaná data jsou zřejmě z části výsledkem použité metody přípravy knihovny, kdy vzhledem k malému množství výchozího materiálu izolovaného ze směsné kultury byla použita metoda celogenomové amplifikace (WGA), a z části možná opět přítomností repetice a velkého množství intronů (Soukal, Hrdá et al., v přípravě).

Genomová data však společně s daty z transkriptomu byla použita k predikci intronů v jaderných genech. Na přítomnost intronů byly analyzovány všechny dříve publikované geny – *nop1p*, *tubA*, *tubB*, *tubG*, *hsp90* a *gapC*, ve kterých bylo nalezeno celkem 29 intronů s kompletními konci, které jsou všechny konvenční s hranicemi GT/AG. Introny v genech *tubA* a *tubB* osmotrofních eugen *R. costata* a *Menoidium bibacillatum* jsou ve stejných místech, ale pouze jedno umístění je shodné i u rodu *Euglena*. Podobně je to i u dalších genů. Namapováním celého transkriptomu na genom bylo detekováno více než 11 tisíc pravděpodobných intronů, z nichž 90 % má konvenční hranice. Nekonvenčních intronů bylo nalezeno sedm (Soukal, Hrdá et al., v přípravě).

### 7.3 Analýza celkového transkriptomu

Na rozdíl od neúplně osekvenovaného genomu je transkriptom *E. gracilis* již poměrně kompletní i proto, že jsou k dispozici data ze tří transkriptomických studií. V první byly získány dva datasety z buněk pěstovaných za světla a za tmy (O'Neill et al., 2015). V druhé byly získány opět dva datasety z buněk pěstovaných v aerobních a anaerobních podmínkách (Yoshida et al., 2016). Ve třetí byl kromě transkriptomu sekvenován a porovnáván i genom (Ebenezer et al., 2019). Jednotlivé transkriptomty byly v poslední práci reanalyzovány a pomocí programu BUSCO byla posouzena kompletnost spojeného transkriptomu, která dosáhla cca 88 %. Kromě zjištění, že ve spojeném datasetu chybí pouze 19 z 303 univerzálních eukaryotických genů, bylo konstatováno, že všechny tři původní datasety byly také poměrně kompletní (Ebenezer et al., 2019). Rozdíl mezi transkriptomty z buněk pěstovaných za světla a za tmy byl poměrně velký, kromě 17,5 tisíce genů nalezených za obou stavů, bylo za světla přítomno dalších 5,4 tisíce a za tmy jiných 9,3 tisíce genů (O'Neill et al., 2015). Transkriptomty z kultur buněk držných po 24 hodin za anaerobních podmínek a těch pěstovaných aerobně si



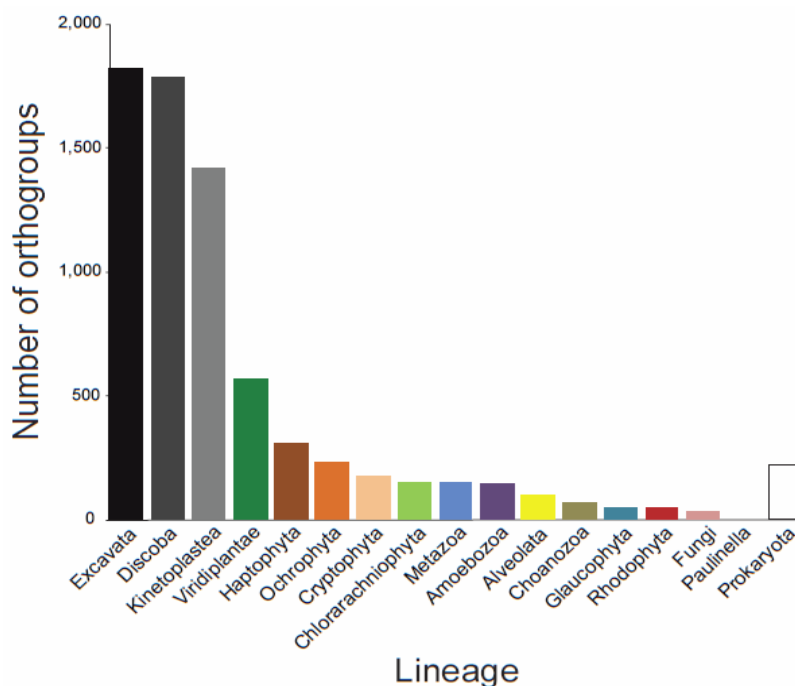
odpovídali z 95 %. U 2080 transkriptů bylo možné pozorovat rozdíly v expresi. Signifikantně se snížilo zastoupení genů zahrnutých v některých metabolických drahách (fotosyntéza, metabolismus nukleotidů, TCA cyklus a oxidativní fosforylace). Navzdory očekávání se pouze částečně zvýšila exprese genů zodpovědných za degradaci zásobního paramylonu a syntézu esterů vosků a předpokládá se, že nejen tyto metabolické dráhy jsou regulovány až post-transkripčně, stejně jako u kinetoplastid (Ebenezer et al., 2019; Yoshida et al., 2016).

Transkriptom *E. gracilis* je velice bohatý, což souvisí s komplexním původem genů. *E. gracilis* má všechny základní metabolické dráhy zahrnující glykolýzu, glukoneogenezi, citrátový cyklus, pentozovou dráhu, Calvinův cyklus, ale také dráhy pro biosyntézu karotenoidů, izoprenoidů, vitamínů a esterů vosků (Ebenezer et al., 2019; O'Neill et al., 2015; Yoshida et al., 2016). U mnoha genů bylo nalezeno množství paralogů a dokonce pro syntézu některých klíčových metabolitů existují dvě rozdílné metabolické dráhy s rozdílnými prekurzory, které mají rozdílný evoluční původ. Příkladem je biosyntéza isoprenoidů, která zahrnuje dráhu methylethritol fosfátu (MEP) v plastidu a dráhu mevalonátu (MEV) v cytoplasmě (O'Neill et al., 2015). Biosyntéza tetrapyrólů také probíhá buď v plastidu z glutamyl-tRNA jako vstupního prekurzoru (tzv. C5 dráha), nebo biosyntéza probíhá částečně v mitochondrii a částečně v cytosolu, počínaje kondenzací výchozích reaktantů succinyl-CoA a glycinu (tzv. C4 dráha) (Kořený & Oborník, 2011). Zachování obou těchto drah je u řas s komplexním plastidem spíše výjimečné a bylo popsáno už jen u *Bigeloviella natans*. U eukaryot s primárním plastidem je vždy přítomná jen plastidová dráha (Cihlář et al., 2016). Přestože bylo u eugleny popsáno velké množství poměrně úplných metabolických drah, ve všech transkriptomických studiích nebyla pro více než polovinu nalezených transkriptů nalezena a anotována žádná funkce.

Příspěvek genů do celkového transkriptomu od jednotlivých předků byl zatím nejpřesněji odhadnut z cca 4000 genů, pro něž bylo možné vytvořit dostatečně robustní fylogenetické stromy, kde bylo možné jednoznačně definovat sesterskou linii euglen. Nejpočetnější skupinu tvoří geny příbuzné linii Excavata (44 %), které tvoří vertikálně zděděnou část genomu. Endosymbiotický genový přenos z pohlcené zelené řasy přispěl do genomu přibližně 14 % genů. Další 18 % genomu tvoří geny příbuzné s geny řas s červeným komplexním plastidem, které byly získány buď postupně horizontálním přenosem nebo během kryptického endosymbiotického přenosu. Asi 5 % genů je patrně výsledkem HGT z prokaryot a zbytek pochází buď z dalších HGT z jiných skupin nebo jsou to nerelevantní artefakty (Ebenezer et al., 2019) (obr.11.).

Osekvenováním transkriptomu *R. costata* bylo získáno 39,585 transkriptů, z nichž pouze 26,052 má BLAST hit v databázi NCBI a jen části z těchto genů bylo přiřazena anotace z KEGG databáze (13,130). Kompletnost transkriptomu byla pomocí programu BUSCO odhadnuta na 88,5 % (Soukal, Hrdá et al., v přípravě). Podle stejné metodiky jako u *E. gracilis* bylo zkonstruováno 3445 fylogenetických stromů, kde je *R. costata* s vysokou podporou sesterským kládem pro některou z definovaných monofyletických linií. Geny příbuzné linii Excavata tvoří 56 %. Druhou největší skupinu tvoří původem prokaryotické geny (15 %), zde je ovšem možné, že část z nich tvoří kontaminace bakteriálními geny pocházejícími z agnotobiotické kultury, které

se nepodařilo odfiltrovat. Necelých 6 % tvoří geny příbuzné chloroplastidům a 7,5 % geny příbuzné s geny řas s červeným komplexním plastidem, což je výrazně méně než u *E. gracilis* (viz obr. 3 v příloženém rukopise).



Obr. 11. Taxonomická příslušnost 4087 predikovaných proteinů *Euglena gracilis*, které tvoří robustní klan s jednotlivými taxonomickými skupinami. Sloupce Discoba a Kinetoplastea jsou podmnožinami superskupiny Excavata (převzato z Ebenezer et al., 2019).

V transkriptomu *R. costata* byly nalezeny tři kopie 1,3- $\beta$ -D-glucan syntázy, které mohou produkovat paramylon, dobře rozpoznatelný v cytoplasmě. Transkripty jsou homologní s enzymem, nalezeným v mitochondriálním proteomu *E. gracilis* (Hammond et al., 2020), ale nemají na rozdíl od něj rozpoznatelný mitochondriální signální peptid. Degradaci paramylonu zajišťuje nalezená endo-1,3(4)- $\beta$ -glukanáza. Cytosolická wax syntáza homologní s enzymem *E. gracilis* nalezena nebyla, což ovšem není překvapivé, protože podle BLAST (Basic Local Alignment Search Tool) je tento enzym příbuzný skupině Chloroplastida a můžeme se domnívat, že pochází z EGT. Bylo ovšem identifikováno 34 transkriptů pro enzym wax ester syntáza/diacylglycerol acyltransferáza, které k syntéze vosků také slouží (Tomiya et al., 2017).

Byly nalezeny všechny enzymy C4 dráhy zajišťující syntézu tetrapyrrolů u heterotrofů, která probíhá částečně v mitochondrii a částečně v cytoplasmě (Kořený & Oborník, 2011; Lakey & Triemer, 2017) (viz obr. 9 v příloženém rukopise). U všech 8 genů pro tyto enzymy tvoří *Euglena* s dalšími fototrofními euglenami a *Rhabdomonas* s dalšími osmotrofními euglenami,

pokud jsou jejich sekvence známy, sesterské větve na fylogenetických stromech. Díky tomu bylo možno potvrdit fylogenetický původ těchto genů. Šest genů, 5-aminolevulinát syntáza (*ALAS*), porfobilinogen syntáza (*ALAD*), porfobilinogen deamináza (*PBGD*), uroporphyrinogen syntáza (*UROS*), uroporphyrinogen decarboxyláza (*UROD*) a protoporphyrinogen oxidáza (*PPOX*), je příbuzných genům eukaryotickým organismům a dva, coproporphyrinogen III oxidáza (*CPOX*) a ferochelatáza (*FeCH*), byly získány HGT. Pro gen *UROS* jsme také odhalili v transkriptomu *E. gracilis* jeho druhý paralog, který je příbuzný skupině Chloroplastida a mohl by pravděpodobně být funkční v plastidu. U *R. costata* nebyla nalezena ani jedna z pěti euglenofytních izoform *CPOX* závislé na kyslíku. Nalezli jsme zato minimálně 3 izoformy nepříbuzné, na kyslíku nezávislé *CPOX*, která by měla zprostředkovávat stejnou reakci. Fylogenetická analýza ukázala jejich příbuznost s bakteriemi. Transkripty pro tento enzym jsme našli i v transkriptomu *E. gracilis*. Žádný homolog s geny plastidové C5 dráhy *E. gracilis* nebyl u *R. costata* nalezen. Společně se zjištěním, že *R. costata* má celkově velice málo genů příbuzných fotosyntetickým organismům, nás vede k přesvědčení, že tento organismus plastid nikdy neměl, což dobře zapadá do hypotézy pozdního získání chloroplastu u euglenofyt.

## 7.4 Plastidový genom a proteom *E. gracilis*

Plastidový genom *E. gracilis* byl publikován již v roce 1993 (Hallick et al., 1993) a byl diskutován v kapitole tři.

Z celkového transkriptomu *E. gracilis* bylo vybráno 1902 proteinů (Ebenezer et al., 2019), které jsou směřovány do plastidu. Bylo nalezeno 70 proteinů s funkcí v elektrontransportních řetězcích a světloběrných anténách, dále celá C5 tetrapyrolová dráha pro syntézu hemu i navazující dráha pro syntézu chlorofylů, a také methylethritol fosfátová dráha biosyntézy terpenoidů. Některé předpokládané geny nebyly nalezeny. Asi polovina proteinů má známou funkci, zbytek proteinů targetovaných do plastidu je bez známé funkce a mohou mít další metabolický potenciál (Novák Vanclová et al., 2020).

Získaný plastidový proteom je také poměrně velký a obsahuje 1345 proteinů, z nichž 48 je kódováno plastidovým genomem. Porovnání s *in silico* predikcí z celkového transkriptomu ovšem ukázalo velké rozdíly v obou data setech, neboť se oba shodovaly jen ve 474 proteinech. Rozdíl je zřejmě způsobený nesprávnou predikcí komplikovaného bipartitního plastidového targetovacího signálu, omezeními danými použitými metodami nebo prostě jen jiným momentálním fyziologickým stavem buněk. Proteom je nicméně podobně velký jako plastidový proteom huseničky a několikrát větší, než jiné získané proteomy (Novák Vanclová et al., 2020).

Předpokládané biosyntetické dráhy byly potvrzeny, ale zároveň byly popsány i některé nečekané vlastnosti plastidu. Byla rekonstruována dráha pro produkci tokoferolu (vitamin E), který se účastní ochrany před oxidativním stresem (Watanabe et al., 2017). Biosyntéza aminokyselin v plastidu probíhá jen v menší míře. Došlo k rozpojení metabolických drah pro syntézu karotenoidů a phytolu, kvůli chybějícímu enzymu spojujícímu tyto dráhy. Naopak byla v proteomu detekována další isoforma alternativní oxidázy, která je jinak přítomna

v mitochondrii. V plastidu jsou přítomné dva sety SUF dráhy pro skládání železo-sírných klastrů, z nichž jedna byla získána HGT z chlamydií (Novák Vanclová et al., 2020).

Na základě proteomických dat mohl být částečně rekonstruován systém mebránových přenašečů, umožňujících vstup proteinů do plastidu. Proteiny jsou uloženy v transportních měchýřcích, které fúzí s vnější membránou, pravděpodobně za účasti plastidových homologů fúzních proteinů. V případě, že je prostřední membrána pozůstatkem vnější membrány primárního plastidu, měly by být proteiny dále přeneseny TOC translokázami, ovšem žádný jejich homolog nebyl nalezen. V proteomu byly naopak nalezeny tzv. derlin-like proteiny, které by mohly vytvářet potřebný kanál a které jsou součástí tzv. SELMA systému ve vnější ze dvou prostředních membrán komplexního červeného plastidu. I když další komponenty tohoto systému nebyly nalezeny a jeho existenci tedy nelze potvrdit, společně s absencí TOC translokonů by to mohlo svědčit pro hostitelský původ této prostřední membrány. Na vnitřní membráně jsou přítomny 3 izoformy TIC22, které jsou zřejmě schopny formovat příslušný kanál, který je velice zjednodušený. Mechanismus přenosu přes membrány se podstatně liší od všech ostatních organismů s plastidem a pravděpodobně využívá další dosud nenalezené přenašeče, které jsou schopny rozpoznat euglenidí TP, které mají několik specifických vlastností (Novák Vanclová et al., 2020).

Fylogenetická analýza evolučního původu jednotlivých plastidových proteinů byla provedena pro 416 proteinů, u nichž bylo možné vytvořit dostatečně robustní fylogenetické stromy, kde bylo možné jednoznačně definovat sesterskou linii *E. gracilis*. Geny příbuzné fototrofním eukaryotům tvoří většinu (346). Celkem 57 genů je příbuzných nefotosyntetickým eukaryotům, především skupině Discoba, a alespoň některé z nich jsou pravděpodobně geny, které pochází z hostitelského organismu a získaly v plastidu novou funkci. Třináct genů je příbuzných prokaryotům a byly získány horizontálním genovým přenosem (Novák Vanclová et al., 2020).

## 7.5 Mitochondriální genom *E. gracilis* a *R. costata*

Mitochondriální genom *E. gracilis* je v porovnání dalšími zástupci skupiny Euglenozoa redukován a obsahuje pouze fragmentované geny pro dvě podjednotky rRNA a sedm genů kódujících proteiny komplexu I (*nad1*, *nad4* a *nad5*), komplexu III (*cob*) a komplexu IV (*coxI*, *coxII* a *coxIII*). Geny pro proteiny komplexu V, ribozomální proteiny a tRNA nebyly nalezeny (Dobáková et al., 2015). Mitochondriální genom je ve formě krátkých úseků DNA o délce mezi 1 – 9 kb (Dobáková et al., 2015; Spencer & Gray, 2011). V celkovém uspořádání mtDNA se Euglenida liší od ostatních skupin kmene Euglenozoa. Kinetoplastida mají unikátní síť propojených maxikroužků a minikroužků s celkem 18 geny kódujícími proteiny, které musí být pro správnou transkripci editovány vložением jednotlivých uridinů. Diplonemida mají mtDNA na kroužcích dvojí délky, které nesou celkem 10 genů kódujících proteiny. Geny jsou velice fragmentované a pro správnou transkripci je nutný jejich trans-sestřih a také určitý druh editace (Faktorová et al., 2016). U *E. gracilis* k editaci RNA nedochází a tato vlastnost tak zřejmě není evolučně původní. Lze však usuzovat, že pro Euglenozoa je typická velká fragmentace mitochondriální DNA (Hammond et al., 2020).

V celkovém genomu *R. costata* byly nalezeny částečné homology všech sedmi genů kódujících mitochondriální proteiny, které byly popsány u *E. gracilis*, nelze však s jistotou říci, zda to nejsou jaderné kopie genů, či dokonce jen jaderné pseudogeny. Mezi transkripty targetovanými do mitochondrie tyto geny nalezeny nebyly, nicméně bylo nalezeno 141 proteinů, které se podílejí na metabolismu RNA a DNA, biogenezi ribosomů a translaci, kterýžto aparát ukazuje na přítomnost funkční mitochondriální DNA (Soukal, Hrdá et al., v přípravě).

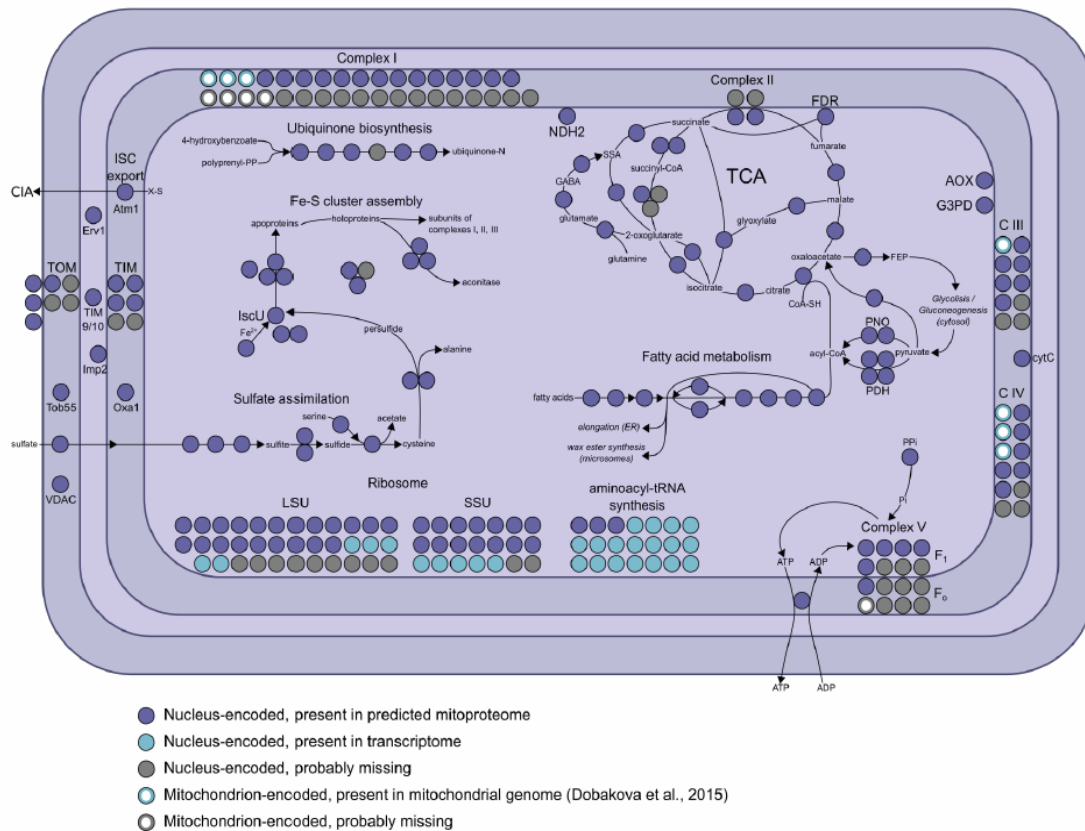
## 7.6 Mitochondriální proteom *E. gracilis*

Z transkriptomu predikovaný mitochondriální proteom *E. gracilis* byl publikován v roce 2019 v rámci genomického projektu (Ebenezer et al., 2019). Celkem bylo identifikováno 1093 proteinů, které mají pravděpodobně mitochondriální funkci. V navazující proteomické studii byl z vyčištěných mitochondrií získán proteom čítající 1786 proteinů včetně pěti ze sedmi proteinů kódovaných přímo v mitochondriálním genomu (Hammond et al., 2020). Vzhledem k tomu, že se oba sety částečně překrývají (552 proteinů), se odhaduje, že celkový mitochondriální proteom *E. gracilis* čítá okolo 2500 proteinů. To je dvakrát více, než bylo nalezeno např. u kinetoplastid, kvasinek nebo myší, a několikrát více než u jiných zkoumaných prvoků. Mitoproteom je velice komplexní a to je dáváno do souvislosti se získáním komplexního plastidu, kdy některé plastidové dráhy byly přesměrovány také do mitochondrie. Podobná velikost mitoproteomu se předpokládá u modelového huseníčku, byť zatím je u této rostliny experimentálně potvrzeno jen necelých 850 proteinů (Lee et al., 2013).

Celkový metabolický potenciál *E. gracilis* je zřejmě obrovský a velice komplexní, neboť celých 57 % proteinů nemá dosud známou funkci, a zároveň bylo odhaleno množství duplikovaných genů a paralogů, které mohou mít částečně rozdílné vlastnosti. Mitochondriální translokázové komplexy jsou poměrně unikátní a obsahují mnoho proteinů nalezených u kinetoplastid, včetně homologů archaických translokáz ATOM. Bylo identifikováno také více než 100 mitochondriální ribozomálních proteinů, které jsou jednak konvenční a jednak Euglenozoa nebo *Euglena* specifické. Část těchto proteinů má patrně duální signální peptidy a jsou využívány zároveň i v plastidu. Totéž platí pro aminoacyl-tRNA syntetázy. *E. gracilis* je v mitochondrii schopná syntetizovat 11 aminokyselin a zřejmě také asimilovat a metabolizovat sulfát, byť některé klíčové enzymy nebyly v mitoproteomu identifikovány. K biosyntéze železo-sírných klastrů využívá *Euglena* klasickou ISC dráhu. Dráha biosyntézy ubichinonu je doplněna o následný krok, kdy je připojena aminová skupina za vzniku rhodochinonu (Hammond et al., 2020).

Další studie mitoproteomu *E. gracilis* byla provedena pro dva stavy – v aerobním a hypoxickém prostředí (Tamaki et al., 2019). Získaný set obsahuje 714 unikátních proteinů, z nich u 116 bylo možno signifikantně určit rozdílnou expresi ve dvou experimentálních stavech. Práce je zaměřena na energetický metabolismus mitochondrie a doplňuje a potvrzuje výsledky získané v předchozích studiích. *E. gracilis* má nebyvalou schopnost využívat velmi různorodé energetické zdroje a je při tom velmi flexibilní. V aerobním prostředí funguje mitochondrie *E. gracilis* podobně jako mitochondrie jiných organismů, ale má schopnost v případě nedostatku

kyslíku přepnout metabolismus směrem k syntéze mastných kyselin, jejichž přeměnou na vosky jsou ze systému odebrány nadbytečné elektrony, aniž by se zastavila syntéza ATP. Kromě toho využívá ještě několika ne zcela běžných enzymů a vedlejších drah, které dále zvyšují flexibilitu mitochondrií (Hammond et al., 2020; Zimorski et al., 2017) (obr. 12).



Obr.12. Mapa mitochondriálního metabolismu *Euglena gracilis*. Barevné kroužky znázorňují jednotlivé proteiny. Tmavomodře jsou transkripty kódované v jádře s predikovanou mitochondriální lokalizací; světlemodře jsou transkripty bez mitochondriálního signálního peptidu; šedé kroužky obvyklé mt proteiny, jejichž transkripty nebyly nalezeny; světlemodré kroužky s bílým středem jsou proteiny kódované v mitochondriálním genomu (Dobáková et al., 2015), šedobíle jsou ztracené geny kódované původně v mitochondriálním genomu (převzato z Ebenezer et al., 2019).

Aerobní dráha začíná aerobní dekarboxylací pyruvátu kompletním pyruvát dehydrogenázovým komplexem. Vzniklý acetyl-CoA vstupuje do modifikovaného Krebsova cyklu, kde jsou dva klíčové enzymy  $\alpha$ -ketoglutarát dehydrogenáza (chybí podjednotky E1 a E2) a sukcinyl-CoA syntetáza nahrazeny dvěma enzymy  $\alpha$ -ketoglutarát dekarboxylázou a sukcinát semialdehydu dehydrogenázou. Ta produkuje ze sukcinátsemialdehydu sukcinát a zbytek cyklu pak probíhá konvenčně (Shigeoka et al., 1986). Kromě této netradiční spojky, někdy nazývané  $\alpha$ -proteobakteriální "shunt", byla v proteomu a transkriptomu identifikována i další vedlejší dráha, tzv. GABA "shunt", kdy je  $\alpha$ -ketoglutarát sérií tří enzymů převeden přes glutamát opět na SSA a cyklus pokračuje, jak je uvedeno výše. Elektrony jsou předávány na kyslík

přes mnohasubjednotkové části dýchacího řetězce, které kromě konvenčních podjednotek sdílí mnoho specifických podjednotek s kinetoplastidy (Ebenezer et al., 2019; Hammond et al., 2020; Perez et al., 2014). S kinetoplastidy *E. gracilis* sdílí také alternativní respirační dráhu, která zahrnuje glycerol-3-fosfát dehydrogenázu, alternativní NADH dehydrogenázu II a alternativní oxidázu, která je schopná přenášet elektrony z ubichinonu přímo na kyslík (Perez et al., 2014). V mitochondrii eugleny je také zcela neočekávaně lokalizován glyoxalátový cyklus, který umožňuje syntézu čtyřuhlíkatého sukcinátu (a potažmo dalších produktů) z dvouuhlíkatého acetyl-CoA, což umožňuje využití dalších substrátů, především ethanolu a acetátu, jako zdrojů pro syntézu organických látek (Nakazawa, 2017).

Při nedostatku kyslíku klesá hladina podjednotek E1 $\alpha$  a E1 $\beta$  pyruvát dehydrogenázového komplexu, ale zato se zvyšuje množství pyruvát:NADP<sup>+</sup>oxidoreduktázy (*PNO*), enzymu obsahujícího pyruvát:ferredoxin oxidoreduktázovou doménu, známou z hydrogenozómů (Nakazawa et al., 2000; Tamaki et al., 2019). Také *PNO* produkuje acetyl-CoA, který je opakovaně vstupním substrátem při syntéze mastných kyselin. Série čtyř enzymů cyklicky prodlužují řetězce acylovaného koenzymu A až do délky 8 – 18 uhlíků. Pro syntézu mastných kyselin s lichým počtem uhlíků je v prvním kroku kondenzován acetyl-CoA s propionyl-CoA, který je produkován metylmalonyl-CoA dráhou. Ta využívá jako substrát malát, který v anaerobním prostředí více exprimovaná NADP<sup>+</sup> dependentní malát dehydrogenáza konvertuje na fumarát (Tamaki et al., 2019). Fumarát je pak fumarát reduktázou zakotvenou v membráně redukován na sukcinát, kdy donorem elektronů je elektronový přenašeč rhodochinon, který má nižší redox potenciál než ubichinon (Hoffmeister et al., 2004). Dvoupodjednotková sukcinyl-CoA syntetáza, která je v mitochondrii přítomna, ale pravděpodobně se neúčastní Krebsova cyklu, běží v opačném směru a produkuje sukcinyl-CoA. Z něho přes metylmalonyl-CoA vzniká propionyl-CoA. Výsledné řetězce acylovaného CoA jsou transportovány do cytosolu a mikrozómů, kde je bez přítomnosti kyslíku fatty acyl-CoA reductáza a následně wax syntáza konvertuje na vosky, které *Euglena* využívá jako zásobní látky (Zimorski et al., 2017). Téměř u všech enzymů metylmalonylové dráhy a syntézy mastných kyselin byla nalezena zvýšená produkce těchto enzymů za anaerobiózy (Tamaki et al., 2019).

## 7.7 Mitochondriální proteom *R. costata* (*in silico*)

Mitochondriální proteom *R. costata* získaný *in silico* z celkového transkriptomu je připraven k publikaci a je součástí této práce. Celkem bylo identifikováno 1518 potenciálních mitochondriálních proteinů, které mají signální peptid nebo jsou homologní s anotovanými mitochondriálními enzymy *E. gracilis*. Celkově lze říci, že bylo nalezeno o něco méně ribozomálních proteinů, translokátorů TOM/TIM i podjednotek komplexů dýchacího řetězce než u fotosyntetického protějšku. Nicméně hlavní metabolické dráhy se podařilo anotovat celé, s výjimkou některých proteinů typických pravděpodobně jen pro *E. gracilis*. *R. costata* žije v mikroaerofilním prostředí a proto nepřekvapuje, že u něj byla nalezena aerobní i anaerobní dráha energetického metabolismu (viz obr. 3 v příloženém rukopise) (viz obr. 8 v příloženém rukopise).

Na rozdíl od *E. gracilis* není v mitochondrii přítomen pyruvát dehydrogenázový komplex, jmenovitě chybí PDH E1  $\alpha$  i  $\beta$  podjednotky a stejně tak E2 podjednotka. Byla nalezena pouze podjednotka E3 – dihydrolipoyl dehydrogenáza, která je ovšem součástí kompletního glycin cleavage systému. *Rhabdomonas costata* má pět izoform *PNO*, které mohou produkovat acetyl-CoA. Je možné, že tyto izoformy se mohou lišit v kinetických vlastnostech umožňujících optimalizaci reakce za různých podmínek, což by jistě stálo za experimentální ověření. Enzymy Krebsova cyklu byly identifikovány všechny včetně enzymů  $\alpha$ -ketoglutarát decarboxylázy a sukcinát semialdehyde dehydrogenázy ( $\alpha$ -proteobakteriální "shunt"), přičemž i zde chybí E1 a E2 podjednotka  $\alpha$ -ketoglutarát dehydrogenázy. Oba tyto enzymy mají  $\alpha$ -proteobakteriální původ a byly pravděpodobně získány HGT. Je přítomný GABA "shunt". V komplexu II byly stejně jako u *E. gracilis* identifikovány jen dvě konvenční subjednotky 1 a 2, z nichž druhá je transkribována ve dvou kusech, stejně jako u eugleny a kinetoplastid (Gawryluk & Gray, 2010; Hammond et al., 2020). Chybí alternativní oxidáza přítomná nejen u kinetoplastid a *E. gracilis*, ale i u *Eutreptiella gymnastica* (Hammond et al., 2020). Enzymy metylmalonyl-CoA dráhy i dráhy syntézy mastných kyselin jsou u *R. costata* v transkriptomu přítomny (stejně tak u *Etl. gymnastica*).

V dráze syntézy ubiquinonu nebyl nalezen enzym zajišťující předposlední krok dráhy ani u jednoho zkoumaného euglenida. Je pravděpodobné, že tento gen byl ztracen a proto, aby dráha mohla běžet, byl nahrazen genem jiným, jehož identita není známa. Pokud se tak stalo, došlo ke ztrátě u společného předka euglenidů. Rhodochinon je syntetizován na membráně rhodochinon methyltransferázou. Asimilace sulfátu není vyloučena, ale u *R. costata* chybí několik enzymů. U *E. gracilis* se usuzuje, že dráha byla doplněna geny pocházejícími z plastidu (Ebenezer et al., 2017), který ovšem *R. costata* nemá a dráha tak skutečně nemusí být kompletní a jejím konečným produktem je kofaktor PAPS (fosfoadenosyl fosfosulfát). ISC dráha pro tvorbu železo-sírných klastrů je přítomna celá. U *R. costata* jsme nenalezli bifunkční enzym isocitrát lyáza/malát syntáza, který je zapojen do glyoxalátového cyklu u *E. gracilis*. Jelikož nebyl nalezen ani u *Etl. gymnastica*, lze tuto dráhu považovat u *E. gracilis* za unikátní v rámci euglenidů.



## 7.8 Publikace III

Soukal, P., Hrdá, Š., Karnkowska, A., Milanowski, R., Szabová, J., Hroudová, M., Strnad, H., Vlček, Č., Čepička, I. & Hamp, V. (2020).

**Genome and transcriptome draft of the heterotrophic euglenid *Rhabdomonas costata*.**

Rukopis v přípravě.



# Genome and transcriptome draft of the heterotrophic euglenid *Rhabdomonas costata*

Petr Soukal<sup>1#</sup>, Štěpánka Hrdá<sup>1#</sup>, Anna Karnkowska<sup>2</sup>, Rafał Milanowski<sup>2</sup>, Jana Szabová<sup>1</sup>, Miluše Hroudová<sup>3</sup>, Hynek Strnad<sup>3</sup>, Čestmír Vlček<sup>3</sup>, Ivan Čepička<sup>4</sup> and Vladimír Hampl<sup>1\*</sup>

<sup>1</sup>Department of Parasitology, BIOCEV, Charles University in Prague, Czech Republic.

<sup>2</sup>Department of Plant Systematics & Geography, University of Warsaw, Poland.

<sup>3</sup>Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

<sup>4</sup>Department of Zoology, Charles University in Prague, Czech Republic.

<sup>#</sup>Authors contributed equally to the manuscript

\*Corresponding author: [vlada@natur.cuni.cz](mailto:vlada@natur.cuni.cz)

**Abstract**

Euglenids represent a group of protists with diverse modes of feeding including phagotrophy, osmotrophy, phototrophy and mixotrophy. Until today, only partial genomic sequence of *E. gracilis* and transcriptomes of several phototrophic and secondary osmotrophic species are available, while primarily heterotrophic euglenids are undersampled in this respect. In this work we fill this gap by presenting genomic and transcriptomic drafts of a primary osmotroph *Rhabdomonas costata*. To reduce bacterial contamination, *R. costata* cells were purified prior to gDNA isolation by fluorescence-activated cell sorting (FACS) in combination with laser microdissection and the extracted DNA was subjected to whole genome amplification. The current genomic assembly length 100 Mbp is 14× smaller than is *E. gracilis*, the difference partly caused by incompleteness. It is too fragmented for comprehensive gene prediction, nevertheless the comparison of transcriptomic and genomic data allowed us to estimate features of its introns including several candidates for non-conventional intron types. A set of 39,585 putative *R. costata* proteins was predicted from the transcriptome. Only 26,052 predicted proteins have any homologue in NCBI and 13,130 were annotated by KEGG. 16% of transcripts bear recognizable splice leader sequence at their 5' terminus. Annotation of the mitochondrial core metabolism provides the first data on the facultatively anaerobic mitochondrion of *R. costata*, which in most respects resembles the mitochondrion of *Euglena gracilis* but is better adapted to anaerobiosis. *R. costata* synthesises heme by a mitochondrial-cytoplasmatic C4 pathway with enzymes orthologous to those from *E. gracilis*. This, together with a low (overall) percentage of green algae-affiliated genes supports the ancestrally osmotrophic status of this species.

## Introduction

Euglenids are a relatively species-rich group (~1,500 described species) of unicellular eukaryotes classified together with kinetoplastids, diplomonads and symbionts into the phylum Euglenozoa (Cavalier-Smith 1981). Euglenids are defined by both ultrastructural and molecular features. They possess one or two flagella inserted in a paraflagellar pocket and reinforced by a paraflagellar rod. Their surface is formed by a distinctive pellicle consisting of three layers – the plasma membrane, a proteinaceous belt supported by microtubules, and the vesicles of endoplasmic reticulum. The pellicle enables some euglenids to move characteristically – by undulated shifts in the shape of the cell – which is referred to as metaboly or euglenoid movement (Leander, Esson, and Breglia 2007).

All major types of eukaryotic modes of nutrition are present in Euglenida – phagotrophy, osmotrophy, phototrophy and mixotrophy. Phagotrophic euglenids form several independent clades in the phylogeny of this group, including the deepest branches in the tree (Lax et al. 2019). Phototrophic euglenids (Euglenophyceae) that arose from one of these clades are by all means the best studied subgroup. Their cell contains a triple membrane-bound plastid derived from a secondary endosymbiosis with a green alga. Since Euglenophyceae constitute a monophyletic clade, it is assumed that this endosymbiosis occurred in the exclusive common ancestor of this clade (Leander 2004), although early acquisition of plastid in the Euglenozoa lineage was also proposed to account for plant-like traits in trypanosomatids (Hannaert et al. 2003). Although all known members of Euglenophyceae contain plastids, six species lost the ability to photosynthesise and became secondarily osmotrophic. The best-known case is *Euglena longa* (formerly *Astasia longa*, Pringsheim, 1936) (Marin-Melkonian, 2003).

Primarily osmotrophic euglenids (genera *Rhabdomonas*, *Menoidium*, *Rhabdospira*, *Gyropaigne*, *Parmidium*, *Distigma* and *Astasia*) form another distinct monophyletic clade branching within phagotrophic euglenids (Lax et al. 2019). Members of the order Rhabdomonadales are distinguished from the rest of the osmotrophic clade (*Distigma*, *Astasia*) by the inability of euglenoid movement. The genus *Rhabdomonas*, containing 11 described species, is defined by one emergent flagellum, non-flattened body and shallowly fluted periplast with six to nine steeply helical ridges (Leander et al. 2017). The species *Rhabdomonas costata* (Korshikow, 1928) Pringsheim 1942 differs from other species by a one-to-one ratio of the lengths of the cell and the flagellum, which are about 20-30  $\mu\text{m}$  (Cabała 2003; Busse and Preisfeld 2003; Lax and Simpson 2013). *R. costata* inhabits peaty waters, particularly with decaying leaves and needles; if present then abundant and usually without any other osmotrophic species (Cabała 2003). This species was occasionally found in Europe (e.g. Wołowski, 2011, Cabała 2003), Australia and South America (this study), but is probably cosmopolitan.

The sister groups of Euglenida are poorly studied marine Symbiontida (e.g. *Bihospites*), Diplonemida, that despite being largely unknown have surprised scientists by their massive abundance in ocean waters (Gawryluk et al. 2016; Flegontova et al. 2016), and finally Kinetoplastea, that comprise many infamous parasites, among which species of genera *Trypanosoma* and *Leishmania* are one of the most studied protists. Mitochondrial and nuclear genomes of 40 species and strains of kinetoplastids have been sequenced until present, but outside kinetoplastids, there are no complete and well annotated nuclear genome sequences available for Euglenozoa. Partial genomes of marine diplomonids were obtained by single-cell approaches (Gawryluk et al. 2016). In Euglenida, the genomic studies so far only covered the iconic model flagellate *Euglena gracilis*, still yielding only the mitochondrial genome (Dobáková et al. 2015) and a very fragmented and unannotated nuclear genome (Ebenezer et al. 2019). Quite in contrast, 30 chloroplast genomes of Euglenophyceae have been published until present (Gockel and Hachtel 2000; Hallick et al. 1993; Kasiborski, Bennett, and Linton 2016; Karnkowska, Bennett, and Triemer 2018). Recently, proteomes of *E. gracilis* plastid (Novák Vanclová et al. 2020) and mitochondria (Hammond et al. 2020) have been revealed by mass spectrometry proteomics.

Unlike their phototrophic relatives, heterotrophic euglenids have drawn substantially less attention with all focusing only on taxonomy, morphology and phylogeny. Here we start filling this major gap by presenting the genomic and transcriptomic draft of the primarily osmotrophic *Rhabdomonas costata*. We use the data to describe some of its features – genome size, intron types, frequency of splice leaders, metabolism of its mitochondrion and heme synthesis.

## Methods

### *Strain origin and cultivation*

*Rhabdomonas costata* strain PANT2 was isolated from a freshwater body in Pantanal (Paraguay) and grown with a non-characterised mixture of bacteria in a standard 802 Sonnenborn's Paramecium medium (pH = 7.4) at room temperature and inoculated every 3 to 4 weeks. For the purpose of this project we have prepared by serial dilution a clonal lineage that was then used for preparation of DNA and RNA.

### *Microscopy*

The DIC microscopy was performed on live culture using Olympus BX51 microscope with Olympus DP70 camera. For the scanning electron microscopy, the pelleted cells were dropped on a filter paper and fixed using 2.5% glutaraldehyde in 0.1M cacodylate buffer for 24 hours. The further processing was done by service lab. The samples were observed using the JEOL JSM-6380LV microscope (JEOL, Akishima, Japan). For transmission electron microscopy, the

pelleted cells were fixed using 2.5% glutaraldehyde in 0.1M cacodylate buffer, postfixed with OsO<sub>4</sub> and the ultrathin sections were contrasted by uranyl acetate and observed using JEOL JEM-1011 microscope.

#### *Isolation of DNA*

We used two methods to obtain high-quality non-contaminated *R. costata* DNA. (1) The cells in the culture were sorted by FACS with the value of drop diameter equal to 70  $\mu\text{m}$  (length of the cell is around 25  $\mu\text{m}$ ). Around 90,000 positive drops, most of them containing a single cell of *R. costata*, were collected and used for DNA extraction. (2) By combination of FACS and laser microdissection in which 69 drops from FACS were subsequently used for microdissection of individual cells, from which DNA was extracted. Both samples were then subjected to whole genome amplification (WGA) (Sigma-Aldrich WGA4-10Rxn) in order to increase the DNA amount. After amplification, the samples contained 27.7  $\mu\text{g}$  and 32.0  $\mu\text{g}$  of DNA, respectively. PCR using general prokaryotic primers for 16S rRNA have produced a specific product when sample 1 was used as a template, but not product with sample 2 (not shown). This supported our expectation that while sample 1 is still contaminated by bacterial DNA, sample 2 is likely contamination-free.

#### *Genome sequencing and assembly*

Both samples were sequenced with Illumina. Initially, the two MiSeq runs of sample 2 (l=250 bp) produced 3.9 Gbp of sequences. Unfortunately, we were not able to produce a reasonable assembly from these data. Therefore, we have sequenced also the sample 1 in using HiSeq (l=100 bp) and obtained 8.2 Gbp of data. The raw reads from sequencing of both samples were assembled by SPAdes v3.7.0 into 143,763 contigs (36,105 contigs longer than 1,000 bp).

#### *RNA isolation, transcriptome sequencing and assembly*

RNA was isolated three times using three slightly different approaches from three different specimen of our *R. costata* clone. 1) Direct isolation of mRNA (Dynabeads mRNA Direct Kit, including polyA-selection; Thermo Fisher Scientific, Waltham, MA, USA) from which the library was prepared, including a polyA-selection step; 2) Isolation of total RNA (GeneAll Hybrid-R RNA purification kit; GeneAll Biotechnology, Seoul, South Korea) followed by mRNA purification (Dynabeads mRNA Direct Kit) from which the library was prepared, including a polyA-selection step; 3) Isolation of total RNA (GeneAll Hybrid-R RNA purification kit) and from 4  $\mu\text{g}$  of the total RNA, the library was prepared according to the standard TruSeq Stranded mRNA Sample Preparation Guide including a polyA-selection step. Unlike the former, the last approach contained only one step of polyA selection. The total RNA was always quantified by Quantus Fluorometer (Promega, Madison, WI, USA) and its quality was checked on an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA). All three

samples were sequenced on an Illumina MiSeq instrument (Illumina, San Diego, CA, USA) using 150 base-length read chemistry in the paired-end mode. As no marked differences in data completeness among the libraries were observed, the 3 libraries were assembled together with Trinity v2.0.6 using default parameters and 93,852 contigs were created. We used Transdecoder for basal protein prediction and obtained 55,783 putative proteins. Online version of software package cd-hit v.6 (Li et al., 2001) was applied to avoid redundancy and yielded 39,585 non-redundant amino acid sequences. The transcriptomic and genomic reads and assemblies are available in GenBank under the BioProject ID PRJNA550357.

#### *Prediction of introns*

Introns were automatically predicted by mapping the assembled transcripts to genome contigs using Exonerate (version: est2genome 2.2.0). Introns with predicted non-conventional boundaries were manually inspected. In selected genes, the introns were inspected manually by aligning transcripts to genome contigs. The position of these introns was compared to other euglenids.

#### *Phylogenetic ancestry of R. costata proteins*

The ancestry of proteins was assessed with the same methodology as in Ebenezer et al (Ebenezer et al. 2019). Briefly, homologues with e-value  $< 10^{-2}$  were retrieved from a custom database, aligned by MAFFT 7.273 (Katoh and Standley 2013) (default parameters) and trimmed in trimAl 1.2 (Capella-Gutiérrez et al. 2009) (default parameters). 13,696 alignments with more than 3 taxa and longer than 74 amino acid residues were used for tree reconstruction in RAxML (Stamatakis 2014) (v8.1.17; 100 rapid bootstraps) in Metacentrum (The National Grid Infrastructure in the Czech Republic). Custom scripts (Python 3.4) were used to sort the trees into bins based on the taxonomic affiliation of the clan in which *R. costata* branched. In 3,445 cases, the tree was included in a taxonomic bin, because it contained a bipartition supported by bootstrap 75% and higher comprised of *R. costata* and members of one defined taxonomic group only.

#### *Prediction of mitochondrial proteome*

The proteome of the mitochondrion was predicted using the following procedure. The complete set of proteins predicted from the transcriptome was BLAST-searched against the Mitominer database and 7,985 proteins with e-value  $< 10^{-4}$  were selected. To lower the redundancy, only one *R. costata* protein per one particular Mitominer-database-hit was kept yielding a set of 1,501 proteins. For every protein, the probability of mitochondrial targeting was predicted using TargetP and Mitoprot tools. Only proteins with probability equal or higher than 0.5 in one or both tools were kept producing 347 candidates. This initial set was enriched by 1,149 *R. costata* orthologues of *Euglena gracilis* proteins enriched in the mitochondrial fraction



(Hammond et al., 2020, Suppl. Table 4, sheet identifying strategies, column B). The orthology was established OrthoMCL (Li, Stoeckert, and Roos 2003) software package as described in Soukal et al. (in preparation). For every gene, the presence of a spliced leader (SL) at the 5' terminus was established. 162 probable contaminants were removed from the list that lacked a SL and at the same time had a prokaryotic best hit in the local database. To each entry, the KEGG ID was assigned using the single-directional best hit method in KAAS (Moriya et al. 2007) or transferred from *E. gracilis* homologue annotation, and the best BLAST hit was identified using blastp against the NCBI nr database. During the manual annotation process, 223 additional proteins were included into the set based on metabolic or molecular function (e.g. enzymes, membrane carriers, protein translocases). The final predicted mitoproteome consists of 1,557 proteins.

The conspicuous absence of alternative oxidase in the data set was verified by gene-specific PCR using degenerate primers based on sequence information from *E. gracilis* (forw. primer: GARGARGCNGARAAYGARAGRATGCA; rev. primer GCRAANGTRTGRTTNACRTSNCGRTG) and *E. gracilis* and *Eutreptiella gymnastica* gDNAs as positive controls.

### *Phylogenies*

Phylogenetic trees of proteins of interest were generated by the following procedure. *Rhabdomonas costata* protein(s) and their homologues were downloaded from databases. In the case of tetrapyrrole biosynthesis pathway, the dataset of (Lakey and Triemer 2016) were used as the seed and enriched by *E. gracilis* and *R. costata* transcripts and their best NCBI blast hits and the best hits from the local database. All entries were aligned by MAFFT (<https://mafft.cbrc.jp/alignment/server/>; (Kato, Rozewicki, and Yamada 2019)) using the automated strategy, trimmed in BMGE version 1.12 (Criscuolo and Gribaldo 2010), manually inspected for mis-aligned positions, and the phylogeny including 1000 ultrafast bootstraps was inferred in IQ-Tree 2.0 (Nguyen et al. 2015) using the model TIM2e+G4, and visualised in FigTree 1.4.4.

## **Results**

### *Light and electron microscopy and phylogenetic position*

The cells were initially observed under the light and electron microscopy (Figure 1). They possessed two flagella inserted in the flagellar pocket but only one extending beyond. The surface of the cell was shaped into 7-9 ridges supported by pellicle and microtubules. Conspicuous paramylon grains were observed in the cytoplasm, along with many mitochondria with developed discoidal cristae. Nucleus contained a large nucleolus and multiple

heterochromatin regions. Based on the microscopic observations and the phylogenetic analysis of the gene for 18S rRNA (Figure 2) the organism was identified as *Rhabdomonas costata*.

#### *General characteristics of the genome and transcriptome*

The basic measures of *R. costata* genome assembly is given in Table 1. The quality of the genome draft is low. The assembly is very fragmented: 36,105 contigs above 1 kb with N50 being 1,194 and the average coverage being  $\sim 112\times$ . Total length of the assembly is 106.9 Mbp. Due to the extreme fragmentation of the genomic draft, we have not proceeded with gene prediction.

The transcriptome sequencing resulted in 39,585 non-redundant protein predictions. The completeness of the transcriptome measured by BUSCO (complete BUSCOs: 76.0%, fragmented BUSCOs: 12.5%, missing BUSCOs: 11.5%, n:303) was satisfying, suggesting that this data set is sufficient for the description of selected features of this organism. The basic characteristics of the transcriptome are given in Table 2. The proteins were automatically annotated by BLAST against NCBI nr and only 26,052 proteins have at least one good homologue ( $e\text{-value} < 10^{-5}$ ). The taxonomic affiliations of *R. costata* proteins are summarised in Figure 3. As expected, most proteins out of 3,445 where the relationship could be robustly established (bootstrap 75 % and higher), affiliated with Ecavata (56%), out of which 53 %, 49 %, 5 % and 1 % are specifically affiliated with Discoba, Kinetoplastea, Heterolobosea and Metamonada, respectively. 15 % of proteins branched with prokaryotes and less than 30 % affiliated with any of the other bins that included e.g. Viridiplantae (5,6 %), Alveolata (4,2 %), Amoebozoa (3,9 %) and others. The proteins affiliated with prokaryotes may represent contamination, but some may be *bona fide* *R. costata* proteins having strong affiliation to prokaryotic homologues due to increased divergence, absence of eukaryotic homologues in the nr database or origin via horizontal gene transfer. The KEGG analyses of the set of 26,052 putative proteins led to 13,130 KEGG annotations. Summary of individual KEGG categories of annotated proteins is given in Figure 4.

All 93,852 transcripts were searched for the presence of the published sequence of *R. costata* spliced leader sequences (ACATTACTGGGTGTAATCATTTTTTCG) (Kuo et al., 2013). Only 16.4 % (15,384) of transcripts contained the 10-nt partial SL (CATTTTTTCG) or a longer fraction of the full-length SL.

#### *Intron characteristics*

Although it was not suitable for gene prediction, we used the genomic assembly for an analysis of introns' types. The introns were detected by mapping of transcripts to genomic contigs. First, this mapping was done manually in the cases of genes, for which the presence of introns was already reported in other euglenids – tubulin alpha (tubA), tubulin beta (tubB), tubulin gama

(tubG), heat shock protein 90 (hsp90), glyceraldehyde phosphate dehydrogenase (gapC) and fibrillarin (nop1p) (Henze et al. 1995; Ebel et al. 1999; Canaday et al. 2001; Milanowski et al. 2014; Milanowski et al. 2016). The information on introns are summarised in the Table 3. We have identified 29 complete introns in these six genes. Some gene regions were not covered by transcripts, and so the presence of additional introns in these genes cannot be excluded. All complete detected introns have conventional GT/AG boundaries. In 14 introns, only one end was found in the data and these were marked as incomplete. The type of these introns could not be determined with certainty but all of them have at least one end with conventional boundaries.

The positions of intron in two best studies genes in this respect (tubA and tubB) were compared with their homologues in other euglenids (Figure 5). All *R. costata* introns were in the same positions as those described by (Milanowski et al. 2014) in tubulin genes of the primary heterotroph *Menoidium bibacillatum*. The second conventional intron in tubA of the two heterotrophs is in identical position as the intron in phototrophic euglenids. Positions of all other introns of heterotrophs and phototrophs are different. Only one of the 15 introns in hsp90 is in the same position as the intron in *Euglena agilis* hsp90, though on the other hand, the gene for *R. costata* fibrillarin shares 3 of the 4 intron sites with the fibrillarin of *E. gracilis* (not shown).

Secondly, we have predicted the introns exhaustively by mapping the transcriptome to the genome assembly and the output was searched for the regions with potential introns. 11,418 putative introns were detected (Table S-introns), out of which 1,164 had non-conventional boundaries (not GT/AG). These were manually inspected. The manual inspection revealed seven cases of putative non-conventional introns with boundaries confirmed by transcript and reads mapping (Table S-introns) of which four are very likely non-conventional introns as no alternative transcripts have been observed. One example intron in the hypothetical protein encoded in the genomic contig NODE\_718 and the transcript TR27401 is shown in the Figure 6.

#### *Mitochondrial proteome*

We used the set of proteins predicted from transcriptome to determine *in silico* the mitochondrial proteome of this primarily osmotrophic euglenid. *R. costata* grows in a polyxenic culture at the bottom of stationary 15 ml tubes, indicating that it thrives in low-oxygen environment, in spite of containing numerous mitochondria with well-developed cristae (Figure 1). In our experience, it was able to survive only short-term (one month) cultivation in complete anaerobiosis after which it dies.

From the 1,557 proteins of the predicted mitochondrial proteome, 1,081 were assigned to categories adopted from KEGG (Table S-mitoproteome, Figure 8). 1,316 proteins have

orthologues in the experimentally established and manually curated proteome of the *E. gracilis* mitochondrion (Hammond et al. 2020; Novák Vanclová et al. 2020). Contigs containing sequences homologous to all seven protein-coding genes reported from the *E. gracilis* mitochondrial genome (Dobáková et al. 2015) were detected in the genome assembly (Table S-mitogenes). Given that in all cases the contigs contained only a single gene or its part, it could not be established whether they originate from mitochondrial or nuclear genome. In addition, most of the coding sequences were interrupted by multiple stop codons indicating that they are pseudogenes. Although we have not detected any contigs that could be confidently considered parts of the mitochondrial genome, the presence of this genome is expected as the predicted mitoproteome set contains a repertoire of over 140 proteins involved in DNA and RNA metabolism, ribosome biogenesis and translation (Table S-mitoproteome). These include two of the already published DNA polymerases I (Harada et al. 2020).

Predicted functions and metabolic pathways of the mitochondrion are summarised in the Figure 8. Pyruvate and malate are probable substrates for the energy metabolism. Malic enzyme (RCo00811 and 12435) catalyzes oxidative decarboxylation of malate to pyruvate. Pyruvate:NADH oxidoreductase (PNO) present in five transcripts that are not identical in sequence (RCo08351, 08357, 21209, 32638, and 52779) is the only enzyme for the oxidative decarboxylation of pyruvate to acetyl-CoA detected in the transcriptome. The canonical mitochondrial pyruvate dehydrogenase complex is apparently absent with only the E3 subunit (dihydrolipoamide dehydrogenase, RCo010764) recovered in the data set (transcript RCo00646 for subunit E1 is likely a contamination). However, the E3 subunit is also a component of other mitochondrial enzyme complexes and in *R. costata* it is probably involved in the glycine cleavage system. Pyruvate can be alternatively reduced to L-/D-lactate by lactate dehydrogenases (RCo06066, 11296, 18299, 22174, 38617, 38618). Gluconeogenic enzymes pyruvate carboxylase (RCo04329) and phosphoenolpyruvate carboxykinase (RCo24588, RCo25226) are present. The set contains also three similar copies of 1,3-beta-D-glucan synthase (RCo03309, 03310, 22891). This enzyme is orthologous to the *E. gracilis* glucan synthase-like 2 protein responsible for the synthesis of paramylon (Tanaka et al. 2017) and reported in the mitochondrial proteome of *E. gracilis* (EG\_transcript\_73, Hammond et al. 2020). Paramylon grains are present in the cytoplasm of *R. costata* (Figure 1). Unlike *E. gracilis* proteins, probability of mitochondrial localisation of glucan synthase-like 2 in *R. costata* based on the prediction software is low, and so we consider it putatively cytosolic. Enzyme endo-1,3(4)-beta-glucanase potentially involved in the degradation of paramylon is also present (RCo17523, 19147, 47076, 17521, 18319 and 23166) and *in silico* predicted to localise in the cytosol.

The TCA cycle seems complete. The alpha-ketoglutarate dehydrogenase complex is absent, and this step is bypassed by enzymes alpha-ketoglutarate decarboxylase (RCo09834) and succinate-semialdehyde dehydrogenase (RCo25294). Both subunits of succinyl-CoA synthetase are present (RCo22592, RCo42191) and this enzyme probably works in the direction of succinyl-CoA formation, thus leading to the fatty acid synthesis. GABA shunt is present but the glyoxylate cycle enzymes are absent. Succinate dehydrogenase (SDH, Complex II) is similar in subunit composition to kinetoplastid SDH (Morales et al. 2009). Six subunits have been found in total, including the conserved eukaryotic subunits SDH1-2 (SDHA-B) and Euglenozoa-specific subunits 6-9. The FeS cluster-containing subunit SDH2 splits in two polypeptides (N- and C- terminus), similarly to trypanosomes and *E. gracilis* (Gawryluk and Gray 2009). The Euglenozoa-specific subunits SDH5, 10 and 11 are missing in the data and are missing also from the *E. gracilis* transcriptome. Other components of the respiratory chain detected in the transcriptome are complex I (21 subunits including the 3 Euglenozoa-specific), complex III (5 subunits), complex IV (8 subunits) and FoF1 ATPase (subunits of F1 part  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  and OSCP and three Euglenozoa-specific) and electron-transferring-flavoprotein dehydrogenase (RCo00411 and 00414). Alternative oxidase, reported from other Euglenozoa, is absent from the transcriptome and was not detected by PCR with specific primers. Three types of membrane-associated electron carriers are present, cytochrome c, ubiquinone (UQ; most enzymes involved in its synthesis are present in the transcriptome) and rholiquinone (RQ), which is formed from ubiquinone by rholiquinone methyltransferase (RCo43299, 43301, 43302 and 43305). Soluble electron-transferring flavoprotein (ETF, RCo28350, 35495 and 11455) may serve as electron donor for fatty acid metabolism.

The ability to synthesize RQ provides *R. costata* the opportunity to transfer electrons from Complex I via Complex II to fumarate, but the same reaction can be performed by FAD-dependent fumarate reductase (RCo13263, 47402, 47403 and 48726) which uses Q for fumarate reduction. The produced succinate is the substrate of succinyl-CoA synthetase producing succinyl-CoA, which may enter the synthesis of wax esters described in *E. gracilis*. *R. costata* contains all enzymes needed for this process. Propionyl-CoA, the first committed substrate for wax ester synthesis, is formed by methylmalonyl-CoA mutase (RCo28046), methylmalonyl-CoA epimerase (RCo48567), and propionyl-CoA carboxylase  $\alpha$  and  $\beta$  (RCo05823 and RCo14953). The condensation of propionyl-CoA and acetyl-CoA can be in principle catalyzed by acetyl-CoA acyltransferase (RCo38684, 11994, 24188, 31732 and 35370) instead of the missing alpha-ketoacyl synthase. *R. costata* contains 3-hydroxyacyl-CoA dehydrogenase (RCo47669) and enoyl-CoA hydratase (RCo02785, 16455, 30891) and enzymes for the reduction of trans-enoyl-CoA: trans-2-enoyl-CoA reductase (RCo48015, 48574), acyl-CoA dehydrogenase (RCo29909) and ETF that could provide electrons. The pathway further

proceeds outside the mitochondrion, where carnitine O-palmitoyltransferases 1 a 2 (RCo04453, 08998, 16397, 33121) export the acyl-CoA. Neither fatty acyl-CoA reductase (ER enzyme) nor wax ester synthase (cytosolic enzyme) were detected, however, 34 transcripts encoding ester synthases/diacylglycerol acyltransferases (WSD), which are used for the wax ester synthesis, were detected (Figure S-wsdtree).

The organelle may be able to import sulphate via a putative transporter (RCo08807) although the identity of this protein is uncertain. It also contains enzymes needed for sulphate activation, sulphate adenylyltransferase (RCo49652) and adenylylsulphate kinase (CysC) (RCo11997), to produce phosphoadenosyl-5'-phosphosulphate (PAPS). The enzymes metabolising inorganic sulphur compounds thiosulphate sulphur transferase and sulphite oxidase are present (RCo00455, RCo10220), however, the production of sulphide is undetected. Still the transcripts of the sulphide-dependent enzyme cysteine synthase (RCo09036, 16104, 16108, 13733 and 51020) are present as is the L-serine producing serine O-acetyltransferase (RCo00499, 28429 and 42343). The mitochondrion also contains a rich set of enzymes for the early and late synthesis of FeS clusters including the export system from the mitochondrion.

The predicted mitochondrial proteome contains enzymes involved in the synthesis and metabolism of 10 proteinogenic amino acids (S, C, T, G, A, V, L, I, Q and P). A complete glycine cleavage system and serine/glycine hydroxymethyl transferase involved in folate one carbon pool are also present. The set contains 136 entries involved in metabolite and ion transport across membranes, out of which 31 are ABC transporters (including Atm1 involved in FeS cluster export) and 80 are mitochondrial carrier family proteins (solute carrier family 25) that cluster into 71 distinct clades (Figure S-MCFtree).

More than 30 transcripts of proteins putatively involved in protein transport and maturation have been detected. These encode four outer membrane proteins – the translocation pore Tom40 (RCo05874), its insertase Tob55 (RCo35243 and 50551) and Euglenozoa specific proteins Atom69 (RCo12450 and 12448) and Atom46 (RCo30973), two distinct homologues of small Tim (RCo00369 and 22175) localised in the intermembrane space, and eight proteins localised in the inner membrane – inner membrane protease subunit 2 (RCo35651), Tim22 (RCo35157), Tim17 (RCo03131 and 18944), Tim44 (RCo35673), Tim16 (RCo40212), Pam16 (RCo35852) and Oxal (RCo 02990, 23198 and 30475). Homologues of both subunits of mitochondrial processing peptidase were detected (RCo00876 and 39561), however the beta subunit may also be a part of Complex III. Soluble chaperones Mge1 (RCo18360), Hsp70 and Hsp60 (RCo05843) are present.

*Tetrapyrrole synthesis pathways*

We found transcripts for the full set of heme biosynthesis enzymes in the transcriptome (Figure 9). All enzymes form monophyletic clades with *E. gracilis* mitochondrial-cytosolic C4 pathway enzymes with various statistical support and were most likely present in their common ancestor. Mitochondrial 5-aminolevulinic synthase (ALAS) homologs branch together within the eukaryotic clade and sister to alpha-proteobacteria suggesting a mitochondrial origin (RCo53079). Although ALAS should localise in the mitochondrion, *in silico* prediction of this localisation is weak. Following four steps take place in cytosol and the pathway ends in the mitochondrion. Porphobilinogen synthase (ALAD) (RCo46560) and porphobilinogen deaminase (PBGD) (RCo16092) homologs are closely related to homologs of the other bacterivorous euglenid *Distigma* sp. and together with the cytosolic isoforms of *E. gracilis* branch within eukaryotic genes. *R. costata* uroporphyrinogen synthase (UROS) (RCo31923) branches together with other photosynthetic euglenids and the group seems to be sister to oomycetes (FigureS-UROStree). Although this protein was annotated as a plastidial form in *E. gracilis*, it is probably cytosolic or dual-localised in both cytosol and plastid. The cytosolic localisation is supported by the presence of a second, so far unnoticed and putatively plastidial UROS homolog in transcriptome of *E. gracilis* that is robustly branching with green algae and cyanobacteria though lacking a clear plastidial targeting signal. Next enzyme, uroporphyrinogen decarboxylase (UROD), has three isoforms in *E. gracilis* and *E. gymnastica*. One of them is plastidial and does not originate from green algae but more likely from cryptophytes. The other two are probably cytosolic and branching within the eukaryotic clade. These isoforms seem to originate from an ancient gene duplication and *R. costata* has only one of these duplicated genes (RCo52619).

The most complicated situation is in the case of coproporphyrinogen oxidase (CPOX) (FigureS-CPOXtree). *E. gracilis* has at least eight isoforms. Five of them are plastidial oxygen-dependent CPOX (HemF) with no homolog found in *R. costata*. Three other isoforms of *E. gracilis* are oxygen-independent CPOX (HemN) of different origins. Two of these enzymes form branches both with close *R. costata* (RCo19415, 29168) and *E. gracilis* homologs that branch within the deltaproteobacterial clade and were probably gained via HGT. These two isoforms have a clear mitochondrial targeting signal. The third one is more divergent and the *R. costata* homolog (RCo45123) is not closely related to *E. gracilis* but rather to alphaproteobacteria. A putatively mitochondrial isoform of protoporphyrinogen oxidase (PPOX) (RCo26828) is branching within a well-supported eukaryotic clade together with *Distigma* and *E. gracilis*, but the targeting prediction is weak. This contrasts with *E. gracilis* and *Eutreptiella gymnastica* PPOX plastid isoforms, which are placed within plastid genes with “chromalveolate” origin. Lastly, the mitochondrial isoform of ferrochelatase FECH is not derived from eukaryotes; instead, all

euglenid sequences including the *R. costata* homolog (RCo12206) branch together with *Naegleria* within bacterial clades and were probably obtained by HGT. Consistently with the absence of plastid there are no traces of the plastidial C5 pathway.

## Discussion

In this work we contribute to knowledge about non-photosynthetic euglenids by presenting draft genome and transcriptome assemblies of *R. costata*. Although the genome assembly is very fragmented, we demonstrate its usefulness by deducing novel information about the intron composition in *R. costata*. We believe that the fragmentation is not caused by insufficient sequence depth, but rather is an artefact of uneven whole genome amplification (WGA) in the DNA sample preparation. Other reasons for extreme fragmentation (e.g. presence of repeats) cannot be excluded at this point. Similarly to *E. gracilis*, there is a GC bias between the transcriptome and the genome, 58 vs 51% respectively. This may suggest the presence of significant AT-rich repeats in the genome or simply reflect the GC content difference between coding and non-coding regions of the genome. Total length of the assembly is slightly over 100 Mbp, which is an order of magnitude less than the total length of the genome assembly of *E. gracilis*. Its haploid genome size is estimated to be 500 Mbp while the total assembly size is over 1.4 Gbp (Ebenezer et al. 2019).

In comparison, the quality of the transcriptome data set was sufficient for functional annotation. The number of non-redundant predicted proteins (39,585) is comparable to *Euglena gracilis* (36,526) (Ebenezer et al. 2019). SL sequences longer than 9 nt were detected at the 5' end in less than 17 % of transcripts. The absence of the SL suggests substantial truncation at the 5' terminus; however, many transcripts lack a SL naturally (Záhonová et al. 2018).

As almost all eukaryotic genomes, also the genome of *Euglena gracilis* contains introns that are removed by the ribonucleoprotein complex spliceosome. These conventional introns have the consensus sequence GT/AG at the ends, and they are excised by two consequential transesterifications. They were described from genes encoding various proteins, some of which are targeted to chloroplast: fibrillar (Breckenridge et al. 1999; Russell et al. 2005), tubulin alpha (tubA), beta (tubB) and gamma (Ebel et al. 1999; Canaday et al. 2001; Milanowski et al. 2014), heat shock protein 90 (hsp90), glyceraldehyde-3-phosphate dehydrogenase (gapC) (Henze et al. 1995; Milanowski et al. 2016), enolase, glyceraldehyde-3-phosphate dehydrogenase A (gapA), porphobilinogen deaminase (pbgD), ubiquinol-cytochrome *c* reductase iron-sulphur subunit (petA), photosystem I reaction centre subunit III (psaF), photosystem II reaction centre protein M (psbM), and photosystem II reaction centre W protein (psbW) (Vesteg et al. 2010). However, the genes of euglenids contain also other types of introns



without the canonical GT/AG boundaries. These non-conventional introns have variable length and no clear pattern of nucleotide sequence at the exon/intron junction. They form the stem-loop RNA structures and their excision is independent of spliceosome and takes place later than the excision of spliceosomal introns (Gumińska et al. 2018; Milanowski et al. 2014). This type of introns was found in *tubA*, *tubB*, *gapC* and *hsp90*. Except these main types of introns so-called “intermediated” introns that combine the features of both types were reported (Rafal Milanowski et al. 2014). The presence of non-conventional introns would complicate the gene predictions.

The genome of *R. costata* seems to be relatively intron rich. We have confirmed the presence of introns in all genes, for which the introns have been investigated in other euglenids (Henze et al. 1995; Breckenridge et al. 1999; Russell et al. 2005; Ebel et al. 1999; Canaday et al. 2001; Milanowski et al. 2014) and our automatic search for introns in the fragmented genomic draft revealed thousands of putative introns. Seven of these introns are likely non-conventional. The presence of non-conventional introns in *R. costata* is expected as they were reported from several euglenid species and recently also from a marine diplomonid (Gawryluk et al. 2016).

The phylogenetic affiliation of *R. costata* predicted proteins is similar to that assessed in *E. gracilis* (Ebenezer et al. 2019) with the notable but expected difference in the fraction of genes affiliated to Viridiplantae. This is much higher in *E. gracilis* than in *R. costata* (14% as compared to 5%) and reflects the endosymbiotic history of *E. gracilis*, which acquired the chloroplast from green alga. The *in silico* predicted proteome of *R. costata* mitochondrion is smaller in size to the experimental proteome of *Euglena gracilis* (1,557 in *Rhabdomonas* vs approx. 2,500 in *E. gracilis*, Hammond et al. 2020).

The mitochondrion of *E. gracilis* bears a unique combination of metabolic features. It contains a set of enzymes for the facultatively anaerobic metabolism, which in the presence of oxygen metabolises pyruvate or malate by pyruvate dehydrogenase complex, slightly modified TCA cycle and the full set of respiratory complexes including alternative oxidase described also from kinetoplastids. In the absence of oxygen the pyruvate is oxidatively decarboxylated by pyruvate:NADH oxidoreductase and the mitochondrial NADH is recycled by respiratory complex I and rhodoquinone-dependent fumarate reductase producing succinate and further propionyl-CoA. The latter is condensed with acetyl-CoA into fatty acids and wax esters stored in the cytoplasm in high concentrations (Tucci et al. 2010). These are recycled under aerobic conditions for the ATP production or for the synthesis of organic compounds thanks to functional glyoxylate cycle uniquely localised in the mitochondrion (Müller et al. 2012).

The biochemistry of the *Rhabdomonas* mitochondrion resembles in several aspects functioning of *E. gracilis* organelle but seems more streamlined. It contains a complete TCA cycle with euglenid-specific bypass of alpha-ketoglutarate decarboxylase and succinate-semialdehyde dehydrogenase. It also uses rhodoquinone to reverse electron flow in the truncated electron transport chain under low oxygen condition to fumarate as the final electron acceptor and produced succinate is then consumed during the synthesis of wax monoesters as in *E. gracilis*. Unlike *E. gracilis*, *R. costata* mitochondrion uses only PNO for oxidative decarboxylation of pyruvate and it does not contain alternative oxidase in its ETC, which is consistent with our inability to amplify these transcripts by PCR. Intriguingly, *R. costata* mitochondrion does not contain enzymes of the glyoxylate cycle, a shortcut of the TCA cycle, which is used to generate four-carbon molecules for anabolic reactions from acetyl-CoA released after degradation of lipids and wax esters. How this is solved in *R. costata* remains to be elucidated.

Interesting difference from the situation in *E. gracilis* concerns the sulphate metabolism. *E. gracilis* is apparently capable to assimilate sulphate into cysteine in the mitochondrion (Hammond et al. 2020; Saidha et al. 1985) and the sulphite reductase has been detected also in its chloroplast fraction (Novák Vanclová et al. 2020). In contrary, *R. costata* can only activate sulphate to the form of PAPS, an important coenzyme in sulphotransferase reactions, but the internal production of sulphide from PAPS is impossible due to the absence of adenylyl-sulphate and sulphite reductases. Still, the presence of sulphide-dependent cysteine synthase suggests that *R. costata* may be able to synthesize cysteine from sulphide putatively produced via another reaction or taken up from the anaerobic environment.

Heme ranks among important cofactors in the cellular metabolism because it is involved in many key biochemical processes. In most heterotrophic eukaryotes, heme synthesis C4 pathway involves eight steps that are localized partially in the mitochondrion and partially in cytosol. The first step is the condensation of glycine and succinyl-CoA into 5-amino-levulinate by ALAS in the mitochondrial matrix. Next four (or five) steps take place in cytosol and the pathway ends in the mitochondrion with two reactions that take place in the intermembrane space and the final reaction ongoing in the mitochondrial matrix (Hamza and Dailey 2012). In most eukaryotes with a plastid, the alternative C5 pathway is present, and all steps are localized in the plastid. 5-amino-levulinate is here synthesised from glutamic acid by three consecutive enzymes: glutamyl-tRNA synthetase (GltX), glutamyltRNA reductase (GTR), and glutamate-1-semialdehyde 2,1-aminomutase (GSA-AT). The pathway then follows the same steps like the classical pathway but is localised in the plastid and catalysed by enzymes of mostly cyanobacterial origin (Kořený et al. 2011; Cenci et al. 2016). *E. gracilis* and *Bigelowiella natans* (algae with a complex plastid) have unexpectedly both pathways and

mosaic evolutionary origin of their enzymes reflects complex evolutionary history of these eukaryotes (Kořený et al. 2011; Woo et al. 2015; Cihlář et al. 2016; Lakey and Triemer, 2017).

The *R. costata* transcriptome contains a complete set of enzymes of the mitochondrial/cytosolic C4 pathway that are orthologous to the *E. gracilis* enzymes. The comparison of these two euglenids helped to reveal that one UROS and two UROD isoforms of *E. gracilis* examined so far are probably cytosolic C4 pathway enzymes as they all have *R. costata* orthologs. Interestingly, we have recovered an unnoticed UROS homolog from the *E. gracilis* transcriptome related to green algae and putatively involved in the C5 pathway in its plastid. While several plastidial isoforms of the oxygen-dependent CPOX (HemF) in *E. gracilis* has two distinct origins in eukaryotic kingdom the newly found isoforms of the oxygen independent CPOX (HemN) that function in mitochondrion have deltaproteobacterial origin. The complete absence of the plastidial C5 pathway for the tetrapyrrole synthesis together with the overall limited number of transcripts affiliated to green algae is consistent with the absence of chloroplast in *R. costata* and suggests that the osmotrophic lifestyle of this euglenid is the primary state.

### Acknowledgment

The project was financially supported by the Czech Science Foundation project nr. 13-24983S and by the project of the Ministry of Education, Youth and Sports of CR within the National Sustainability Program II (Project BIOCEV-FAR) LQ1604 and by the project “BIOCEV” (CZ.1.05/1.1.00/02.0109). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated. The authors would like to thank Martina Johnson Pokorná for help with FACS sorting and laser microdissection.

### References:

- Breckenridge, David G., Yoh-ichi Watanabe, Spencer J. Greenwood, Michael W. Gray, and Murray N. Schnare. 1999. “U1 Small Nuclear RNA and Spliceosomal Introns in *Euglena Gracilis*.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (3): 852–56. <https://doi.org/10.1073/pnas.96.3.852>.
- Busse, Ingo, and Angelika Preisfeld. 2003. “Systematics of Primary Osmotrophic Euglenids: A Molecular Approach to the Phylogeny of *Distigma* and *Astasia* (Euglenozoa).” *International Journal of Systematic and Evolutionary Microbiology* 53 (Pt 2): 617–24. <https://doi.org/10.1099/ijs.0.02295-0>.

- Cabała, Jolanta. 2003. "Some Interesting Colourless Euglenophytes Found in Southern Poland." *Acta Societatis Botanicorum Poloniae* 72 (3): 243–47. <https://doi.org/10.5586/asbp.2003.032>.
- Canaday, J, L H Tessier, P Imbault, and François Paulus. 2001. "Analysis of *Euglena Gracilis* Alpha-, Beta- and Gamma-Tubulin Genes: Introns and Pre-mRNA Maturation." *Molecular Genetics and Genomics: MGG* 265 (1): 153–60. <https://doi.org/10.1007/s004380000403>.
- Capella-Gutiérrez Salvador, José M. Silla-Martínez, Toni Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972-1973. doi:10.1093/bioinformatics/btp348.
- Cavalier-Smith, Thomas. 1981. "Eukaryote Kingdoms: Seven or Nine?" *BioSystems* 14 (3–4): 461–81. [https://doi.org/10.1016/0303-2647\(81\)90050-2](https://doi.org/10.1016/0303-2647(81)90050-2).
- Cenci, Ugo, Daniel Moog, Bruce A. Curtis, Goro Tanifuji, Laura Eme, Julius Lukeš, and John M. Archibald. 2016. "Heme Pathway Evolution in Kinetoplastid Protists." *BMC Evolutionary Biology* 16 (1): 1–18. <https://doi.org/10.1186/s12862-016-0664-6>.
- Cihlář, Jaromír, Zoltán Füßy, Aleš Horák, and Miroslav Oborník. 2016. "Evolution of the Tetrapyrrole Biosynthetic Pathway in Secondary Algae: Conservation, Redundancy and Replacement." *PLoS ONE* 11 (11). <https://doi.org/10.1371/journal.pone.0166338>.
- Criscuolo, Alexis, and Simonetta Gribaldo. 2010. "BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments." *BMC Evolutionary Biology* 10 (1): 210. <https://doi.org/10.1186/1471-2148-10-210>.
- Dobáková, Eva, Pavel Flegontov, Tomáš Skalický, and Julius Lukeš. 2015. "Unexpectedly Streamlined Mitochondrial Genome of the Euglenozoan *Euglena gracilis*." *Genome Biology and Evolution* 7 (12): 3358–67. <https://doi.org/10.1093/gbe/evv229>.
- Ebel, Chantal, Christian Frantz, François Paulus, and P Imbault. 1999. "Trans-Splicing and Cis-Splicing in the Colourless Euglenoid, *Entosiphon Sulcatum*." *Current Genetics* 35 (5): 542–50. <https://doi.org/10.1007/s002940050451>.
- Ebenezer, Thankgod E., Martin Zoltner, Alana Burrell, Anna Nenarokova, Anna M.G. Novák Vanclová, Binod Prasad, Petr Soukal, et al. 2019. "Transcriptome, Proteome and Draft Genome of *Euglena Gracilis*." *BMC Biology* 17 (1): 1–23. <https://doi.org/10.1186/s12915-019-0626-8>.
- Flegontova, Olga, Pavel Flegontov, Shruti Malviya, Stephane Audic, Patrick Wincker, Colomban de Vargas, Chris Bowler, Julius Lukeš, and Aleš Horák. 2016. "Extreme Diversity of Diplonemid Eukaryotes in the Ocean." *Current Biology* 26 (22): 3060–65. <https://doi.org/10.1016/j.cub.2016.09.031>.

- Gawryluk, Ryan M.R., Javier del Campo, Noriko Okamoto, Jürgen F.H. F.H. Strassert, Julius Lukeš, Thomas A. A. Richards, Alexandra Z. Z. Worden, et al. 2016. “Morphological Identification and Single-Cell Genomics of Marine Diplonemids.” *Current Biology* 26 (22): 3053–59. <https://doi.org/10.1016/j.cub.2016.09.013>.
- Gawryluk, Ryan M.R., and Michael W Gray. 2009. “A Split and Rearranged Nuclear Gene Encoding the Iron-Sulfur Subunit of Mitochondrial Succinate Dehydrogenase in Euglenozoa” 7: 1–7. <https://doi.org/10.1186/1756-0500-2-16>.
- Gawryluk, Ryan M.R., Javier del Campo, Noriko Okamoto, Jürgen F.H. Strassert, Julius Lukeš, Thomas A. Richards, Alexandra Z. Worden, Alyson E. Santoro, and Patrick J. Keeling. 2016. “Morphological Identification and Single-Cell Genomics of Marine Diplonemids.” *Current Biology* 26 (22): 3053–59. <https://doi.org/10.1016/j.cub.2016.09.013>.
- Gockel, Gaby, and Wolfgang Hachtel. 2000. “Complete Gene Map of the Plastid Genome of the Nonphotosynthetic Euglenoid Flagellate *Astasia Longa*.” *Protist* 151 (4): 347–51. [https://doi.org/10.1078/S1434-4610\(04\)70033-4](https://doi.org/10.1078/S1434-4610(04)70033-4).
- Gumińska, Natalia, Magdalena Płecha, Bożena Zakryś, and Rafał Milanowski. 2018. “Order of Removal of Conventional and Nonconventional Introns from Nuclear Transcripts of *Euglena Gracilis*.” *PLoS Genetics* 14 (10). <https://doi.org/10.1371/journal.pgen.1007761>.
- Hallick, Richard B., Ling Hong, Robert G. Drager, Mitchell R. Favreau, Amparo Monfort, Bernard Orsat, Albert Spielmann, and Erhard Stutz. 1993. “Complete Sequence of *Euglena Gracilis* Chloroplast DNA.” *Nucleic Acids Research* 21 (15): 3537–44. <https://doi.org/10.1093/nar/21.15.3537>
- Hammond, Michael J, Anna Nenarokova, Anzhelika Butenko, Martin Zoltner, Eva Lacová Dobáková, Mark C Field, and Julius Lukeš. 2020. “A Uniquely Complex Mitochondrial Proteome from *Euglena Gracilis*.” *Molecular Biology and Evolution*, March. <https://doi.org/10.1093/molbev/msaa061>.
- Hamza, Iqbal, and Harry A. Dailey. 2012. “One Ring to Rule Them All: Trafficking of Heme and Heme Synthesis Intermediates in the Metazoans.” *Biochimica et Biophysica Acta - Molecular Cell Research* 1823 (9): 1617-1632 <https://doi.org/10.1016/j.bbamcr.2012.04.009>.
- Hannaert, Véronique, Emma Saavedra, Francis Duffieux, Jean Pierre Szikora, Daniel J. Rigden, Paul A.M. Michels, and Fred R. Opperdoes. 2003. “Plant-like Traits Associated with Metabolism of Trypanosoma Parasites.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (3): 1067–71. <https://doi.org/10.1073/pnas.0335769100>.
- Harada, Ryo, Yoshihisa Hirakawa, Akinori Yabuki, Yuichiro Kashiyama, Moe Maruyama, Ryo Onuma, Petr Soukal, et al. 2020. “Inventory and Evolution of Mitochondrion-Localized Family A DNA Polymerases in Euglenozoa.” *Pathogens* 9 (4): 257. <https://doi.org/10.3390/pathogens9040257>.

- Henze, Katrin, Abdelfattah Badr, Michael Wettern, Rudiger Cerff, , and William Martin. 1995. "A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. " *Proceedings of the National Academy of Sciences of the United States of America*, 92(20), 9122–9126. <https://doi.org/10.1073/pnas.92.20.9122>.
- Kanehisa, Minoru, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. 2016. "KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs." *Nucleic Acids Research* 45: 353–61. <https://doi.org/10.1093/nar/gkw1092>.
- Karnkowska, Anna, Matthew S. Bennett, and Richard E. Triemer. 2018. "Dynamic Evolution of Inverted Repeats in Euglenophyta Plastid Genomes." *Scientific Reports* 8 (1). <https://doi.org/10.1038/s41598-018-34457-w>.
- Kasiborski, Beth A, Matthew S Bennett, and Eric W Linton. 2016. "The Chloroplast Genome of *Phacus orbicularis* ( Euglenophyceae ): An Initial Datum Point for the Phacaceae" 411: 404–11. <https://doi.org/10.1111/jpy.12403>.
- Katoh, Kazutaka, John Rozewicki, and Kazunori D Yamada. 2019. "MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization" 20 (4): 1160–66. <https://doi.org/10.1093/bib/bbx108>.
- Kořený, Luděk, and Miroslav Oborník. 2011. "Sequence Evidence for the Presence of Two Tetrapyrrole Pathways in *Euglena Gracilis*." *Genome Biology Evolution* 3: 359–364. <https://doi.org/10.1093/gbe/evr029>.
- Kořený, Luděk, Roman Sobotka, Jan Janouškovec, Patrick J Keeling, and Miroslav Oborník. 2011. "Tetrapyrrole Synthesis of Photosynthetic Chromerids Is Likely Homologous to the Unusual Pathway of Apicomplexan Parasites." *The Plant Cell* 23 (9): 3454–62. <https://doi.org/10.1105/tpc.111.089102>.
- Lakey, Bryan, and Richard Triemer. 2016. " The Tetrapyrrole Synthesis Pathway as a Model of Horizontal Gene Transfer in Euglenoids." *Journal of Phycology* 51(1): 198-217. <https://doi.org/10.1111/jpy.12491>.
- Lax, Gordon, Won Je Lee, Yana Eglit, and Alastair Simpson. 2019. "Ploetids Represent Much of the Phylogenetic Diversity of Euglenids." *Protist* 170 (2): 233–57. <https://doi.org/10.1016/j.protis.2019.03.001>.
- Lax, Gordon, and Alastair G B Simpson. 2013. "Combining Molecular Data with Classical Morphology for Uncultured Phagotrophic Euglenids (Excavata): A Single-Cell Approach." *The Journal of Eukaryotic Microbiology* 60 (6): 615–25. <https://doi.org/10.1111/jeu.12068>.
- Leander, Brian S., Gordon Lax, Anna Karnkowska, and Alastair G. B. Simpson. 2017. "Euglenida." In *Handbook of the Protists*, 1–42. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-32669-6\\_13-1](https://doi.org/10.1007/978-3-319-32669-6_13-1).
- Leander, Brian S. 2004. "Did Trypanosomatid Parasites Have Photosynthetic Ancestors?" *Trends in Microbiology* 12 (6): 251–58. <https://doi.org/10.1016/j.tim.2004.04.001>.

- Leander, Brian S, Heather J Esson, and Susana A Breglia. 2007. "Macroevolution of Complex Cytoskeletal Systems in Euglenids." *BioEssays News and Reviews in Molecular Cellular and Developmental Biology* 29 (10): 987–1000. <https://doi.org/10.1002/bies.20645>.
- Li, L., Christian J. Stoeckert, and David S. Roos. 2003. "OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes." *Genome Research* 13 (9): 2178–89. <https://doi.org/10.1101/gr.1224503>.
- Milanowski, Rafał, Natalia Gumińska, Anna Karnkowska, Takao Ishikawa, and Bożena Zakryś. 2016. "Intermediate Introns in Nuclear Genes of Euglenids - Are They a Distinct Type?" *BMC Evolutionary Biology* 16 (1): 49. <https://doi.org/10.1186/s12862-016-0620-5>.
- Milanowski, Rafał, Anna Karnkowska, Takao Ishikawa, and Bożena Zakryś. 2014. "Distribution of Conventional and Nonconventional Introns in Tubulin (a and b) Genes of Euglenids." *Molecular Biology and Evolution* 31 (3): 584–593. <https://doi.org/10.1093/molbev/mst227>.
- Morales, Jorge, Tatsushi Mogi, Shigeru Mineki, Eizo Takashima, Reiko Mineki, Hiroko Hirawake, Kimitoshi Sakamoto, Satoshi Omura, and Kiyoshi Kita. 2009. "Novel Mitochondrial Complex II Isolated from Trypanosoma Cruzi Is Composed of 12 Peptides Including a Heterodimeric Ip Subunit." *Journal of Biological Chemistry* 284 (11): 7255–63. <https://doi.org/10.1074/jbc.M806623200>.
- Moriya, Yuki, Masumi Itoh, Shujiro Okuda, Akiyasu C. Yoshizawa, and Minoru Kanehisa. 2007. "KAAS: An Automatic Genome Annotation and Pathway Reconstruction Server." *Nucleic Acids Research* 35 (SUPPL.2). <https://doi.org/10.1093/nar/gkm321>.
- Müller, Miklós, Marek Mentel, Jaap J van Hellemond, Katrin Henze, Christian Woehle, Sven B Gould, Re-Young Yu, Mark van der Giezen, Aloysius G M Tielens, and William F Martin. 2012. "Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes." *Microbiology and Molecular Biology Reviews: MMBR* 76 (2): 444–95. <https://doi.org/10.1128/MMBR.05024-11>.
- Nguyen, Lam Tung, Heiko A. Schmidt, Arndt Von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74. <https://doi.org/10.1093/molbev/msu300>.
- Novák Vanclová, Anna M.G., Martin Zoltner, Steven Kelly, Petr Soukal, Kristína Záhonová, Zoltán Füssy, Thank God E. Ebenezer, et al. 2020. "Metabolic Quirks and the Colourful History of the Euglena Gracilis Secondary Plastid." *New Phytologist* 225 (4):1578-1592. <https://doi.org/10.1111/nph.16237>.
- Russell, Anthony G, Timothy E Shutt, Russell F Watkins, and Michael W Gray. 2005. "An Ancient Spliceosomal Intron in the Ribosomal Protein L7a Gene (Rpl7a) of Giardia Lamblia." *BMC Evolutionary Biology* 5 (1): 45. <https://doi.org/10.1186/1471-2148-5-45>.

- Saidha, T., A. I. Stern, D. H. Lee, and J. A. Schiff. 1985. "Localization of a Sulphate-Activating System within Euglena Mitochondria." *The Biochemical Journal* 232 (2): 357–65. <https://doi.org/10.1042/bj2320357>.
- Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics (Oxford, England)* 30 (9): 1312–13. <https://doi.org/10.1093/bioinformatics/btu033>.
- Tanaka, Yuji, Takahisa Ogawa, Takanori Maruta, Yuta Yoshida, Kazuharu Arakawa, and Takahiro Ishikawa. 2017. "Glucan Synthase-like 2 Is Indispensable for Paramylon Synthesis in Euglena Gracilis." *FEBS Letters*. Wiley Blackwell. <https://doi.org/10.1002/1873-3468.12659>.
- Vesteg, Matej, Rostislav Vacula, Jürgen M. Steiner, Bianka Mateášiková, Wolfgang Löffelhardt, Broňa Brejová, and Juraj Krajčovič. 2010. "A Possible Role for Short Introns in the Acquisition of Stroma-Targeting Peptides in the Flagellate Euglena Gracilis." *DNA Research* 17 (4): 223–31. <https://doi.org/10.1093/dnares/dsq015>.
- Wołowski, Konrad. 2011. Euglenophyta (Euglenoids). In: The freshwater algal flora of the British Isles. An identification guide to freshwater and terrestrial algae. Second edition. (John, D.M., Whitton, B.A. & Brook, A.J. Eds), pp. 181-239. Cambridge: Cambridge University Press.
- Woo, Yong H, Hifzur Ansari, Thomas D Otto, Christen M Klinger, Martin Kolísko, Jan Michálek, Alka Saxena, et al. 2015. "Chromerid Genomes Reveal the Evolutionary Path from Photosynthetic Algae to Obligate Intracellular Parasites." *ELife* 4: 1–41. <https://doi.org/10.7554/eLife.06974>.
- Záhonová, Kristína, Zoltán Füssy, Erik Birčák, Anna M.G. Novák Vanclová, Vladimír Klimeš, Matej Vesteg, Juraj Krajčovič, Miroslav Oborník, and Marek Eliáš. 2018. "Peculiar Features of the Plastids of the Colourless Alga Euglena Longa and Photosynthetic Euglenophytes Unveiled by Transcriptome Analyses." *Scientific Reports* 8 (1): 1–15. <https://doi.org/10.1038/s41598-018-35389-1>.



**Table 1:** Parameters of the genomic assembly.

Number of reads Miseq	11 624 864
Number of reads Hiseq	82 258 718
Number of contigs	143 763
Number of contigs ( $\geq 500$ bp)	82 982
Number of contigs ( $\geq 1000$ bp)	36 105
Number of contigs ( $\geq 5000$ bp)	153
Median contig length	661
Median contig length ( $\geq 500$ bp)	924
Mean contig length	744
Mean contig length ( $\geq 500$ bp)	1 135
Max contig length	8 093
Total length	106 888 161
Total length ( $\geq 500$ bp)	94 209 074
Total length ( $\geq 1000$ bp)	58 674 546
N50	1 194
L50	25 493
GC (%)	51,63
Number of N's per 100 kbp	31,22

**Table 2:** Parameters of the transcriptomic assembly.

Number of reads	12 103 119
Number of contigs	93 852
Number of contigs ( $\geq 1000$ bp)	19 335
Number of contigs ( $\geq 5000$ bp)	73
Median contig length	430
Mean contig length	651
Max contig length	17 913
Total length	66 880 466
N50	856
GC (%)	58
Putative proteins	55 783
Unique proteins	39 585
Any homologue (e-value $> 10^{-5}$ )	26 052

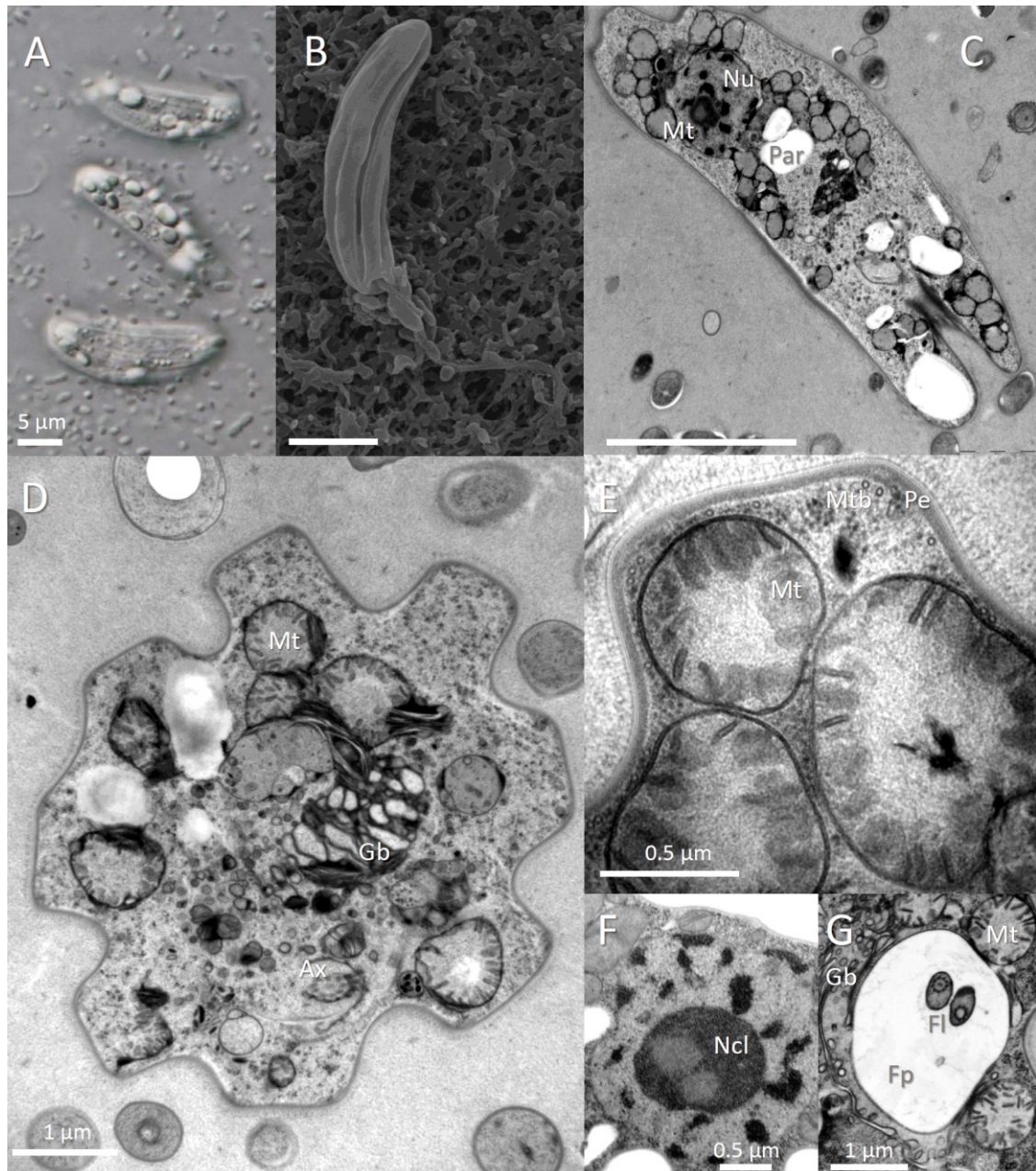
**Table 3:** Introns identified in the selected genes of *Rhabdomonas costata*.

Gene	Gene abb.	Complete Conventional introns	Incomplete intron *	ORF length (nt)	gDNA coverage (%)**
Tubulin alfa	tubA	2	0	1356	100
Tubulin beta	tubB	4	0	1335	100
Tubulin gamma	tubG	6	7	1680	76
Heatshock protein 90	Hsp90	10	5	2112	68
Glyceraldehyde-3- phosphate dehydrogenase	gapC	3	1	1062	70
Fibrillarin	nop1p	4	1	903	85

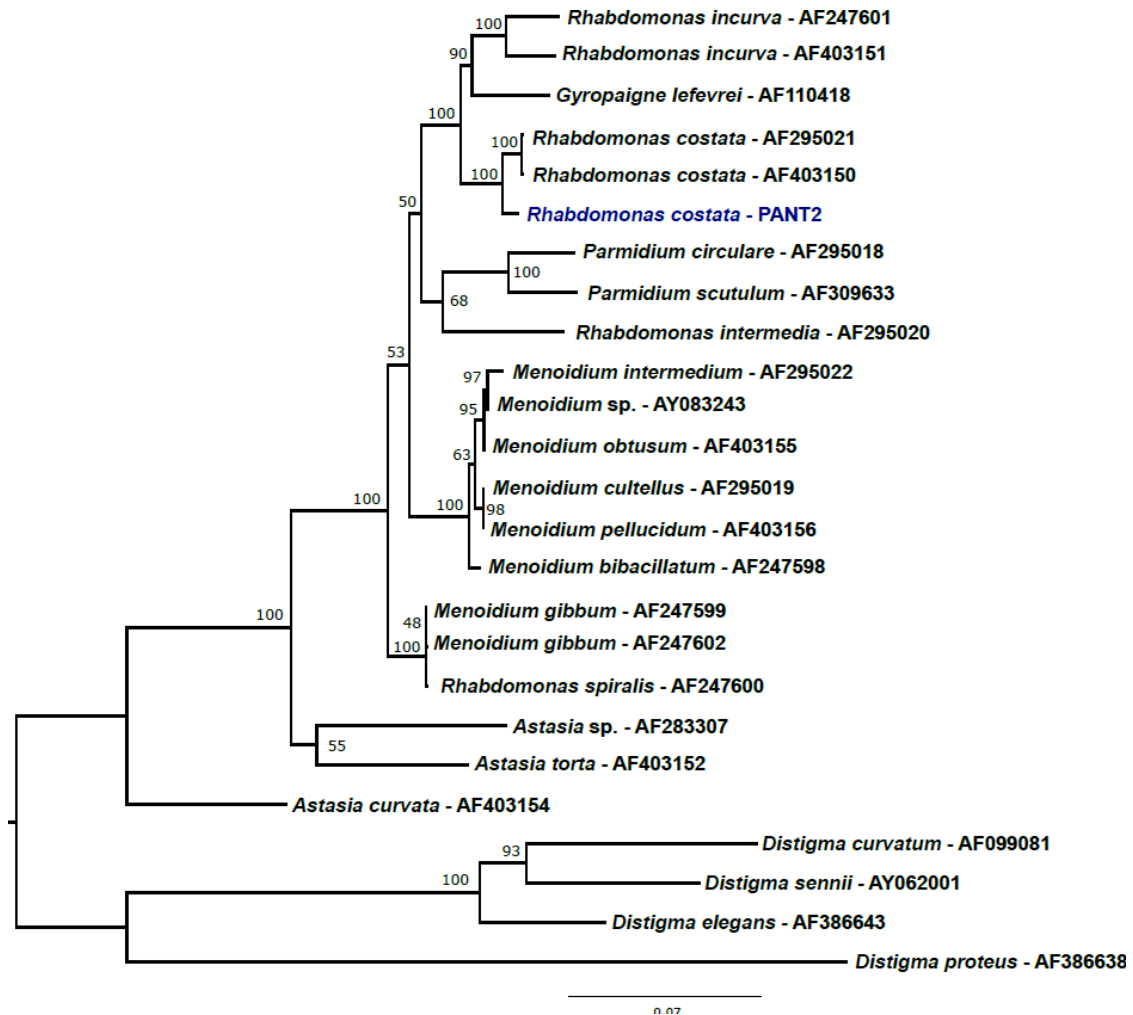
\* Only one intron boundary was found in the data, thus the intron type could not be determined with certainty.

\*\* Part of ORF length mapped to gDNA. Number of introns may not be definitive in low-percentage coverage.

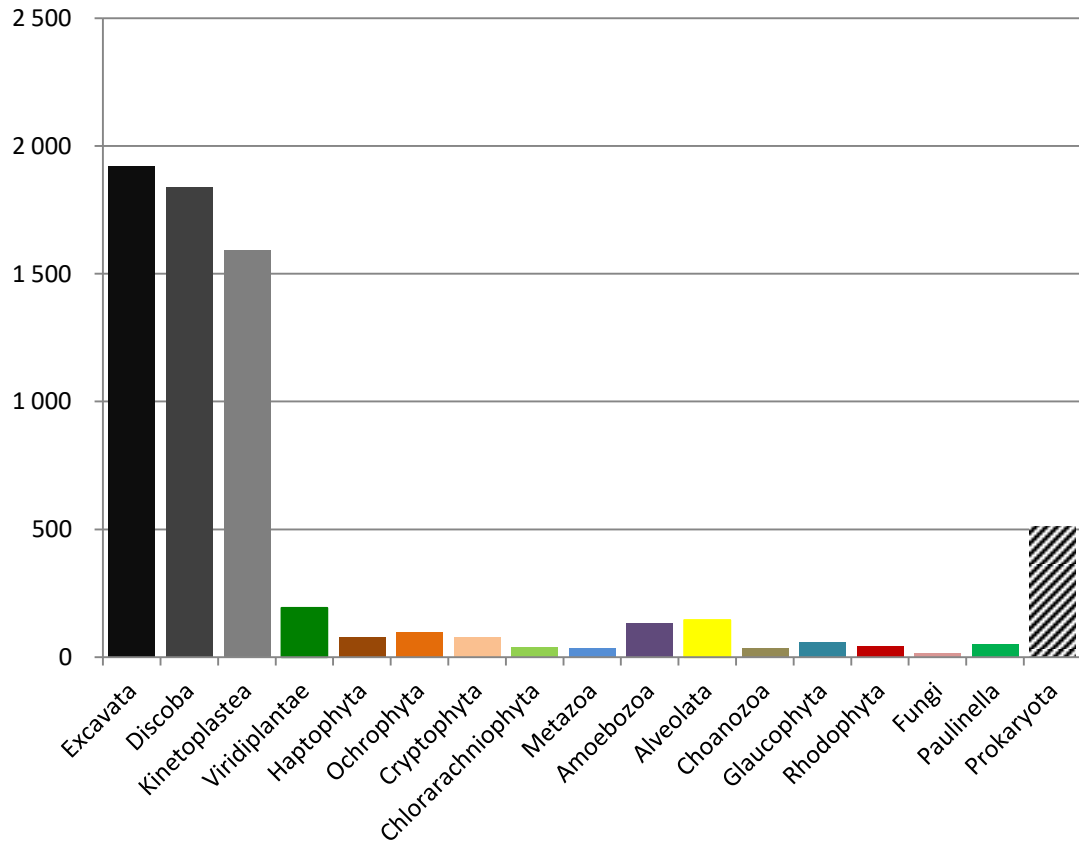
**Figure 1: Microscopic investigation of *Rhabdomonas costata*.** Cells in DIC contrast (A) with visible pellicular stripes and paramylon grains. SEM microscopy (B) of a cell showing surface invaginations and a flagellum inserted in the flagellar pocket. Longitudinal (C) and transverse (D) TEM sections and details of pellicle and mitochondria (E), nucleus (F) and flagellar pocket with two flagella (G). Ax – axoneme, Gb – Golgi body, Fl – flagella, Fp – flagellar pocket, Mt – mitochondria, Mtb – subpellicular microtubules, Nu – nucleus, Ncl – nucleolus, Par – paramylon grain, Pe – pellicle.



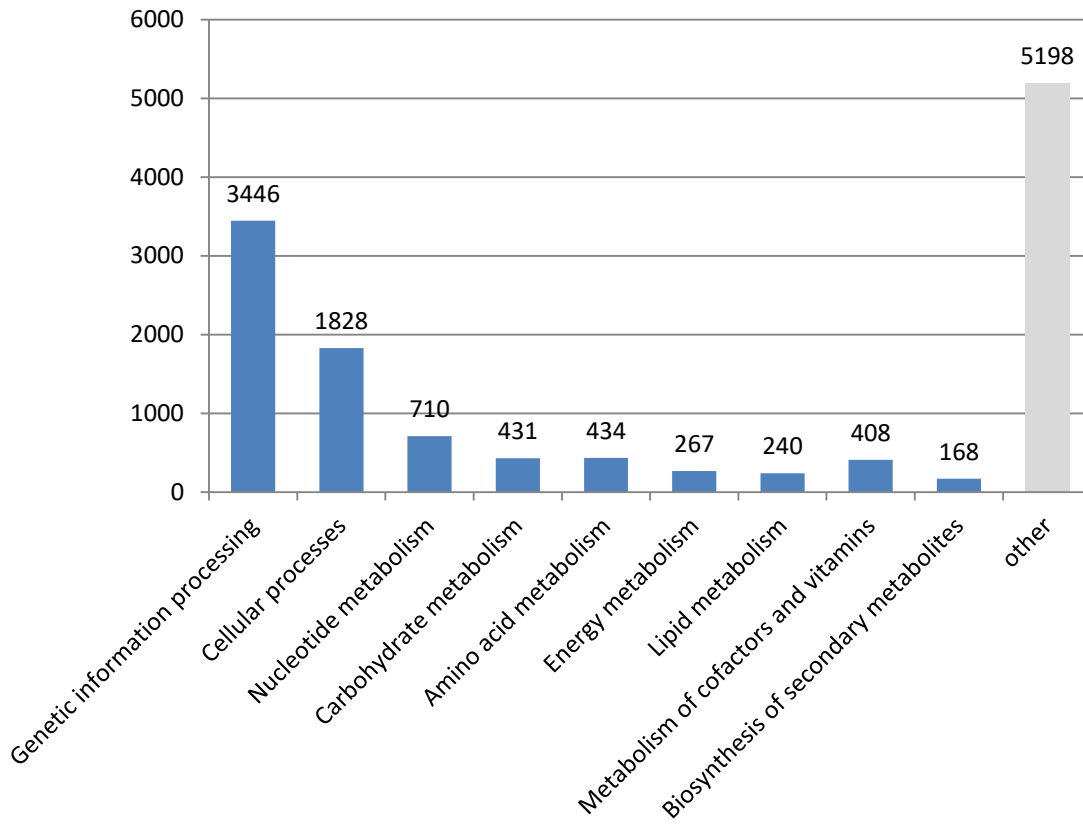
**Figure 2: Phylogenetic tree of euglenids based on 18S rRNA gene.** The tree was constructed in IQ-Tree using TIM2e+G4 model selected by Bayesian information criterion from a trimmed alignment containing 1,569 positions. The values at the nodes represent ultrafast bootstraps from 1,000 repetitions. The strain analysed in this study is shown in blue.



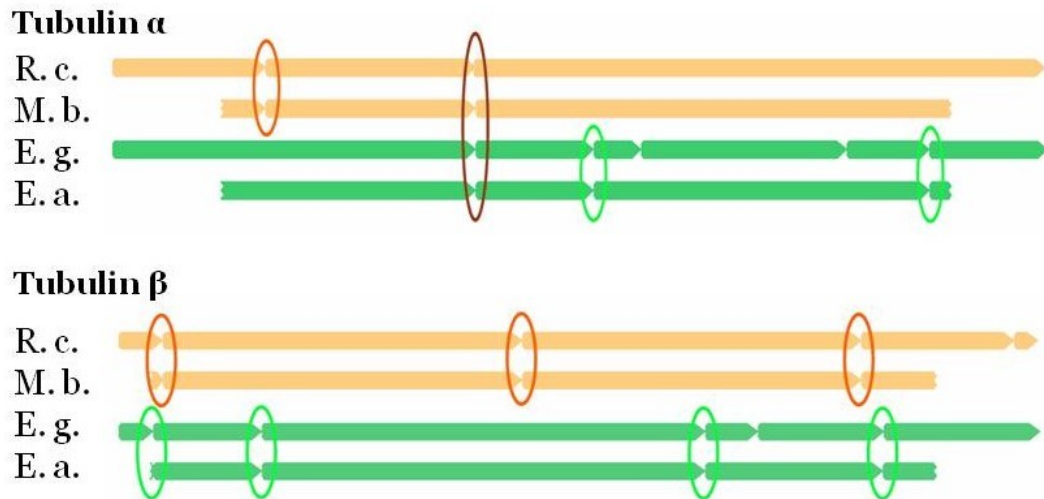
**Figure 3: Graph summarising taxonomical affinities of the predicted proteins of *Rhabdomonas costata*.** 3,445 protein phylogenies, in which *R. costata* was robustly placed into a taxonomically homogeneous clan, were sorted accordingly into taxonomic bins. Discoba and Kinetoplastea represent the subgroups of Excavata.



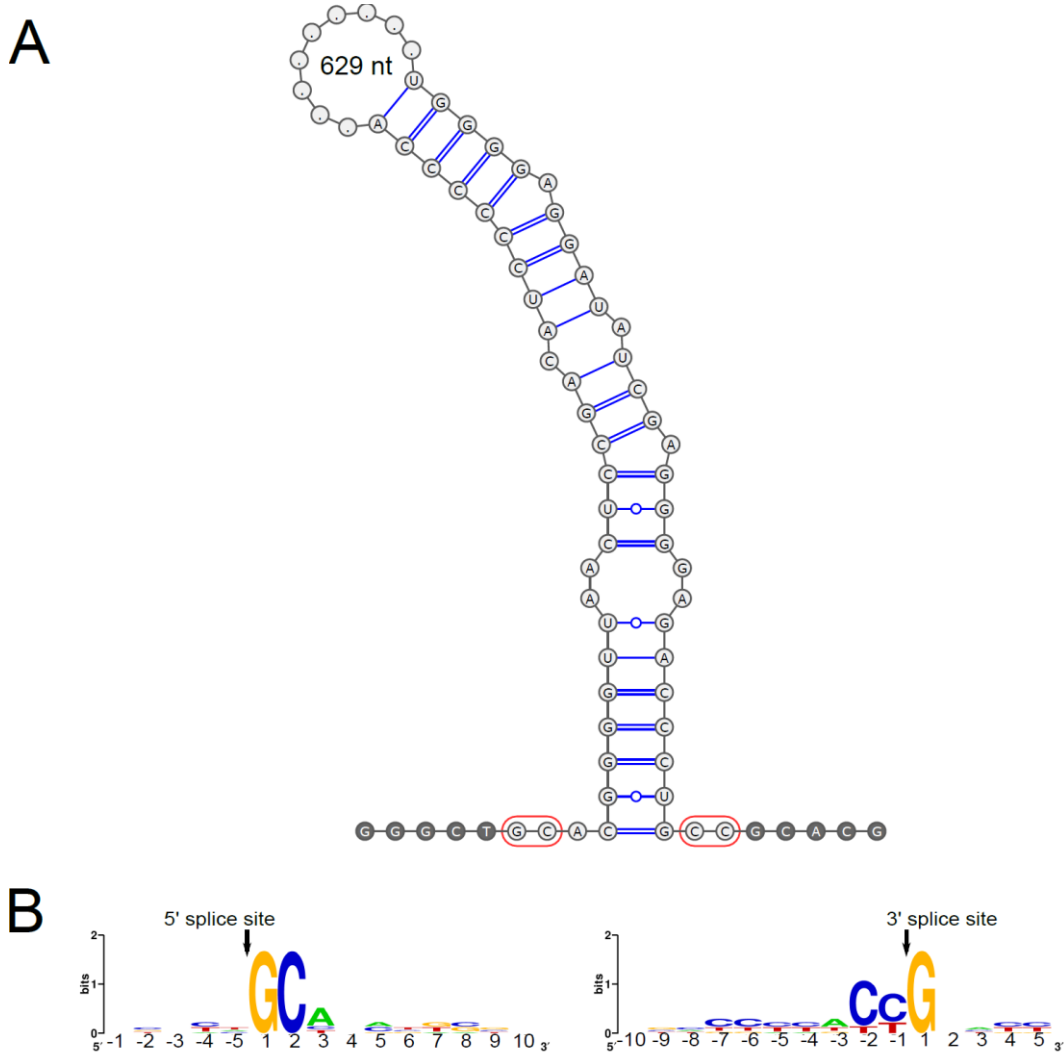
**Figure 4: KEGG Functional categories of predicted proteins of *R. costata*.** The 13,130 proteins (33 % out of 39,585) were ascribed to functional categories according to KEGG. The categories with less than 10 members were fused to category “other”.



**Figure 5: Comparison of positions of introns in tubulin  $\alpha$  and  $\beta$  genes in four euglenids.** Heterotrophic euglenids *Rhabdomonas costata* and *Menoidium bibacillatum* (orange lines) and phototrophic euglenids *Euglena gracilis* and *Euglena agilis* (green lines). Introns shared between heterotrophs – orange ellipses, phototrophs – green ellipses, all four euglenids – brown ellipse.

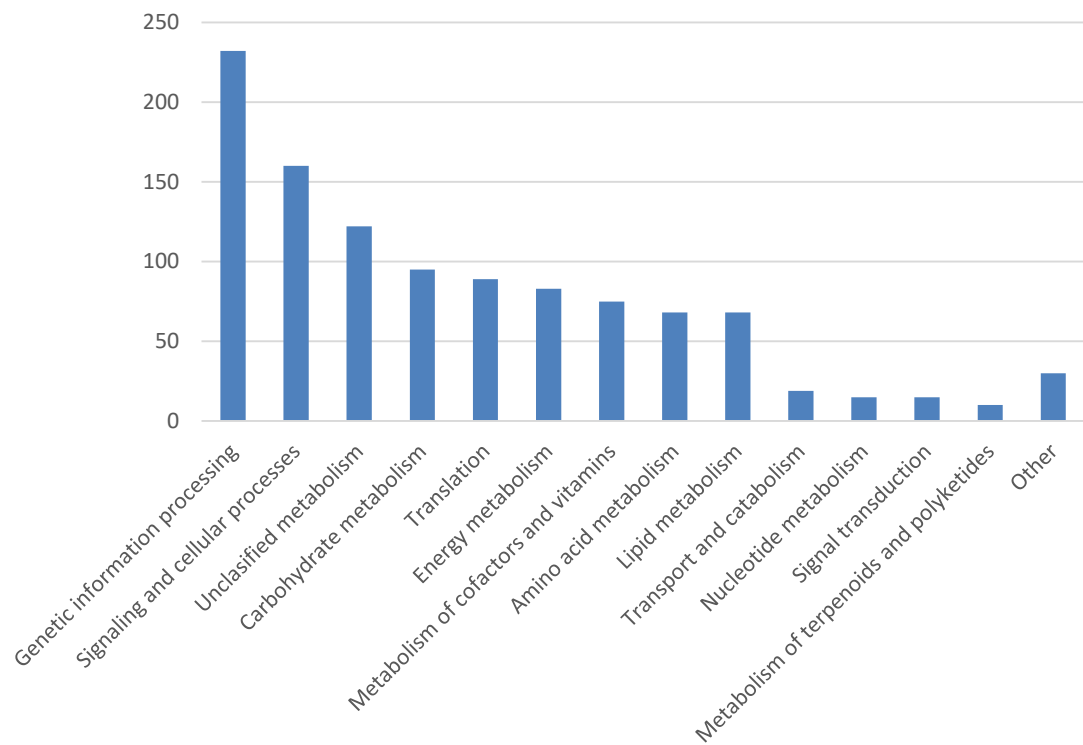


**Figure 6: Example of the non-conventional intron of *R. costata*.** The secondary structure of a non-conventional intron from NODE\_718 (A) and the sequence logo of boundaries of putative non-conventional intron detected in *R. costata* (B).

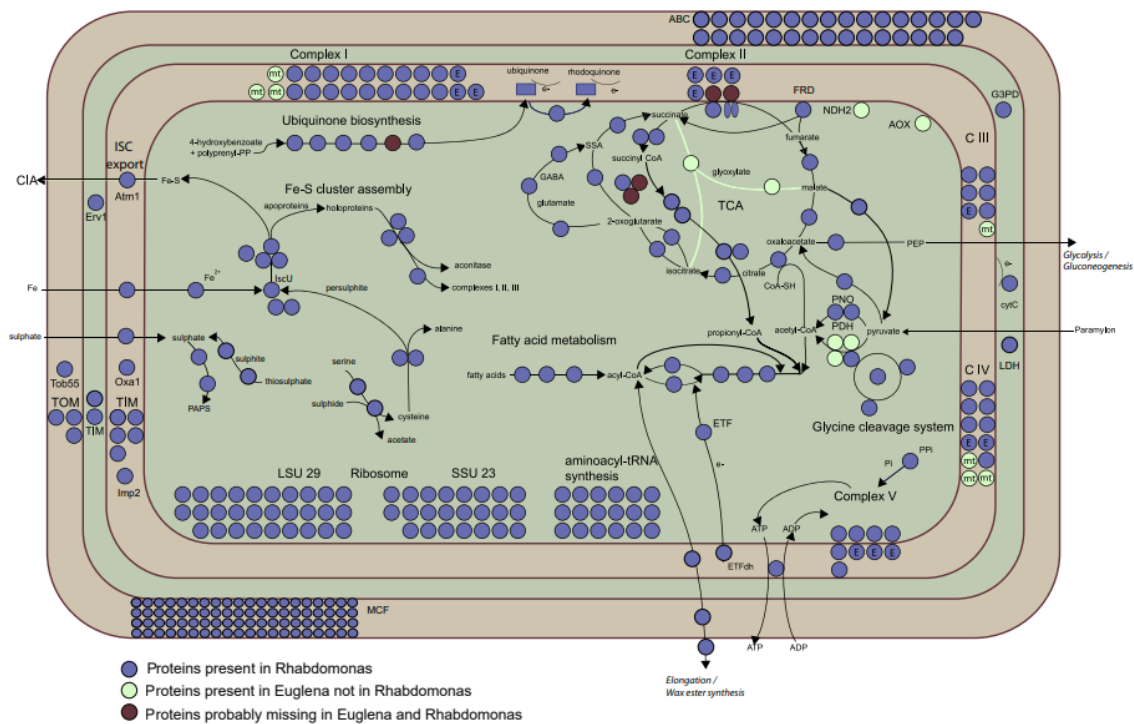




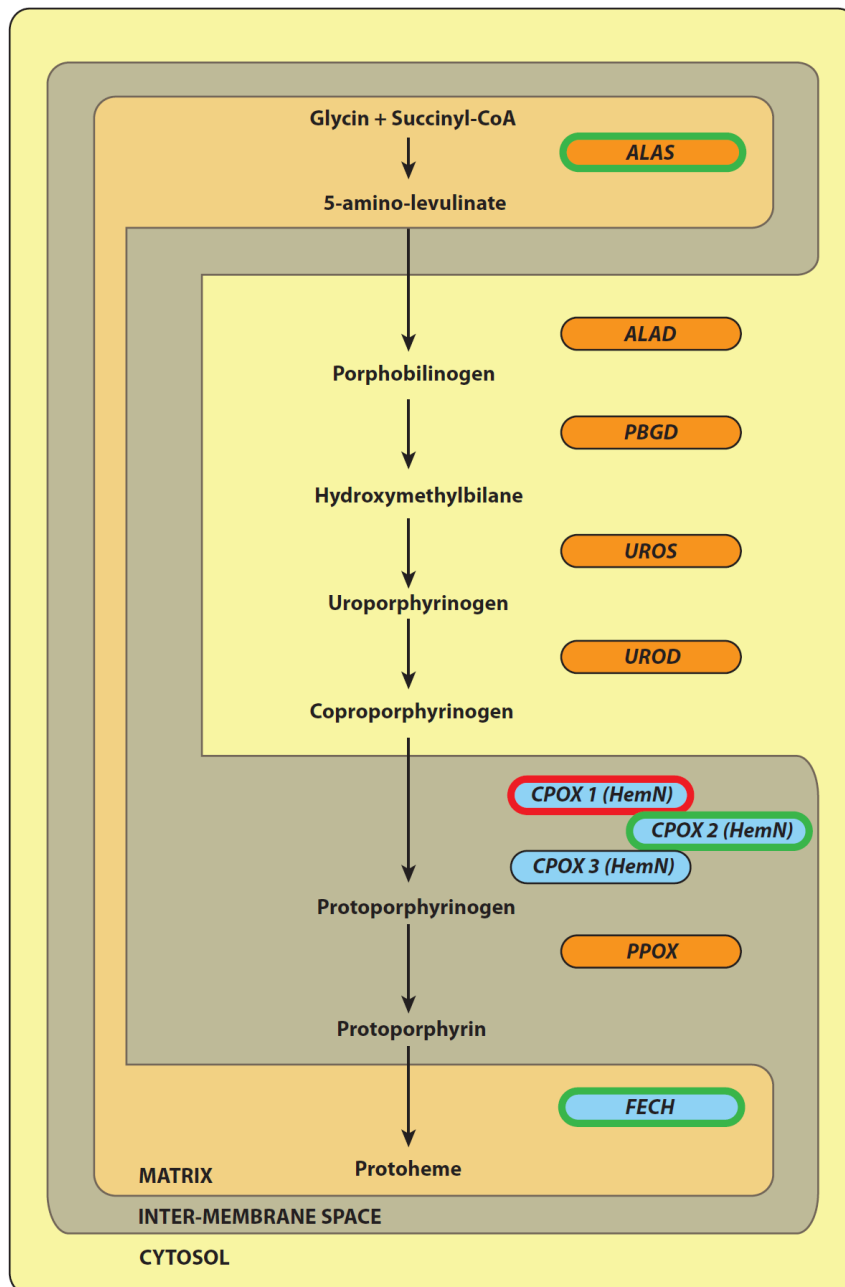
**Figure 7: KEGG Functional categories of proteins predicted to mitochondrial proteome of *R. costata*.** The 1,081 proteins of the mitochondrial proteome were ascribed to functional categories according to KEGG. The categories with less than 10 members were fused to category “other”.



**Figure 8: Metabolic map of *R. costata* mitochondrion.** The map is based on *in silico* prediction of mitochondrial proteome from the transcriptomic data set. Blue circles represent enzymes present in *R. costata* and *E. gracilis*, grey circles represent enzymes absent in *R. costata* but present in *E. gracilis* and brown circles represent typical eukaryotic enzymes absent in both euglenids. To allow comparison inspired by Ebenezer et al. 2019.



**Figure 9: Origins and subcellular localization of tetrapyrrole synthesis enzymes in *R. costata*.** Predicted localisation of enzyme is indicated by the position in the diagram. Phylogenetic origin is indicated by colour (orange – eukaryotic origin, blue – eubacterial origin), presence of mitochondrial targeting peptide (TP) is indicated by frame (red frame – high TP value in *R. costata*, green frame – high TP value in sister euglenophytes).





## 8 Shrnutí

Předložená práce shrnuje stávající poznatky o endosymbióze a zaměřuje se na sekundární endosymbiózu u euglenofyt. Tři původní práce dokumentují plastidový genom *Eutreptiella gymnastica*, mitochondriální genom *Pyramimonas parkeae* a celkový genom a transkriptom *Rhabdomonas costata*. Přílohou je také kapitola Secondary plastids of Euglenophytes z knihy Secondary Endosymbioses, na které jsem se podílela.

S rozvojem masivního paralelního sekvenování bylo možné v poměrně krátké době osmi let osekvenovat 30 plastidových genomů 13 rodů skupiny Euglenophyceae. Genomy mají prakticky stejný genový obsah (kromě sekundárně osmotrofní *E. longa*) (Vanclová et al., 2017), který se ustavil v procesu endosymbiózy redukcí plastidového genomu prasinofytní řasy, jež byla předkem sekundárního plastidu. Jedná se o soubor 89 genů, které mají plastidovou funkci a jen výjimečně chybí v některém plastidovém genomu euglenofyt. Kromě těchto genů jsou v plastidových genomech přítomny i geny pro maturázy/reverzní transkriptázy, které se účastní sestřihu intronů typu II (Doetsch et al., 1998). Počet těchto proteinů do značné míry koreluje s počtem intronů. U euglenofyt se kromě intronů typu II vyskytují i introny typu III a tzv. twintrony, introny uvnitř jiného intronu (Copertino et al., 1991; Doetsch et al., 1998). Plastidové genomy se také liší přítomností či nepřítomností invertovaných repetit (Karnkowska et al., 2018). Všechny tyto znaky mohou dobře sloužit při fylogenetických analýzách společně se znaky morfologickými a molekulárně biologickými.

Se znalostí plastidového genomu *P. parkeae*, nejbližší známé příbuzné řasy předka endosymbionta euglenofyt, bylo možno určit, které z plastidových genů byly přeneseny do jádra endosymbiotickým genovým přenosem a které zmizely úplně (Hrdá et al., 2012). V transkriptomu *E. gracilis* a *Etl. gymnastica* jsme našli několik takových genů (*ccsA*, *fstH*, *infA*, *YCF3* a *petA*), které zřejmě mají signální peptid, který je navádí do plastidu, kde byla jejich přítomnost potvrzena proteomicky (Novák Vanclová et al., 2020). Zjištění, že tyto konkrétní geny byly přeneseny do jádra, je cenné, neboť u jiných organismů se sekundárním nebo komplexním plastidem nebyl tento přesun dosud popsán. Endosymbióza u euglenofyt proběhla poměrně nedávno a bylo tak snažší najít řasu, která je zřejmě velmi blízce příbuzná předkovi řasy pohlcené společným předkem euglenofyt a identifikovat v jejím plastidovém genomu geny přenesené do jádra euglenofyt.

Další upřesnění by mohlo přinést nalezení ještě bližšího příbuzného původního endosymbionta, kterým by mohla být například prasinofytní řasa *Pterosperma cristatum*, jejíž částečně osekvenovaný plastidový a mitochondriální genom čeká v naší laboratoři na další bioinformatické zpracování.

Zajímavým a nepříliš často diskutovaným fenoménem provázejícím sekundární endosymbiózu je fakt, že se při ní zcela ztrácí mitochondrie endosymbionta. Přítomnost dvou funkčních mitochondrií různého evolučního původu v buňce je zdokumentována u dinoflagelátů *Durinskia* a *Kryptoperidinium* (Imanian et al., 2007), ale u většiny linií s komplexním plastidem byla

mitochondrie endosymbionta eliminována. Snažili jsme se odpovědět na otázku, zda se během formování organely mohly alespoň některé geny z mitochondrie pohlcené řasy přesunout do jádra hostitele. Naše data získaná osekvenováním mitochondriálního genomu *P. parkeae* tuto hypotézu nepodporují, neboť v transkriptomu *E. gracilis* jsme žádné geny, které by mohly pocházet z tohoto genomu, nenalezli.

Endosymbiotický genový přenos byl mnohokrát dokumentován a často bylo zjištěno, že v jaderném genomu hostitelské buňky se nalézají i geny, které jsou jiného původu než získaný plastid (Maruyama et al., 2011; Moustafa et al., 2009; Y. Yang et al., 2011). Vysoký podíl genů z jiné linie řas, než ze které pochází komplexní plastid, vedly k postulování několika teorií o původu těchto genů a procesu jejich přijetí. Teorie tvoří souvislou škálu, která začíná postupným získáváním cizích, prokaryotických i eukaryotických genů z potravy (Doolittle, 1998), přes získávání genů z přechodných endosymbiontů, kterých mohlo být i několik různých, tzv. "shopping bag model" a "red carpet model" (Larkum et al., 2007; Ponce-Toledo et al., 2019), po případy, kdy byla endosymbióza završena, ale původní plastid byl ztracen a místo něho byl získán jiný, jak nyní pozorujeme například u obrněnky *Lepidodinium* (Minge et al., 2010) až po postupné zisky genů z různých zdrojů jako kompenzace redukujícího se genomu organely (Nowack et al., 2016). Podíl genů nepřibuzných endosymbiontovi by mohl napovědět, jakým způsobem, ze tří výše popsaných, byly geny získány, a zda mohlo dojít v evoluční historii organismu k opakované endosymbióze, jak ji například u euglenofyt předpokládá Markunas (Markunas & Triemer, 2016). K analýzám je ale třeba přistupovat velice opatrně, neboť prostá podobnost získaná na základě BLAST může nadhodnocovat určité typy výsledků, ať už v důsledku nerovnoměrného zastoupení různých linií ve srovnávacích databázích, nebo kvůli nesprávné anotaci genů či jiným sekvenčním artefaktům. Reanalýzou některých prvotních výsledků se zjistilo, že podíl cizích jaderných genů nepocházejících z EGT je nižší, než se původně zdálo (Burki et al., 2012; Deschamps & Moreira, 2012). Pro zkoumání tohoto jevu je proto důležité sestavovat pro jednotlivé geny dostatečně robustní fylogenetické stromy, které původ genů určí s větší přesností. Předpokladem je také doplňování databází o genomy dalších organismů, s plastidem nebo bez něj, tak aby byla pokryta celá škála eukaryotických organismů, včetně méně prozkoumaných linií.

Takovou mezeru v databázi jsme se pokusili doplnit osekvenováním částečného genomu a poměrně úplného transkriptomu u heterofníhého euglenida *Rhabdomonas costata*. V jeho genomu jsme našli jen zanedbatelný počet genů, pocházejících z řas, ať už červených nebo zelených, a proto vyvozujeme, že tento organismus nikdy plastid neměl. To vyvrací hypotézu časného získání plastidu u společného předka kmene Euglenozoa (Hannaert et al., 2003; Martin & Borst, 2003). Zároveň znaky nalezené u *R. costata* tvoří pozadí pro znaky nalezené u *E. gracilis* a lze tedy lépe určit, které vlastnosti se pravděpodobně pojí se získáním plastidu. Díky tomu, že jsme rekonstruovali dráhu syntézy tetrapyrólů u *R. costata*, jsme mohli určit, které z paralogů genů *UROS*, *UROD* a *CPOX* u *E. gracilis* pochází pravděpodobně od společného předka euglenidů. Také jsme u *R. costata* našli některé mitochondriální proteiny, které byly doposud v rámci euglenofyt známy jen u *E. gracilis*, například enzymy tzv.  $\alpha$ -proteobakteriálního "shuntu", které modifikují klasický Krebsův cyklus, nebo přítomnost

pyruvát:NADP+oxidoreduktázy, enzymu který využívají i někteří anaerobně žijící prvoci (Muller et al., 2012). Tyto geny lze tedy považovat za geny přítomné u společného předka *R. costata* a *E. gracilis*.

U *Rhabdomonas costata* nejsou přítomné některé geny dráhy asimilace sulfátu a zdá se, že dráha je u *E. gracilis* doplněna o geny pocházející z endosymbionta (Hammond et al., 2020). Stejně tak chybí *R. costata* wax syntáza přítomná u *E. gracilis*, která je dle BLAST podobná zástupcům skupiny Chloroplastida a je tak také zřejmě endosymbiotického původu. Srovnáním transkriptomů *R. costata* a *E. gracilis* bude možné vytipovat i další geny, které by měly projít důkladnější fylogenetickou analýzou, aby bylo možné určit, zda pocházejí skutečně z EGT, či jsou výsledkem HGT.

Vývoj sekvenačních technologií jde velmi rychle dopředu a tak dalších genomických, transkriptomických i proteomických dat bude neustále přibývat. Kromě obohacování databází musí dojít také k vylepšování algoritmů pro jejich analýzu. Naše data i zvolené postupy k jejich vyhodnocování k oběmu výrazně přispěli a jsou připravené pro další studie.

## 9 Literatura

- Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., Brown, M. W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., ... Zhang, Q. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology*, 66(1), 4–119. <https://doi.org/10.1111/jeu.12691>
- Archibald, J. M. (2015). Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), 10147–10153. <https://doi.org/10.1073/pnas.1421374112>
- Baurain, D., Brinkmann, H., Petersen, J., Rodríguez-Ezpeleta, N., Stechmann, A., Demoulin, V., Roger, A. J., Burger, G., Lang, B. F., & Philippe, H. (2010). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Molecular Biology and Evolution*, 27(7), 1698–1709. <https://doi.org/10.1093/molbev/msq059>
- Bell, E. M., & Laybourn-Parry, J. (2003). Mixotrophy in the antarctic phytoflagellate, *Pyramimonas gelidicola* (Chlorophyta: Prasinophyceae). *Journal of Phycology*, 39(4), 644–649. <https://doi.org/10.1046/j.1529-8817.2003.02152.x>
- Bennett, M. S., & Triemer, R. E. (2015). Chloroplast genome evolution in the Euglenaceae. *Journal of Eukaryotic Microbiology*, 62(6), 773–785. <https://doi.org/10.1111/jeu.12235>
- Bennett, M. S., Wiegert, K. E., & Triemer, R. E. (2012). Comparative chloroplast genomics between *Euglena viridis* and *Euglena gracilis* (Euglenophyta). *Phycologia*, 51(6), 711–718. <https://doi.org/10.2216/12-017.1>
- Bennett, M. S., Wiegert, K. E., & Triemer, R. E. (2014). Characterization of *Euglenaformis* gen. nov. and the chloroplast genome of *Euglenaformis* [*Euglena*] *proxima* (Euglenophyta). *Phycologia*, 53(1), 66–73. <https://doi.org/10.2216/13-198.1>
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., ... El-Sayed, N. M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, 309(5733), 416–422. <https://doi.org/10.1126/science.1112642>
- Bicudo, C. E. d. M., & Menezes, M. (2016). Phylogeny and classification of euglenophyceae: A brief review. *Frontiers in Ecology and Evolution*, 4(MAR), 1–15. <https://doi.org/10.3389/fevo.2016.00017>
- Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S., & Maier, U. G. (2009). Protein targeting into secondary plastids. *Journal of Eukaryotic Microbiology*, 56(1), 9–15. <https://doi.org/10.1111/j.1550-7408.2008.00370.x>
- Bonen, L., & Vogel, J. (2001). The ins and outs of group II introns. *Trends in Genetics*, 17(6), 322–331. [https://doi.org/10.1016/S0168-9525\(01\)02324-1](https://doi.org/10.1016/S0168-9525(01)02324-1)
- Breglia, S. A., Yubuki, N., & Leander, B. S. (2013). Ultrastructure and molecular phylogenetic position of heteronema scaphurum: A eukaryovorous euglenid with a cytoproct. *Journal of Eukaryotic Microbiology*, 60(2), 107–120. <https://doi.org/10.1111/jeu.12014>



- Brueckner, J., & Martin, W. F. (2020). Bacterial Genes Outnumber Archaeal Genes in Eukaryotic Genomes. *Genome Biology and Evolution*, *12*(4), 282–292. <https://doi.org/10.1093/gbe/evaa047>
- Burger, G., Plante, I., Lonergan, K. M., & Gray, M. W. (1995). The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. *Journal of Molecular Biology*, *245*(5), 522–537. <https://doi.org/10.1006/jmbi.1994.0043>
- Burger, G., Gray, M. W., Forget, L., & Lang, B. F. (2013). Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biology and Evolution*, *5*(2), 418–438. <https://doi.org/10.1093/gbe/evt008>
- Burger, G., & Nedelcu, A. M. (2012). Mitochondrial genomes of algae. In R. Bock & V. Knoop (Eds.), *Genomics of Chloroplasts and Mitochondria. Advances in Photosynthesis and Respiration*, *35* (127–157). Springer Netherlands. [https://doi.org/10.1007/978-94-007-2920-9\\_6](https://doi.org/10.1007/978-94-007-2920-9_6)
- Burki, F. (2017). The convoluted evolution of eukaryotes with complex plastids. In Y. Hirakawa (Eds.), *Secondary endosymbioses. Advances in Botanical Research*, *84* (1–30). <https://doi.org/10.1016/bs.abr.2017.06.001>
- Burki, F., Flegontov, P., Oborník, M., Cihlář, J., Pain, A., Lukeš, J., Keeling, P. J. (2012). Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biology and Evolution*, *4*(6), 626–635. <https://doi.org/10.1093/gbe/evs049>
- Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The New Tree of Eukaryotes. *Trends in Ecology and Evolution*, *35*(1), 43–55. <https://doi.org/10.1016/j.tree.2019.08.008>
- Cabała, J. (2003). Some interesting colourless euglenophytes found in Southern Poland. *Acta Societatis Botanicorum Poloniae*, *72*(3), 243–247. <https://doi.org/10.5586/asbp.2003.032>
- Canaday, J., Tessier, L. H., Imbault, P., & Paulus, F. (2001). Analysis of *Euglena gracilis* alpha-, beta- and gamma-tubulin genes: Introns and pre-mRNA maturation. *Molecular and General Genetics*, *265*(1), 153–160. <https://doi.org/10.1007/s004380000403>
- Cavalier-Smith, T. (1981). Eukaryote kingdoms: Seven or nine? *BioSystems*, *14*(3–4), 461–481. [https://doi.org/10.1016/0303-2647\(81\)90050-2](https://doi.org/10.1016/0303-2647(81)90050-2)
- Cavalier-Smith, T. (2016). Higher classification and phylogeny of Euglenozoa. *European Journal of Protistology*, *56*, 250–276. <https://doi.org/10.1016/j.ejop.2016.09.003>
- Cavalier-Smith, T. (2017). Euglenoid pellicle morphogenesis and evolution in light of comparative ultrastructure and trypanosomatid biology: Semi-conservative microtubule/strip duplication, strip shaping and transformation. *European Journal of Protistology*, *61*, 137–179. <https://doi.org/10.1016/j.ejop.2017.09.002>
- Cenci, U., Sibbald, S. J., Curtis, B. A., Kamikawa, R., Eme, L., Moog, D., Henrissat, B., Maréchal, E., Chabi, M., Djemiel, C., Roger, A. J., Kim, E., & Archibald, J. M. (2018). Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. *BMC Biology*, *16*(1), 1–23. <https://doi.org/10.1186/s12915-018-0593-5>

- Christopher, D. A., & Hallick, R. B. (1989). *Euglena gracilis* chloroplast ribosomal protein operon: A new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic Acids Research*, 17(19), 7591–7608. <https://doi.org/10.1093/nar/17.19.7591>
- Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., & Jansen, R. K. (2006). The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution*, 23(11), 2175–2190. <https://doi.org/10.1093/molbev/msl089>
- Cihlář, J., Füssy, Z., Horák, A., & Oborník, M. (2016). Evolution of the tetrapyrrole biosynthetic pathway in secondary algae: Conservation, redundancy and replacement. *PLoS ONE*, 11(11): e0166338. <https://doi.org/10.1371/journal.pone.0166338>
- Copertino, D. W., Christopher, D. A., & Hallick, R. B. (1991). A mixed group II/group III twintron in the *Euglena gracilis* chloroplast ribosomal protein S3 gene: Evidence for intron insertion during gene evolution. *Nucleic Acids Research*, 19(23), 6491–6497. <https://doi.org/10.1093/nar/19.23.6491>
- Copertino, D. W., Hall, E. T., Van Hook, F. W., Jenkins, K. P., & Hallick, R. B. (1994). A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Research*, 22(6), 1029–1036. <https://doi.org/10.1093/nar/22.6.1029>
- Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M. C., Ball, S. G., Gile, G. H., Hirakawa, Y., Hopkins, J. F., Kuo, A., Rensing, S. A., Schmutz, J., Symeonidi, A., Elias, M., Eveleigh, R. J. M. M., Herman, E. K., Klute, M. J., ... Archibald, J. M. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature*, 492(7427), 59–65. <https://doi.org/10.1038/nature11681>
- Dabbagh, N., Bennett, M. S., Triemer, R. E., & Preisfeld, A. (2017). Chloroplast genome expansion by intron multiplication in the basal psychrophilic euglenoid *Eutreptiella pomquetensis*. *PeerJ*, 2017(8), 1–22. <https://doi.org/10.7717/peerj.3725>
- Dabbagh, N., & Preisfeld, A. (2017). The chloroplast genome of *Euglena mutabilis*—cluster arrangement, intron analysis, and intrageneric trends. *Journal of Eukaryotic Microbiology*, 64(1), 31–44. <https://doi.org/10.1111/jeu.12334>
- Dabbagh, N., & Preisfeld, A. (2018). Intrageneric variability between the chloroplast genomes of *Trachelomonas grandis* and *Trachelomonas volvocina* and phylogenomic analysis of phototrophic euglenoids. *Journal of Eukaryotic Microbiology*, 65(5), 648–660. <https://doi.org/10.1111/jeu.12510>
- Deschamps, P., & Moreira, D. (2012). Reevaluating the green contribution to diatom genomes. *Genome Biology and Evolution*, 4(7), 683–688. <https://doi.org/10.1093/gbe/evs053>
- Dobáková, E., Flegontov, P., Skalický, T., & Lukeš, J. (2015). Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. *Genome Biology and Evolution*, 7(12), 3358–3367. <https://doi.org/10.1093/gbe/evv229>
- Doetsch, N. A., Thompson, M. D., & Hallick, R. B. (1998). A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg? *Mol Biol Evol*, 15(1), 76–86. <https://doi.org/10.1093/oxfordjournals.molbev.a025850>

- Doolittle, W. F. (1998). You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics* 14(8), 307-311. [https://doi.org/10.1016/S0168-9525\(98\)01494-2](https://doi.org/10.1016/S0168-9525(98)01494-2)
- Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., Brillet-Guéguen, L., Freeman, K. D., Richter, D. J., & Bowler, C. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *ELife*, 6, 1–45. <https://doi.org/10.7554/eLife.23717>
- Dorrell, R. G., & Howe, C. J. (2015). Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), 10247–10254. <https://doi.org/10.1073/pnas.1421380112>
- Dorrell, R. G., & Smith, A. G. (2011). Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryotic Cell*, 10(7), 856–868. <https://doi.org/10.1128/EC.00326-10>
- Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., Wu, X., Reith, M., Cavalier-Smith, T., & Maier, U. G. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, 410(6832), 1091–1096. <https://doi.org/10.1038/35074092>
- Ebenezer, T.E., Carrington, M., Lebert, M., Kelly, S., & Field, M. C. (2017). *Euglena gracilis* Genome and Transcriptome: Organelles, Nuclear Genome Assembly Strategies and Initial Features. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, 979, 125–140. Springer Cham., 125–140. doi:10.1007/978-3-319-54910-1\_7
- Ebenezer, T. E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A. M. G., Prasad, B., Soukal, P., Santana-Molina, C., O'Neill, E., Nankissoor, N. N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A. P., Devos, D. P., Lukeš, J., Lebert, M., Vaughan, S., ... Field, M. C. (2019). Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology*, 17(1), 1–23. <https://doi.org/10.1186/s12915-019-0626-8>
- Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., Bryant, D., Steel, M. A., Lockhart, P. J., Penny, D., & Martin, W. (2004). A genome phylogeny for mitochondria among  $\alpha$ -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular Biology and Evolution*, 21(9), 1643–1660. <https://doi.org/10.1093/molbev/msh160>
- Faktorová, D., Dobáková, E., Peña-Díaz, P., & Lukeš, J. (2016). From simple to supercomplex: Mitochondrial genomes of euglenozoan protists. *F1000Research*, 5(0), 1–9. <https://doi.org/10.12688/f1000research.8040.2>
- Flegontov, P., Michálek, J., Janouškovec, J., Lai, D. H., Jirků, M., Hajdušková, E., Tomčala, A., Otto, T. D., Keeling, P. J., Pain, A., Oborník, M., & Lukeš, J. (2015). Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. *Molecular Biology and Evolution*, 32(5), 1115–1131. <https://doi.org/10.1093/molbev/msv021>
- Flori, S., Jouneau, P. H., Finazzi, G., Maréchal, E., & Falconet, D. (2016). Ultrastructure of the periplastidial compartment of the diatom *Phaeodactylum tricornutum*. *Protist*, 167(3), 254–267. <https://doi.org/10.1016/j.protis.2016.04.001>

- Füssy, Z., & Oborník, M. (2018). Complex endosymbioses I: From primary to complex plastids, multiple independent events. *Methods in Molecular Biology*, 1829, 17-35. [https://doi.org/10.1007/978-1-4939-8654-5\\_2](https://doi.org/10.1007/978-1-4939-8654-5_2)
- Gabalón, T., & Huynen, M. A. (2003). Reconstruction of the proto-mitochondrial metabolism. *Science* 301(5633), 609. <https://doi.org/10.1126/science.1085463>
- Gawryluk, R. M. R., & Gray, M. W. (2010). An ancient fission of mitochondrial cox1. *Molecular Biology and Evolution*, 27(1), 7–10. <https://doi.org/10.1093/molbev/msp223>
- Gibbs, S. P. (1978). The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Canadian Journal of Botany*, 56(22), 2883–2889. <https://doi.org/10.1139/b78-345>
- Gile, G. H., & Slamovits, C. H. (2014). Transcriptomic analysis reveals evidence for a cryptic plastid in the colpodellid *Voromonas pontica*, a close relative of chromerids and apicomplexan parasites. *PLoS ONE*, 9(5): e96258. <https://doi.org/10.1371/journal.pone.0096258>
- Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J., & McFadden, G. I. (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 103(25), 9566–9571. <https://doi.org/10.1073/pnas.0600707103>
- Gockel, G., Hachtel, W., & Melkonian, M. (2000). Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist* 151(4):347-351. [https://doi.org/10.1078/S1434-4610\(04\)70033-4](https://doi.org/10.1078/S1434-4610(04)70033-4)
- Gonczarowska-Jorge, H., Zahedi, R. P., & Sickmann, A. (2017). The proteome of baker's yeast mitochondria. *Mitochondrion*, 33, 15–21. <https://doi.org/10.1016/j.mito.2016.08.007>
- Gray, M. W. (2017). Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Molecular Biology of the Cell*, 28(10), 1285-1287. <https://doi.org/10.1091/mbc.E16-07-0509>
- Gray, M. W., & Archibald, J. M. (2012). Origins of mitochondria and plastids. In: Bock R., Knoop V. (Eds.) *Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration*, 35, 1–30. [https://doi.org/10.1007/978-94-007-2920-9\\_1](https://doi.org/10.1007/978-94-007-2920-9_1)
- Green, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant Journal*, 66(1), 34–44. <https://doi.org/10.1111/j.1365-313X.2011.04541.x>
- Griffin, N. J., & Aken, M. E. (1990). Interactions between the prasinophycean tidal pool flagellates *Pyramimonas parkeae* and *Tetraselmis apiculata* in batch culture. *South African Journal of Botany*, 56(6), 593–598. [https://doi.org/10.1016/s0254-6299\(16\)30996-6](https://doi.org/10.1016/s0254-6299(16)30996-6)
- Gualberto, J. M., Milesina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., & Dietrich, A. (2014). The plant mitochondrial genome: Dynamics and maintenance. *Biochimie*, 100(1), 107–120. <https://doi.org/10.1016/j.biochi.2013.09.016>
- Guillou, L., Eikrem, W., Chrétiennot-Dinet, M.-J., Le Gall, F., Massana, R., Romari, K., Pedrós-Alió, C., & Vaultot, D. (2004). Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist*, 155(June), 193–214. <https://doi.org/10.1078/143446104774199592>

- Hadariová, L., Vesteg, M., Hampl, V., & Krajčovič, J. (2018). Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Current Genetics* 64(2), 365–387. <https://doi.org/10.1007/s00294-017-0761-0>
- Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A., & Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research*, 21(15), 3537–3544. <https://doi.org/10.1093/nar/21.15.3537>
- Hammond, M. J., Nenarokova, A., Butenko, A., Dobáková, E. L., Field, M. C., Lukeš, J., Zoltner, M., Dobáková, E. L., Field, M. C., & Lukeš, J. (2020). A Uniquely Complex Mitochondrial Proteome from *Euglena gracilis*. *Molecular Biology and Evolution*, msaa061, 1–19. <https://doi.org/10.1093/molbev/msaa061>
- Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J. P., Rigden, D. J., Michels, P. A. M., & Opperdoes, F. R. (2003). Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 1067–1071. <https://doi.org/10.1073/pnas.0335769100>
- Hehenberger, E., Imanian, B., Burki, F., & Keeling, P. J. (2014). Evidence for the retention of two evolutionary distinct plastids in dinoflagellates with diatom endosymbionts. *Genome Biology and Evolution*, 6(9), 2321–2334. <https://doi.org/10.1093/gbe/evu182>
- Hempel, F., Bullmann, L., Lau, J., Zauner, S., & Maier, U. G. (2009). ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Molecular Biology and Evolution*, 26(8), 1781–90. <https://doi.org/10.1093/molbev/msp079>
- Henze, K., Badr, A., Wettern, M., Cerff, R., & Martin, W. (1995). A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 92(20), 9122–9126. <https://doi.org/10.1073/pnas.92.20.9122>
- Hewadikaramge, M., & Linton, E. (2018). Intrageneric chloroplast genome comparison in the genus *Euglena* (Phylum: Euglenophyta) with annotated chloroplast genomes of *Euglena hiemalis* and *Euglena clara*. *Journal of Applied Phycology*, 30, 3167–3177. <https://doi.org/10.1007/s10811-018-1547-6>
- Hirakawa, Y. (2017). Chlorarachniophytes with complex secondary plastids of green algal origin. In Y. Hirakawa (Eds.), *Secondary endosymbioses. Advances in Botanical Research*, 84 (359–393). <https://doi.org/10.1016/bs.abr.2017.06.006>
- Hoffmeister, M., Van Der Klies, A., Rotte, C., Van Grinsven, K. W. A. A., Van Hellemond, J. J., Henze, K., Tielens, A. G. M. M., & Martin, W. (2004). *Euglena gracilis* rhodoquinone:ubiquinone ratio and mitochondrial proteome differ under aerobic and anaerobic conditions. *Journal of Biological Chemistry*, 279(21), 22422–22429. <https://doi.org/10.1074/jbc.M400913200>
- Hong, L., & Hallick, R. B. (1994). A group III intron is formed from domains of two individual group II introns. *Genes and Development*, 8(13), 1589–1599. <https://doi.org/10.1101/gad.8.13.1589>
- Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012). The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS ONE*, 7(3), e33746. <https://doi.org/10.1371/journal.pone.0033746>

- Hrdá, Š., Hroudová, M., Vlček, Č., & Hampl, V. (2017). Mitochondrial genome of prasinophyte alga *Pyramimonas parkeae*. *Journal of Eukaryotic Microbiology*, 64(3), 360–369. <https://doi.org/10.1111/jeu.12371>
- Imanian, B., Carpenter, K. J., & Keeling, P. J. (2007). Mitochondrial genome of a tertiary endosymbiont retains genes for electron transport proteins. *Journal of Eukaryotic Microbiology*, 54(2), 146–153. <https://doi.org/10.1111/j.1550-7408.2007.00245.x>
- Imanian, B., Pombert, J. F., & Keeling, P. J. (2010). The complete plastid genomes of the two “Dinotoms” *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS ONE*, 5(5). <https://doi.org/10.1371/journal.pone.0010711>
- Inui, H., Ishikawa, T., & Tamoi, M. (2017). Wax ester fermentation and its application for biofuel production. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, 979, 269–283. Springer Cham. [https://doi.org/10.1007/978-3-319-54910-1\\_13](https://doi.org/10.1007/978-3-319-54910-1_13)
- Ishida, K. I., Cao, Y., Hasegawa, M., Okada, N., & Hara, Y. (1997). The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. *Journal of Molecular Evolution*, 45(6), 682–687. <https://doi.org/10.1007/PL00006272>
- Jackson, C., Clayden, S., & Reyes-Prieto, A. (2015). The Glaucophyta: The blue-green plants in a nutshell. *Acta Societatis Botanicorum Poloniae*, 84(2), 149–165. <https://doi.org/10.5586/asbp.2015.020>
- Jackson, C., Knoll, A. H., Chan, C. X., & Verbruggen, H. (2018). Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Scientific Reports*, 8(1), 1–10. <https://doi.org/10.1038/s41598-017-18805-w>
- Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., Bright, K. J., Imanian, B., Strom, S. L., Delwiche, C. F., Waller, R. F., Fensome, R. A., Leander, B. S., Rohwer, F. L., & Saldarriaga, J. F. (2017). Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America*, 114(2), E171–E180. <https://doi.org/10.1073/pnas.1614842114>
- Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24), 10949–10954. <https://doi.org/10.1073/pnas.1003335107>
- Janouškovec, J., Liu, S. L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., & Keeling, P. J. (2013). Evolution of red algal plastid genomes: Ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS ONE*, 8(3), e59001. <https://doi.org/10.1371/journal.pone.0059001>
- Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A. T., Kolísko, M., Mylnikov, A. P., & Keeling, P. J. (2015). Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), 10200–10207. <https://doi.org/10.1073/pnas.1423790112>

- Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A. T., Rohwer, F. L., Mylnikov, A. P., & Keeling, P. J. (2017). A new lineage of eukaryotes illuminates early mitochondrial genome reduction. *Current Biology*, 27(23), 3717–3724.e5. <https://doi.org/10.1016/j.cub.2017.10.051>
- Jenkins, K. P., Hong, L., & Hallick, R. B. (1995). Alternative splicing of the *Euglena gracilis* chloroplast roaA transcript. *Rna* 1(6), 624–633. PMID: 7489521
- Jiroutová, K., Kořený, L., Bowler, C., & Oborník, M. (2010). A gene in the process of endosymbiotic transfer. *PLoS ONE*, 5(10), e13234. <https://doi.org/10.1371/journal.pone.0013234>
- John, U., Lu, Y., Wohlrab, S., Groth, M., Janouškovec, J., Kohli, G. S., Mark, F. C., Bickmeyer, U., Farhat, S., Felder, M., Frickenhaus, S., Guillou, L., Keeling, P. J., Moustafa, A., Porcel, B. M., Valentin, K., & Glöckner, G. (2019). An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Science Advances*, 5(4). <https://doi.org/10.1126/sciadv.aav1110>
- Kamikawa, R., Tanifuji, G., Kawachi, M., Miyashita, H., Hashimoto, T., & Inagaki, Y. (2015). Plastid genome-based phylogeny pinpointed the origin of the green-colored plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Genome Biology and Evolution*, 7(4), 1133–1140. <https://doi.org/10.1093/gbe/evv060>
- Karlberg, O., Canbäck, B., Kurland, C. G., & Andersson, S. G. E. (2000). The dual origin of the yeast mitochondrial proteome. *Yeast*, 1(3), 170–187. [https://doi.org/10.1002/1097-0061\(20000930\)17:3<170::aid-yea25>3.0.co;2-v](https://doi.org/10.1002/1097-0061(20000930)17:3<170::aid-yea25>3.0.co;2-v)
- Karnkowska, A., Bennett, M. S., & Triemer, R. E. (2018). Dynamic evolution of inverted repeats in Euglenophyta plastid genomes. *Scientific Reports*, 8(1), 16071. <https://doi.org/10.1038/s41598-018-34457-w>
- Karnkowska, A., Bennett, M. S., Watza, D., Kim, J. I., Zakryś, B., & Triemer, R. E. (2015). Phylogenetic relationships and morphological character evolution of photosynthetic euglenids (excavata) inferred from taxon-rich analyses of five genes. *Journal of Eukaryotic Microbiology*, 62(3), 362–373. <https://doi.org/10.1111/jeu.12192>
- Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S. C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L. D., Herman, E. K., Soukal, P., Hroudová, M., Doležal, P., Stairs, C. W., Roger, A. J., Eliáš, M., Dacks, J. B., Vlček, Č., & Hampl, V. (2016). A eukaryote without a mitochondrial organelle. *Current Biology*, 26(10), 1274–1284. <https://doi.org/10.1016/j.cub.2016.03.053>
- Kasiborski, B. A., Bennett, M. S., & Linton, E. W. (2016). The chloroplast genome of *Phacus orbicularis* (Euglenophyceae): an initial datum point for the phacaceae. *Journal of Phycology*, 52(3), 404–411. <https://doi.org/10.1111/jpy.12403>
- Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), 729–748. <https://doi.org/10.1098/rstb.2009.0103>
- Kim, J. I., Linton, E. W., & Shin, W. (2015). Taxon-rich multigene phylogeny of the photosynthetic euglenoids (Euglenophyceae). *Frontiers in Ecology and Evolution*, 3(AUG). <https://doi.org/10.3389/fevo.2015.00098>

- Kim, J. I., Shin, W., & Triemer, R. E. (2010). Multigene analyses of photosynthetic euglenoids and new family, phacaceae (euglenales). *Journal of Phycology*, 46(6), 1278–1287. <https://doi.org/10.1111/j.1529-8817.2010.00910.x>
- Kim, J. I., Yoon, H. S., Yi, G., Kim, H. S., Yih, W., & Shin, W. S. (2015). The plastid genome of the cryptomonad *Teleaulax amphioxeia*. *PLoS ONE*, 10(6), 1–17. <https://doi.org/10.1371/journal.pone.0129284>
- Kořený, L., & Oborník, M. (2011). Sequence evidence for the presence of two tetrapyrrole pathways in *Euglena gracilis*. *Genome Biology and Evolution*, 3(1), 359–364. <https://doi.org/10.1093/gbe/evr029>
- Kosmala, S., Karnkowska, A., Milanowski, R., Kwiatowski, J., & Zakryś, B. (2005). Phylogenetic and taxonomic position of *Lepocinclis fusca* comb. nov. (= *Euglena fusca*) (Euglenaceae): Morphological and molecular justification. *Journal of Phycology*, 41(6), 1258–1267. <https://doi.org/10.1111/j.1529-8817.2005.00141.x>
- Krajčovič, J., Vesteg, M. & Schwartzbach, S. D. (2015). Euglenoid flagellates: A multifaceted biotechnology platform. *Journal of Biotechnology*, 20(202), 135–45. <https://doi.org/10.1016/j.jbiotec.2014.11.035>
- Kuo, R. C., Zhang, H., Zhuang, Y., Hannick, L., & Lin, S. (2013). Transcriptomic study reveals widespread spliced leader trans-splicing, short 5'-UTRs and potential complex carbon fixation mechanisms in the euglenoid Alga *Eutreptiella* sp. *PLoS ONE*, 8(4), e60826. <https://doi.org/10.1371/journal.pone.0060826>
- Lakey, B., & Triemer, R. (2017). The tetrapyrrole synthesis pathway as a model of horizontal gene transfer in euglenoids. *Journal of Phycology*, 53(1), 198–217. <https://doi.org/10.1111/jpy.12491>
- Lang, B. F., Laforest, M.-J. J., & Burger, G. (2007). Mitochondrial introns: a critical view. *Trends in Genetics*, 23(3), 119–125. <https://doi.org/10.1016/j.tig.2007.01.006>
- Lang, B. F., & Nedelcu, A. M. (2012). Plastid Genomes of Algae. In R. Bock & V. Knoop (Eds.), *Genomics of Chloroplasts and Mitochondria. Advances in Photosynthesis and Respiration*, 35 (59–87). Springer Netherlands. [https://doi.org/10.1007/978-94-007-2920-9\\_3](https://doi.org/10.1007/978-94-007-2920-9_3)
- Larkum, A. W. D., Lockhart, P. J., & Howe, C. J. (2007). Shopping for plastids. *Trends in Plant Science*, 12(5), 189–195. <https://doi.org/10.1016/j.tplants.2007.03.011>
- Lax, G., Lee, W. J., Eglit, Y., & Simpson, A. (2019). Ploetids represent much of the phylogenetic diversity of Euglenids. *Protist*, 170(2), 233–257. <https://doi.org/10.1016/j.protis.2019.03.001>
- Leander, B. S. (2004). Did trypanosomatid parasites have photosynthetic ancestors? *Trends in Microbiology*, 12(6), 251–258. <https://doi.org/10.1016/j.tim.2004.04.001>
- Leander, B. S., Esson, H. J., & Breglia, S. A. (2007). Macroevolution of complex cytoskeletal systems in euglenids. *BioEssays*, 29(10), 987–1000. <https://doi.org/10.1002/bies.20645>
- Leander, B. S., Lax, G., Karnkowska, A., & Simpson, A. G. B. (2017). Euglenida. In J.M. Archibald et al. (Eds.), *Handbook of the Protists*, 2nd edition. Springer. <https://doi.org/10.1007/978-3-319-32669-6>



- Leander, B. S., Triemer, R. E., & Farmer, M. A. (2001). Character evolution in heterotrophic euglenids. *European Journal of Protistology*, 37(3), 337–356. <https://doi.org/10.1078/0932-4739-00842>
- Lee, C. P., Taylor, N. L., & Harvey Millar, A. (2013). Recent advances in the composition and heterogeneity of the *Arabidopsis* mitochondrial proteome. *Frontiers in plant science*, 4, 4. <https://doi.org/10.3389/fpls.2013.00004>
- Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., & De Clerck, O. (2012). Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences*, 31(1), 1–46. <https://doi.org/10.1080/07352689.2011.615705>
- Lemieux, C., Otis, C., & Turmel, M. (2014a). Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evolutionary Biology*, 14(1), 211. <https://doi.org/10.1186/s12862-014-0211-2>
- Lemieux, C., Otis, C., & Turmel, M. (2014b). Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics*, 15, 857. <https://doi.org/10.1186/1471-2164-15-857>
- Lhee, D., Ha, J. S., Kim, S., Park, M. G., Bhattacharya, D., & Yoon, H. S. (2019). Evolutionary dynamics of the chromatophore genome in three photosynthetic *Paulinella* species. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-38621-8>
- Li, L., Wang, S., Wang, H., Sahu, S. K., Marin, B., Li, H., Xu, Y., Liang, H., Li, Z., Cheng, S., Reder, T., Çebi, Z., Wittek, S., Petersen, M., Melkonian, B., Du, H., Yang, H., Wang, J., Wong, G. K. S., ... Liu, H. (2020). The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nature Ecology and Evolution*. <https://doi.org/10.1038/s41559-020-1221-7>
- Lim, L., & McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), 749–763. <https://doi.org/10.1098/rstb.2009.0273>
- Linton, E. W., Karnkowska-Ishikawa, A., Kim, J. I., Shin, W., Bennett, M. S., Kwiatowski, J., Zakryś, B., & Triemer, R. E. (2010). Reconstructing euglenoid evolutionary relationships using three genes: Nuclear SSU and LSU, and chloroplast SSU rDNA sequences and the description of *Euglenaria* gen. nov. (Euglenophyta). *Protist*, 161(4), 603–619. <https://doi.org/10.1016/j.protis.2010.02.002>
- Lloyd, A. H., Rousseau-Gueutin, M., Timmis, J. N., Sheppard, A. E., & Ayliffe, M. A. (2012). Promiscuous Organellar DNA. In R. Bock & V. Knoop (Eds.), *Genomics of Chloroplasts and Mitochondria. Advances in Photosynthesis and Respiration*, 35 (201–221). Springer Netherlands. [https://doi.org/10.1007/978-94-007-2920-9\\_9](https://doi.org/10.1007/978-94-007-2920-9_9)
- Löffelhardt, W. (2014). The single primary endosymbiotic event. In W. Löffelhardt (Eds.) *Endosymbiosi*, 39–52. Springer, Vienna. [https://doi.org/10.1007/978-3-7091-1303-5\\_3](https://doi.org/10.1007/978-3-7091-1303-5_3)
- Lopes Dos Santos, A., Pollina, T., Gourvil, P., Corre, E., Marie, D., Garrido, J. L., Rodríguez, F., Noël, M. H., Vaultot, D., & Eikrem, W. (2017). Chloropicophyceae, a new class of picophytoplanktonic prasinophytes. *Scientific Reports*, 7(1), 1–20. <https://doi.org/10.1038/s41598-017-12412-5>

- Lukešová, S., Karlicki, M., Tomečková Hadariová, L., Szabová, J., Karnkowska, A., & Hampl, V. (2019). Analyses of environmental sequences and two regions of chloroplast genomes revealed the presence of new clades of photosynthetic euglenids in marine environments. *Environmental Microbiology Reports*, *12*(1), 78–91. <https://doi.org/10.1111/1758-2229.12817>
- Mahapatra, D. M., Chanakya, H. N., & Ramachandra, T. V. (2013). *Euglena* sp. as a suitable source of lipids for potential use as biofuel and sustainable wastewater treatment. *Journal of Applied Phycology* *25*, 855–865. <https://doi.org/10.1007/s10811-013-9979-5>
- Makai, S., Li, X., Hussain, J., Cui, C., Wang, Y., Chen, M., Yang, Z., Ma, C., Guo, A. Y., Zhou, Y., Chang, J., Yang, G., & He, G. (2015). A census of nuclear cyanobacterial recruits in the plant kingdom. *PLoS ONE*, *10*(3):e0120527. <https://doi.org/10.1371/journal.pone.0120527>
- Marcotte, E. M., Xenarios, I., Van Der Blik, A. M., & Eisenberg, D. (2000). Localizing proteins in the cell from their phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 12115–12120. <https://doi.org/10.1073/pnas.220399497>
- Margulis, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, *14*(3), 225–274. [https://doi.org/10.1016/0022-5193\(67\)90079-3](https://doi.org/10.1016/0022-5193(67)90079-3)
- Marin, B. (2004). Origin and fate of chloroplasts in the euglenoida. *Protist*, *155*(1), 13–14. <https://doi.org/10.1078/1434461000159>
- Marin, B., Palm, A., Klingberg, M., & Melkonian, M. (2003). Phylogeny and taxonomic revision of plastid-containing Euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist*, *154*(1), 99–145. <https://doi.org/10.1078/143446103764928521>
- Markunas, C. M., & Triemer, R. E. (2016). Evolutionary history of the enzymes involved in the Calvin-Benson cycle in Euglenids. *Journal of Eukaryotic Microbiology*, *63*(3), 326–339. <https://doi.org/10.1111/jeu.12282>
- Martijn, J., Vosseberg, J., Guy, L., Offre, P., & Ettema, T. J. G. (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*, *557*, 101–105. <https://doi.org/10.1038/s41586-018-0059-5>
- Martin, W., & Borst, P. (2003). Secondary loss of chloroplasts in trypanosomes. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, *100*(3), 765–767. <https://doi.org/10.1073/pnas.0437776100>
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., & Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(19), 12246–12251. <https://doi.org/10.1073/pnas.182432999>
- Martínez-Alberola, F., Barreno, E., Casano, L. M., Gasulla, F., Molins, A., & del Campo, E. M. (2019). Dynamic evolution of mitochondrial genomes in Trebouxiophyceae, including the first completely assembled mtDNA from a lichen-symbiont microalga (*Trebouxia* sp. TR9). *Scientific Reports*, *9*(1). <https://doi.org/10.1038/s41598-019-44700-7>

- Maruyama, S., Suzaki, T., Weber, A. P. M., Archibald, J. M., & Nozaki, H. (2011). Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology*, *11*(1), 105. <https://doi.org/10.1186/1471-2148-11-105>
- Matsuda, F., Hayashi, M., & Kondo, A. (2011). Comparative profiling analysis of central metabolites in *Euglena gracilis* under various cultivation conditions. *Bioscience, Biotechnology and Biochemistry*, *75*(11), 2253–2256. <https://doi.org/10.1271/bbb.110482>
- McFadden, G. I., & Van Dooren, G. G. (2004). Evolution: Red algal genome affirms a common origin of all plastids. *Current Biology*, *14*(13), 514–516. <https://doi.org/10.1016/j.cub.2004.06.041>
- McLachlan, J. L., Seguel, M. R., & Fritz, L. (1994). *Tetretreptia pomquetensis* gen. et. sp. nov. (Euglenophyceae): a quadriflagellate, phototropic, marine euglenoid. *Journal of Phycology*, *30*(3), 538–544. <https://doi.org/10.1111/j.0022-3646.1994.00538.x>
- Méndez-Leyva, A. B., Guo, J., Mudd, E. A., Wong, J., Schwartz, J. M., & Day, A. (2019). The chloroplast genome of the marine microalga *Tisochrysis lutea*. *Mitochondrial DNA Part B: Resources*, *4*(1), 253–255. <https://doi.org/10.1080/23802359.2018.1547140>
- Mereschkowsky, K. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Centralbl.*, *25*: 593–604. English translation in Martin, W., Kowallik, K. V. (1999). Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. *European Journal of Phycology*, *34*(3), 287–295. <https://doi.org/10.1080/09670269910001736342>
- Michel, F., & Ferat, J. L. (1995). Structure and activities of group II introns. *Annual Review of Biochemistry*, *64*, 435–461. <https://doi.org/10.1146/annurev.biochem.64.1.435>
- Michel, F., Umesono, K., & Ozeki, H. (1989). Comparative and functional anatomy of group II introns--a review. *Gene*, *82*(1), 5–30. [https://doi:10.1016/0378-1119\(89\)90026-7](https://doi:10.1016/0378-1119(89)90026-7).
- Milanowski, R., Gumińska, N., Karnkowska, A., Ishikawa, T., & Zakryś, B. (2016). Intermediate introns in nuclear genes of euglenids - are they a distinct type? *BMC Evolutionary Biology*, *16*(1), 1–11. <https://doi.org/10.1186/s12862-016-0620-5>
- Milanowski, R., Karnkowska, A., Ishikawa, T., & Zakryś, B. (2014). Distribution of conventional and nonconventional introns in tubulin ( $\alpha$  and  $\beta$ ) genes of euglenids. *Molecular Biology and Evolution*, *31*(3), 584–593. <https://doi.org/10.1093/molbev/mst227>
- Minge, M. A., Shalchian-Tabrizi, K., Tørresen, O. K., Takishita, K., Probert, I., Inagaki, Y., Klaveness, D., & Jakobsen, K. S. (2010). A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evolutionary Biology*, *10*(1). <https://doi.org/10.1186/1471-2148-10-191>
- Moore, C. E., Curtis, B., Mills, T., Tanifuji, G., & Archibald, J. M. (2012). Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. *Genome Biology and Evolution*, *4*(11), 1162–1175. <https://doi.org/10.1093/gbe/evs090>
- Moore, R. B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J. S., Heimann, K., Šlapeta, J., Hoegh-Guldberg, O., Logsdon, J. M., & Carter, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, *451*(7181), 959–963. <https://doi.org/10.1038/nature06635>

- Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., & Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*, 324(5935), 1724–1726. <https://doi.org/10.1126/science.1172983>
- Müller, M., Mentel, M., van Hellemond, J. J., Henze, K., Woehle, C., Gould, S. B., Yu, R.-Y., van der Giezen, M., Tielens, A. G. M., & Martin, W. F. (2012). Biochemistry and evolution of anaerobic energy metabolism in Eukaryotes. *Microbiology and Molecular Biology Reviews*, 76(2), 444–495. <https://doi.org/10.1128/membr.05024-11>
- Nakayama, T., Takahashi, K., Kamikawa, R., Iwataki, M., Inagaki, Y., & Tanifuji, G. (2020). Putative genome features of relic green alga-derived nuclei in dinoflagellates and future perspectives as model organisms. *Communicative & Integrative Biology*, 13(1), 84–88. <https://doi.org/10.1080/19420889.2020.1776568>
- Nakazawa, M. (2017). C2 metabolism in *Euglena*. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, 979, 39–45. Springer Cham. [https://doi.org/10.1007/978-3-319-54910-1\\_3](https://doi.org/10.1007/978-3-319-54910-1_3)
- Nakazawa, M., Inui, H., Yamaji, R., Yamamoto, T., Takenaka, S., Ueda, M., Nakano, Y., & Miyatake, K. (2000). The origin of pyruvate: NADP<sup>+</sup> oxidoreductase in mitochondria of *Euglena gracilis*. *FEBS Letters*, 479(3), 155–156. [https://doi.org/10.1016/S0014-5793\(00\)01882-2](https://doi.org/10.1016/S0014-5793(00)01882-2)
- Nieuwenhuis, M., Van De Peppel, L. J. J., Bakker, F. T., Zwaan, B. J., Aanen, D. K., & Baldauf, S. (2019). Enrichment of G4DNA and a Large Inverted Repeat Coincide in the Mitochondrial Genomes of *Termitomyces*. *Genome Biology and Evolution*, 11(7), 1857–1869. <https://doi.org/10.1093/gbe/evz122>
- Novák Vanclová, A. M. G., Zoltner, M., Kelly, S., Soukal, P., Záhonová, K., Ebenezer, T. E., Dobáková, E. L., Eliáš, M., Lukeš, J., & Field, M. C. (2020). Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytologist* 225(4), 1578–1592. doi:10.1111/nph.16237
- Nowack, E. C. M., Melkonian, M., & Glöckner, G. (2008). Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by Eukaryotes. *Current Biology*, 18(6), 410–418. <https://doi.org/10.1016/j.cub.2008.02.051>
- Nowack, E. C. M., Price, D. C., Bhattacharya, D., Singer, A., Melkonian, M., & Grossman, A. R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), 12214–12219. <https://doi.org/10.1073/pnas.1608016113>
- O’Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., Zimba, P. V., Henrissat, B., & Field, R. A. (2015). The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular BioSystems*, 11(10), 2808–2820. <https://doi.org/10.1039/c5mb00319a>
- Oborník, M., & Lukeš, J. (2015). The organellar genomes of *Chromera* and *Vitrella*, the phototrophic relatives of apicomplexan parasites. *Annual Review of Microbiology*, 69, 129–44. <https://doi.org/10.1146/annurev-micro-091014-104449>

- Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohya, K., & Michel, F. (1993). Group-I introns in the Liverwort mitochondrial genome - the gene coding for subunit-1 of cytochrome-oxidase shares 5 intron positions with its fungal counterparts. *Nucleic Acids Research*, *21*(5), 1297–1305. <http://nar.oxfordjournals.org/content/21/5/1297.full.pdf>
- Patron, N. J., Waller, R. F., Archibald, J. M., & Keeling, P. J. (2005). Complex protein targeting to dinoflagellate plastids. *Journal of Molecular Biology*, *348*(4), 1015–1024. <https://doi.org/10.1016/j.jmb.2005.03.030>
- Peikert, C. D., Mani, J., Morgenstern, M., Käser, S., Knapp, B., Wenger, C., Harsman, A., Oeljeklaus, S., Schneider, A., & Warscheid, B. (2017). Charting organellar importomes by quantitative mass spectrometry. *Nature Communications*, *8*, 15272. <https://doi.org/10.1038/ncomms15272>
- Pérez-Brocal, V., Shahar-Golan, R., & Clark, C. G. (2010). A linear molecule with two large inverted repeats: The mitochondrial genome of the stramenopile *Proteromonas lacertae*. *Genome Biology and Evolution*, *2*(1), 257–266. <https://doi.org/10.1093/gbe/evq015>
- Perez, E., Lapaille, M., Degand, H., Cilibrasi, L., Villavicencio-Queijeiro, A., Morsomme, P., González-Halphen, D., Field, M. C., Remacle, C., Baurain, D., & Cardol, P. (2014). The mitochondrial respiratory chain of the secondary green alga *Euglena gracilis* shares many additional subunits with parasitic Trypanosomatidae. *Mitochondrion*, *19*(PB), 338–349. <https://doi.org/10.1016/j.mito.2014.02.001>
- Pombert, J. F., James, E. R., Janouškovec, J., & Keeling, P. J. (2012). Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the *Monomorpha aenigmatica* plastid genome. *PLoS ONE*, *7*(12), 338–349. <https://doi.org/10.1371/journal.pone.0053433>
- Ponce-Toledo, R. I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K., & Moreira, D. (2017). An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology*, *27*(3), 386–391. <https://doi.org/10.1016/j.cub.2016.11.056>
- Ponce-Toledo, R. I., López-García, P., & Moreira, D. (2019). Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytologist*, *224*(2), 618–624. <https://doi.org/10.1111/nph.15965>
- Ponce-Toledo, R. I., Moreira, D., López-García, P., & Deschamps, P. (2018). Secondary plastids of euglenids and chlorarachniophytes function with a mix of genes of red and green algal ancestry. *Molecular Biology and Evolution*, *35*(9), 2198–2204. <https://doi.org/10.1093/molbev/msy121>
- Qiu, H., Price, D. C., Weber, A. P. M., Facchinelli, F., Yoon, H. S., & Bhattacharya, D. (2013). Assessing the bacterial contribution to the plastid proteome. *Trends in Plant Science*, *18*(12), 680–687. <https://doi.org/10.1016/j.tplants.2013.09.007>
- Reyes-Prieto, A., Weber, A. P. M., & Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. *Annual Review of Genetics*, *41*(1), 147–168. <https://doi.org/10.1146/annurev.genet.41.110306.130134>
- Riisberg, I., Orr, R., Kluge, R., Shalchian-Tabrizi, K., Bowers, H., Patil, V., Edvardsen, B., & Jakobsen, K. (2009). Seven gene phylogeny of heterokonts. *Protist*, *160*, 191–204. <https://doi.org/10.1016/j.protis.2008.11.004>

- Roger, A. J., Muñoz-Gómez, S. A., & Kamikawa, R. (2017). The origin and diversification of mitochondria. *Current Biology*, 27(21), 1177–1192. <https://doi.org/10.1016/j.cub.2017.09.015>
- Rogers, M. B., Gilson, P. R., Su, V., McFadden, G. I., & Keeling, P. J. (2007). The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Molecular Biology and Evolution*, 24(1), 54–62. <https://doi.org/10.1093/molbev/msl129>
- Russell, A. G., Watanabe, Y. I., Charette, J. M., & Gray, M. W. (2005). Unusual features of fibrillarin cDNA and gene structure in *Euglena gracilis*: Evolutionary conservation of core proteins and structural predictions for methylation-guide box C/D snoRNPs throughout the domain Eucarya. *Nucleic Acids Research*, 33(9), 2781–2791. <https://doi.org/10.1093/nar/gki574>
- Santillán Torres, J. L., Atteia, A., Claros, M. G., & González-Halphen, D. (2003). Cytochrome f and subunit IV, two essential components of the photosynthetic bf complex typically encoded in the chloroplast genome, are nucleus-encoded in *Euglena gracilis*. *Biochimica et Biophysica Acta - Bioenergetics*, 1604(3), 180–189. [https://doi.org/10.1016/S0005-2728\(03\)00058-6](https://doi.org/10.1016/S0005-2728(03)00058-6)
- Sarai, C., Tanifuji, G., Nakayama, T., Kamikawa, R., Takahashi, K., Yazaki, E., Matsuo, E., Miyashita, H., Ishida, K. ichiro, Iwataki, M., & Inagaki, Y. (2020). Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 117(10), 5364–5375. <https://doi.org/10.1073/pnas.1911884117>
- Satjarak, A., Burns, J. A., Kim, E., & Graham, L. E. (2017). Complete mitochondrial genomes of prasinophyte algae *Pyramimonas parkeae* and *Cymbomonas tetramitiformis*. *Journal of Phycology*, 53(3), 601–615. <https://doi.org/10.1111/jpy.12521>
- Satjarak, A., & Graham, L. E. (2017a). Comparative DNA sequence analyses of *Pyramimonas parkeae* (Prasinophyceae) chloroplast genomes. *Journal of Phycology*, 53(2), 415–424. <https://doi.org/10.1111/jpy.12515>
- Satjarak, A., & Graham, L. E. (2017b). Genome-wide analysis of carbohydrate-active enzymes in *Pyramimonas parkeae* (Prasinophyceae). *Journal of Phycology*, 53(5), 1072–1086. <https://doi.org/10.1111/jpy.12566>
- Ševčíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., Jenkins, J., Schmutz, J., Příbyl, P., Fousek, J., Vlček, C., Lang, B. F., Oborník, M., Worden, A. Z., & Eliáš, M. (2015). Updating algal evolutionary relationships through plastid genome sequencing: Did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Scientific Reports*, 5(10134), 1–12. <https://doi.org/10.1038/srep10134>
- Sheiner, L., & Striepen, B. (2013). Protein sorting in complex plastids. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1833(2), 352–359. <https://doi.org/10.1016/j.bbamcr.2012.05.030>
- Sheveleva, E. V., & Hallick, R. B. (2004). Recent horizontal intron transfer to a chloroplast genome. *Nucleic Acids Research*, 32(2), 803–810. <https://doi.org/10.1093/nar/gkh225>
- Shigeoka, S., Onishi, T., Maeda, K., Nakano, Y., & Kitaoka, S. (1986). Occurrence of thiamin pyrophosphate-dependent 2-oxoglutarate decarboxylase in mitochondria of *Euglena gracilis*. *FEBS Letters*, 195(1–2), 43–47. [https://doi.org/10.1016/0014-5793\(86\)80126-0](https://doi.org/10.1016/0014-5793(86)80126-0)

- Singer, A., Poschmann, G., Mühlich, C., Valadez-Cano, C., Hänsch, S., Hüren, V., Rensing, S. A., Stühler, K., & Nowack, E. C. M. (2017). Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. *Current Biology*, 27(18), 2763–2773.e5. <https://doi.org/10.1016/j.cub.2017.08.010>
- Smith, D. R., & Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences*, 112(33), 10177–10184. <https://doi.org/10.1073/pnas.1422049112>
- Spencer, D. F., & Gray, M. W. (2011). Ribosomal RNA genes in *Euglena gracilis* mitochondrial DNA: fragmented genes in a seemingly fragmented genome. *Molecular Genetics and Genomics*, 285(1), 19–31. <https://doi.org/10.1007/s00438-010-0585-9>
- Stegemann, S., Hartmann, S., Ruf, S., & Bock, R. (2003). High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 8828–8833. <https://doi.org/10.1073/pnas.1430924100>
- Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q., & Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nature Communications*, 5(5764), 1–7. <https://doi.org/10.1038/ncomms6764>
- Sun, Y., & Pombert, J. F. (2014). Helicosporidia: A genomic snapshot of an early transition to parasitism. *Acta Societatis Botanicorum Poloniae*, 83(4), 377–385. <https://doi.org/10.5586/asbp.2014.039>
- Suzuki, K. (2017). Large-scale cultivation of *Euglena*. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, 979, 285–293. Springer Cham. [https://doi.org/10.1007/978-3-319-54910-1\\_14](https://doi.org/10.1007/978-3-319-54910-1_14)
- Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H., & Kreel, N. E. (2008). Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1504), 2629–40. <https://doi.org/10.1098/rstb.2008.0023>
- Tamaki, S., Nishino, K., Ogawa, T., Maruta, T., Sawa, Y., Arakawa, K., & Ishikawa, T. (2019). Comparative proteomic analysis of mitochondria isolated from *Euglena gracilis* under aerobic and hypoxic conditions. *PLoS ONE*, 14(12), 1–13. <https://doi.org/10.1371/journal.pone.0227226>
- Tanifuji, G., Onodera, N. T., Wheeler, T. J., Dlutek, M., Donaher, N., & Archibald, J. M. (2011). Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. *Genome Biology and Evolution*, 3, 44–54. <https://doi.org/10.1093/gbe/evq082>
- Thompson, M. D., Copertino, D. W., Thompson, E., Favreau, M. R., & Hallick, R. B. (1995). Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Research*, 23(23), 4745–4752. <https://doi.org/10.1093/nar/23.23.4745>
- Thorsness, P. E., & Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature*, 346, 376–379. <https://doi.org/10.1038/346376a0>

- Thrash, J. C., Boyd, A., Huggett, M. J., Grote, J., Carini, P., Yoder, R. J., Robbertse, B., Spatafora, J. W., Rappé, M. S., & Giovannoni, S. J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Scientific Reports*, *1*(13), 1–9. <https://doi.org/10.1038/srep00013>
- Tomiyama, T., Kurihara, K., Ogawa, T., Maruta, T., Ogawa, T., Ohta, D., Sawa, Y., & Ishikawa, T. (2017). Wax ester synthase/diacylglycerol acyltransferase isoenzymes play a pivotal role in wax ester biosynthesis in *euglena gracilis*. *Scientific Reports*, *7*(1), 1–13. <https://doi.org/10.1038/s41598-017-14077-6>
- Triemer, R. E., & Farmer, M. A. (1991). An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. In M. Melkonian, R. A. Andersen, & E. Schnepf (Eds.) *The Cytoskeleton of Flagellate and Ciliate Protists. Protozoology 164*, 91–104. Springer Vienna. [https://doi.org/10.1007/978-3-7091-6714-4\\_9](https://doi.org/10.1007/978-3-7091-6714-4_9)
- Triemer, R. E., & Lewandowski, C. L. (1994). Ultrastructure of the basal apparatus and putative vestigial feeding apparatuses in a quadriflagellate Euglenoid (Euglenophyta). *Journal of Phycology*, *30*(1), 28–38. <https://doi.org/10.1111/j.0022-3646.1994.00028.x>
- Triemer, R. E., Linton, E., Shin, W., Nudelman, A., Monfils, A., Bennett, M., & Brosnan, S. (2006). Phylogeny of the euglenales based upon combined SSU and LSU rDNA sequence comparisons and description of *Discoplastis* gen. nov. (Euglenophyta). *Journal of Phycology*, *42*(3), 731–740. <https://doi.org/10.1111/j.1529-8817.2006.00219.x>
- Turmel, M. (1999). The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: Two radically different evolutionary patterns within green algae. *The Plant Cell Online*, *11*(9), 1717–1730. <https://doi.org/10.1105/tpc.11.9.1717>
- Turmel, M., Dos Santos, A. L., Otis, C., Sergerie, R., & Lemieux, C. (2019). Tracing the evolution of the plastome and mitogenome in the chloropicophyceae uncovered convergent tRNA gene losses and a variant plastid genetic code. *Genome Biology and Evolution*, *11*(4), 1275–1292. <https://doi.org/10.1093/gbe/evz074>
- Turmel, M., Gagnon, M. C., O’Kelly, C. J., Otis, C., & Lemieux, C. (2009). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Molecular Biology and Evolution*, *26*(3), 631–648. <https://doi.org/10.1093/molbev/msn285>
- Turmel, M., & Lemieux, C. (2018). Evolution of the plastid genome in green algae. *Advances in Botanical Research*, *85*, 157–193. <https://doi.org/10.1016/bs.abr.2017.11.010>
- Turmel, M., Otis, C., & Lemieux, C. (2017). Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. *Scientific Reports*, *7*, 994. <https://doi.org/10.1038/s41598-017-01144-1>
- Turmel, M., Otis, Ch., & Lemieux, C. (1999). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(18), 10248–10253. <https://doi.org/10.1073/pnas.96.18.10248>
- Vanclová, A. M. G., Hadariová, L., Hrdá, Š., & Hampl, V. (2017). Secondary Plastids of Euglenophytes. In Y.Hirakawa (Eds.), *Secondary endosymbioses. Advances in Botanical Research*, *84*, 321–358. <https://doi.org/10.1016/bs.abr.2017.06.008>



- Vesteg, M., Vacula, R., Steiner, J. M., Mateášiková, B., Löffelhardt, W., Brejová, B., & Krajčovič, J. (2010). A possible role for short introns in the acquisition of stroma-targeting peptides in the flagellate *Euglena gracilis*. *DNA Research*, *17*(4), 223–231. <https://doi.org/10.1093/dnares/dsq015>
- Viprey, M., Guillou, L., Ferréol, M., & Vaultot, D. (2008). Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environmental Microbiology*, *10*(7), 1804–1822. <https://doi.org/10.1111/j.1462-2920.2008.01602.x>
- Waller, R. F., & Kořený, L. (2017). Plastid Complexity in Dinoflagellates: A Picture of Gains, Losses, Replacements and Revisions. In Y.Hirakawa (Eds.), *Secondary endosymbioses. Advances in Botanical Research*, *84*, 105–143. <https://doi.org/10.1016/bs.abr.2017.06.004>
- Walne, P. L., Moestrup, Ø., Norris, R. E., & Ettl, H. (1986). Light and electron microscopical studies of *Eutreptiella eupharyngea* sp. nov. (Euglenophyceae) from Danish and American waters. *Phycologia*, *25*(1), 109–126. <https://doi.org/10.2216/i0031-8884-25-1-109.1>
- Wang, B., Xue, J., Li, L., Liu, Y., & Qiu, Y. L. (2009). The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Current Genetics*, *55*(6), 601–609. <https://doi.org/10.1007/s00294-009-0273-7>
- Wang, W., Lanfear, R., & Gaut, B. (2019). Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biology and Evolution*, *11*(12), 3372–3381. <https://doi.org/10.1093/gbe/evz256>
- Wang, Z., & Wu, M. (2015). An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Scientific Reports*, *5*, 7949. <https://doi.org/10.1038/srep07949>
- Watanabe, F., Yoshimura, K., & Shigeoka, S. (2017). Biochemistry and physiology of vitamins in Euglena. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, *979*, 65–90. Springer Cham. [https://doi.org/10.1007/978-3-319-54910-1\\_5](https://doi.org/10.1007/978-3-319-54910-1_5)
- Wetherbee, R., Jackson, C. J., Repetti, S. I., Clementson, L. A., Costa, J. F., van de Meene, A., Crawford, S., & Verbruggen, H. (2019). The golden paradox – a new heterokont lineage with chloroplasts surrounded by two membranes. *Journal of Phycology*, *55*(2), 257–278. <https://doi.org/10.1111/jpy.12822>
- Wiegert, K. E., Bennett, M. S., & Triemer, R. E. (2012). Evolution of the chloroplast genome in photosynthetic Euglenoids: A comparison of *Eutreptia viridis* and *Euglena gracilis* (Euglenophyta). *Protist*, *163*(6), 832–843. <https://doi.org/10.1016/j.protis.2012.01.002>
- Wiegert, K. E., Bennett, M. S., & Triemer, R. E. (2013). Tracing patterns of chloroplast evolution in euglenoids: Contributions from *Colacium vesiculosum* and *Strombomonas acuminata* (Euglenophyta). *Journal of Eukaryotic Microbiology*, *60*(2), 214–221. <https://doi.org/10.1111/jeu.12025>
- Wise, R. R. (2006). The diversity of plastid form and function. In R. R. Wise & J. K. Hooper, *The Structure and Function of Plastids. Advances in Photosynthesis and Respiration*, *23*, 3–26. <https://doi.org/10.1007/978-1-4020-4061-0>

- Wisecaver, J. H., & Hackett, J. D. (2010). Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate *Dinophysis acuminata*. *BMC Genomics*, *11*(1), 366. <https://doi.org/10.1186/1471-2164-11-366>
- Woehle, C., Dagan, T., Martin, W. F., & Gould, S. B. (2011). Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. *Genome Biology and Evolution*, *3*(1), 1220–1230. <https://doi.org/10.1093/gbe/evr100>
- Wollman, F. A. (2016). An antimicrobial origin of transit peptides accounts for early endosymbiotic events. *Traffic*, *17*(12), 1322–1328. <https://doi.org/10.1111/tra.12446>
- Woo, Y. H., Ansari, H., Otto, T. D., Klinger, C. M., Kolísko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., Ali, S., Bernal, A., Del Campo, J., Cihlář, J., Flegontov, P., Gornik, S. G., Hajdušková, E., Horák, A., Janouškovec, J., ... Pain, A. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *ELife*, *4*, e06974, 1–41. <https://doi.org/10.7554/eLife.06974>
- Yamaguchi, A., Yubuki, N., & Leander, B. S. (2012). Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: Description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evolutionary Biology*, *12*(1). <https://doi.org/10.1186/1471-2148-12-29>
- Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J., & Woese, C. R. (1985). Mitochondrial origins. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(13), 4443–4447. <https://doi.org/10.1073/pnas.82.13.4443>
- Yang, Y., Maruyama, S., Sekimoto, H., Sakayama, H., & Nozaki, H. (2011). An extended phylogenetic analysis reveals ancient origin of “non-green” phosphoribulokinase genes from two lineages of “green” secondary photosynthetic eukaryotes: Euglenophyta and Chlorarachniophyta. *BMC Research Notes*, *4*(1), 330. <https://doi.org/10.1186/1756-0500-4-330>
- Yoo, Y. Du, Seong, K. A., Kim, H. S., Jeong, H. J., Yoon, E. Y., Park, J., Kim, J. I., Shin, W., & Palenik, B. (2018). Feeding and grazing impact by the bloom-forming euglenophyte *Eutreptiella eupharyngea* on marine eubacteria and cyanobacteria. *Harmful Algae*, *73*, 98–109. <https://doi.org/10.1016/j.hal.2018.02.003>
- Yoshida, Y., Tomiyama, T., Maruta, T., Tomita, M., Ishikawa, T., & Arakawa, K. (2016). De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*, *17*(1), 1–10. <https://doi.org/10.1186/s12864-016-2540-6>
- Yurchenko, T., Ševčíková, T., Strnad, H., Butenko, A., & Eliáš, M. (2016). The plastid genome of some eustigmatophyte algae harbours a bacteria-derived six-gene cluster for biosynthesis of a novel secondary metabolite. *Open Biology*, *6*(11), 160249. <https://doi.org/10.1098/rsob.160249>
- Záhonová, K., Füssy, Z., Birčák, E., Novák Vanclová, A. M. G., Klimeš, V., Vesteg, M., Krajčovič, J., Oborník, M., & Eliáš, M. (2018). Peculiar features of the plastids of the colourless alga *Euglena longa* and photosynthetic euglenophytes unveiled by transcriptome analyses. *Scientific Reports*, *8*(1), 1–15. <https://doi.org/10.1038/s41598-018-35389-1>

- Záhonová, K., Füssy, Z., Oborník, M., Eliáš, M., & Yurchenko, V. (2016). RuBisCO in Non-Photosynthetic Alga *Euglena longa*: Divergent Features, Transcriptomic Analysis and Regulation of Complex Formation. *PLOS ONE*, *11*(7), e0158790. <https://doi.org/10.1371/journal.pone.0158790>
- Zakryś, B., Milanowski, R., & Karnkowska, A. (2017). Evolutionary origin of *Euglena*. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, *979*, 3–17. Springer Cham. [https://doi.org/10.1007/978-3-319-54910-1\\_1](https://doi.org/10.1007/978-3-319-54910-1_1)
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., & Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, *541*(7637), 353–358. <https://doi.org/10.1038/nature21031>
- Zauner, S., Heimerl, T., Moog, D., & Maier, U. G. (2019). The known, the new, and a possible surprise: A re-evaluation of the nucleomorph-encoded proteome of Cryptophytes. *Genome Biology and Evolution*, *11*(6), 1618–1629. <https://doi.org/10.1093/gbe/evz109>
- Zimmerly, S., & Semper, C. (2015). Evolution of group II introns. *Mobile DNA*, *6*(1), 7. <https://doi.org/10.1186/s13100-015-0037-5>
- Zimorski, V., Rauch, C., van Hellemond, J. J., Tielens, A. G. M., & Martin, W. F. (2017). The mitochondrion of *Euglena gracilis*. In: Schwartzbach S., Shigeoka S. (Eds.) *Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology*, *979*, 19–37. Springer Cham. [https://doi.org/10.1007/978-3-319-54910-1\\_2](https://doi.org/10.1007/978-3-319-54910-1_2)

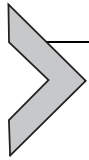
## **10 Příloha**

**Vanclová, A. M. G., Hadariová, L., Hrdá, Š., & Hampl, V. (2017).**

### **Secondary Plastids of Euglenophytes.**

**In Y.Hirakawa (Eds.), Secondary endosymbioses. Advances in Botanical Research, 84, 321–358.**

**<https://doi.org/10.1016/bs.abr.2017.06.008>**



# Secondary Plastids of Euglenophytes

Anna M.G. Vanclová<sup>1</sup>, Lucia Hadariová, Štěpánka Hrdá,  
Vladimír Hampl<sup>1</sup>

Faculty of Science, Charles University in Prague, Prague, Czech Republic

<sup>1</sup>Corresponding authors: e-mail address: vanclova@gmail.com; vlada@natur.cuni.cz

## Contents

1. Introduction: What Are the Euglenophytes and Why to Care About Them	322
2. Origin of Euglenophyte Plastids: Early or Late, Green or Red?	325
3. Plastid Morphology: Display of Diversity	328
4. Plastid Genomes: Sped-Up Evolution and Introns Gone Haywire	330
5. Plastid Biogenesis and Housekeeping: How to Make It and How to Control It	335
6. Plastid Metabolism: A Factory With Redundant Production Lines	340
7. Secondary Osmotrophy and Plastid Bleaching: Plastids That Forgot How to Plastid	343
8. Conclusions	347
Acknowledgements	347
References	347

## Abstract

Euglenophytes obtained their plastids from a primary green alga related to extant genus *Pyramimonas*. The relatively recent establishment of this new organelle is an intriguing evolutionary phenomenon worth studying and comparing with other secondary plastids with a regard to their similarities and differences. A remarkably fast evolution driven by rapid intron gain and diversification is observed in euglenid plastid genomes which often tend to swell in size and rearrange while keeping the gene content stable. As a result of the secondary endosymbiosis, the plastid is wrapped in an additional membrane which makes any protein, metabolite, or ion transporting routes more complicated. In the case of protein import, secretory pathway-derived, signal peptide-dependent mechanism involving the endoplasmic reticulum, Golgi, and vesicular transport were recruited. The plastid endosymbiosis also served as a source of various oddities concerning metabolic pathways as the new organelle contained some of the enzymes and pathways already present in the host. Thus, several cases of division of labour and specialization can be observed, as well as simple redundancies which might be in fact just transitory and will eventually disappear in the future course of evolution. Endosymbiotic and lateral gene transfers were quite common in the

ancestors of euglenophytes, especially in the case of plastid proteins many of which were demonstrated to have originated not only from the green-algal endosymbiont but also from a spectrum of nongreen lineages. The circumstances of the nongreen-algal gene gains are unclear. Another evolutionary phenomenon occurring in euglenophytes is the secondary loss of plastid or its photosynthetic capacity. This process gave rise to a number of distinct species which no longer possess the ability to photosynthesize. Interestingly, this “bleaching” process can be induced in the laboratory, enabling to study the process of plastid loss *in vitro*.

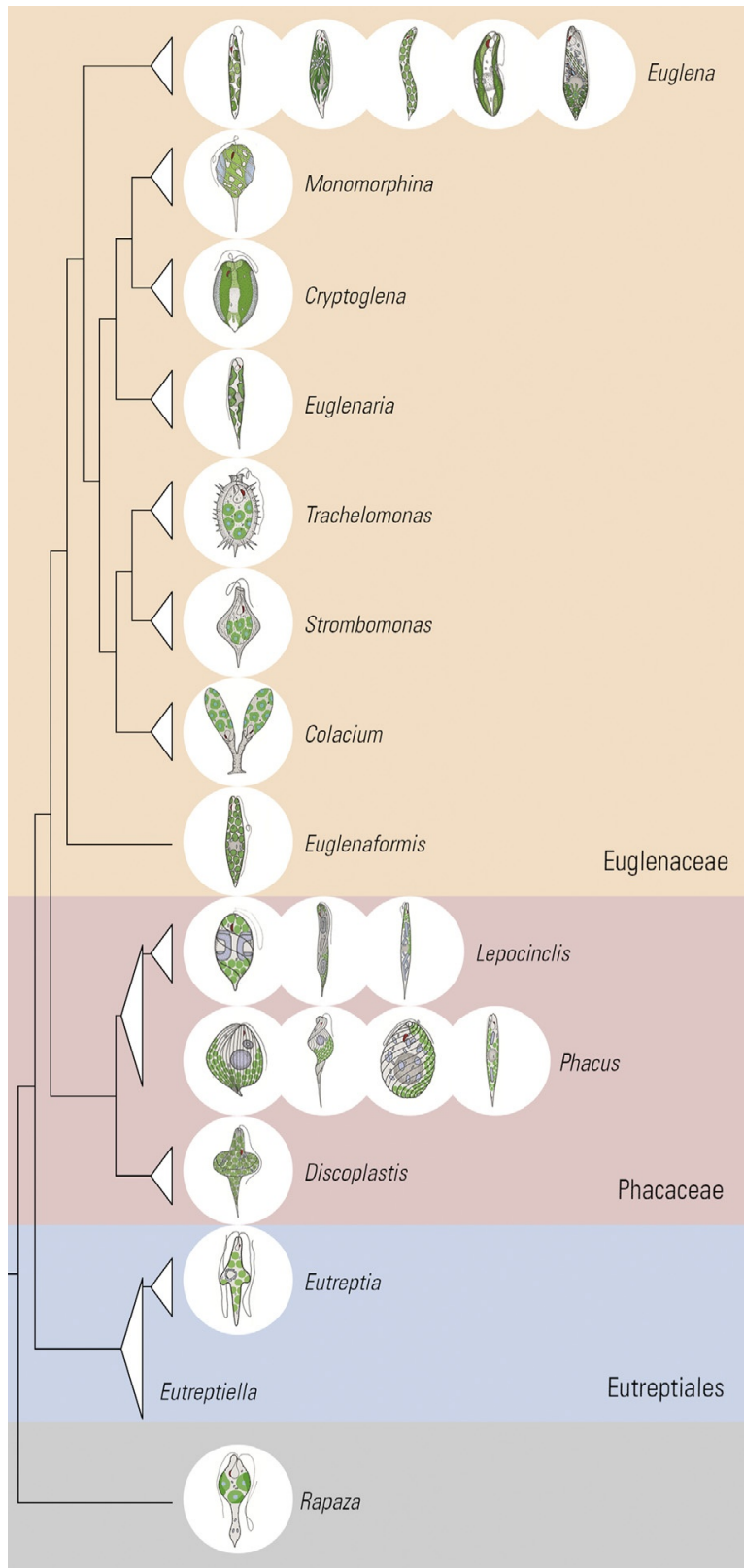


## 1. INTRODUCTION: WHAT ARE THE EUGLENOPHYTES AND WHY TO CARE ABOUT THEM

Euglenids are a group of flagellate protists belonging to the phylum Euglenozoa (Excavata), along with kinetoplastids—a group made famous by its important pathogenic members such as trypanosomes—diplonemids, and symbiontids (also termed Postgaurdea) (Adl et al., 2012; Cavalier-Smith, 2016). Euglenids are a highly diversified taxon in regard to their nutritional strategies. The ancestral mode of nutrition was probably bacteriovory which was replaced by eukaryovory in a large portion of the diversity; this shift is believed to have taken place once in the evolution. One lineage of eukaryovores then lost the phagotrophic ability and became osmotrophic. This lineage is currently termed primary osmotrophs (Leander, 2004; Leander, Esson, & Breglia, 2007; Leander, Triemer, & Farmer, 2001). However, some researchers propose an alternative hypothesis according to which eukaryovory was the ancestral nutritional strategy common not only to euglenids but also to kinetoplastids, diplomemids, and symbiontids, and all other feeding modes are derived from it (Cavalier-Smith, 2016). One lineage within eukaryovorous euglenids gained a green secondary plastid in presumably one endosymbiotic event. All these photoautotrophs form one robust phylogenetic clade termed the Euglenophyta. The term “euglenophytes” thus refers to phototrophic (or secondarily osmotrophic) euglenids that harbour (or once harboured) a plastid, while the term “euglenids” addresses the whole group of organisms of various nutritional strategies. The euglenophytes were one of the first protists to be discovered and described. The first documented observation of an euglenophyte, presumably the genus *Euglena*, and description of its typical slime-like mode of movement, metaboly, was carried out by John Harris at the end of the 17th century (Harris, 1695) and the first species of euglenophytes were described

in the beginning of the 19th century by [Ehrenberg \(1830\)](#). Several isolated lineages within euglenophytes lost the photosynthetic pigments, switching to osmotrophic mode of nutrition independently of the earlier mentioned primary osmotrophs; these organisms are termed secondary osmotrophs and retain a colourless plastid despite not using it for photosynthesis. Additionally, one obligatory–mixotrophic species, *Rapaza viridis*, has a photosynthetically active green plastid but at the same time requires green-algal (*Tetraselmis*) prey to survive and it is the only euglenid which combines photoautotrophy with phagotrophy. It is suspected to represent a link between the eukaryovorous euglenids and the phototrophic euglenophytes. This is also suggested by its phylogenetic position as sister lineage to all other euglenophytes ([Fig. 1](#); [Yamaguchi, Yubuki, & Leander, 2012](#)). However, it is still unclear whether the *R. viridis* plastid is stable or transient and if it is related to the plastids of other euglenophytes. The closest known non-photosynthetic sister of euglenophytes is the eukaryovorous genus *Teloprocta* (former *Heteronema*) *scaphurum* ([Cavalier-Smith, 2016](#); [Lax & Simpson, 2013](#)).

Euglenophytes have been intensively studied for a long time thanks to their relatively large size, distinctive and aesthetically pleasing appearance, and easy collection and cultivation, and so approximately 950 species of euglenophytes have been described to this day ([Guiry & Guiry, 2017](#)). The inner taxonomy of euglenophytes is currently relatively clearly resolved ([Fig. 1](#); [Karnkowska et al., 2015](#); [Kim, Linton, & Shin, 2015](#); [Kim & Shin, 2008](#); [Linton et al., 2010](#); [Triemer et al., 2006](#); [Zakryś, Milanowski, & Karnkowska, 2017](#)). The group splits into two orders: Eutreptiales and Euglenales. The Eutreptiales exhibit a number of features which are considered ancestral, such as the adaptation to marine environment and presence of two emergent flagella, and contain two genera: *Eutreptia* and *Eutreptiella*. The Euglenales are more diverse, they have only one emergent flagellum, and they are thought to be almost exclusively freshwater dwelling. The Eutreptiales are probably paraphyletic, with the Euglenales being their inner group ([Cavalier-Smith, 2016](#)). The Euglenales are further split into two distinctive families, Euglenaceae and Phacaceae. The Euglenaceae include eight genera: *Euglena*, *Euglenaria*, *Euglenaformis*, *Cryptoglana*, *Monomorpha*, *Colacium*, *Trachelomonas*, and *Strombomonas*. This group is highly diverse in regard to the shape, number, and position of the plastids, and also overall morphology of the cell which may exhibit various shapes—some members, i.e., genera *Trachelomonas* and *Strombomonas*, are even surrounded



**Fig. 1** Simplified phylogenetic tree of the 14 euglenophyte genera (including *Rapaza*) with examples of cell and plastid morphologies. One to five different illustrative species are shown for each genus, depending on its relative intrageneric plastid morphological diversity.



by mineralized extracellular lorica—and types and degrees of motility—some members have pellicles of low flexibility and prefer swimming using flagellum to metaboly, while some species in the genus *Colacium* are completely sessile. The Phacaceae include three genera: *Phacus*, *Lepocinclis*, and *Discoplastis*, the first two of which have a rigid pellicle, which is flattened or even helical in shape, and are not able to move using metaboly (Cavalier-Smith, 2016; de M. Bicudo & Menezes, 2016).

Euglenophytes are currently studied for their possible utilization in biotechnology and other applied sciences, especially as potential producers of biofuels and also nutritional supplements due to their capacity to synthesize various lipids and compounds with antioxidative properties such as tocopherols and carotenoids (Krajčovič, Vesteg, & Schwartzbach, 2015; Ogawa et al., 2015; Teerawanichpan & Qiu, 2010).



## 2. ORIGIN OF EUGLENOPHYTE PLASTIDS: EARLY OR LATE, GREEN OR RED?

Before the age of molecular phylogenetics, euglenophytes were traditionally classified as a subgroup or sister of green algae due to their plastid morphology and pigment content, especially the combination of chlorophylls *a* and *b*, which is otherwise unique to green plants and algae and chlorarachniophytes, and the chlorophylls *a*- and *b*-containing plastids were correctly assumed to have arisen only once in the evolution. However, it was suspected for a long time that something is amiss here because most of the nonplastid morphology, physiology, and biochemistry of these organisms were strikingly dissimilar to green algae. With the (re)invention of the endosymbiotic theory by Lynn Margulis in 1967 (Sagan, 1993) and its general acceptance, a completely new paradigm opened for evolutionary biology and taxonomy trying to resolve unclear phenomena and phylogenetic relationships such as this one. A decade later, the secondary endosymbiotic origin of the euglenophyte plastid was proposed by Gibbs (1978, 1981) and euglenids were placed along with kinetoplastids into a new phylum Euglenozoa by Cavalier-Smith (1981). The first sequenced genes from the model euglenid *Euglena gracilis* (Douglas & Turner, 1991; Morden & Golden, 1991), followed shortly by the complete sequencing of its plastid genome (Hallick et al., 1993), brought a molecular evidence for this claim. In 2009, a wider analysis of phylogenetic relationship within green algae with a focus on prasinophytes and also including sequences from secondary plastids was performed by Turmel, Gagnon, O'Kelly, Otis, and Lemieux

(2009) which resulted in pinpointing the origin of the euglenid plastid to the Pyramimonadales clade in the prasinophytes with the genus *Pyramimonas* as the most suspect source. Following studies on other euglenid plastid genomes (Bennett & Triemer, 2015; Bennett, Wiegert, & Triemer, 2014; Dabbagh & Preisfeld, 2016; Hrdá, Fousek, Szabová, Hampl, & Vlček, 2012; Pombert, James, Janouškovec, & Keeling, 2012; Wiegert, Bennett, & Triemer, 2012, 2013) confirmed *Pyramimonas parkeae* as the closest living relative of euglenophyte plastids.

The relative position of the plastid acquisition in the euglenid lineage was disputed in the past (Bodył, Mackiewicz, & Milanowski, 2010; Hannaert et al., 2003) and even though it is currently considered resolved and placed at the root of the extant euglenophytes the evolutionary history of euglenids seems more complex and traces of an interesting story about lateral gains and secondary losses of genes—and possibly endosymbionts or even organelles—can be read from the available sequence data of euglenids and their not-so-distant relatives kinetoplastids.

Kinetoplastids, a group known mainly for their parasitic members of considerable epidemiological significance of the genera *Trypanosoma* and *Leishmania*, have been shown to contain genes related to cyanobacteria or green plastids. Moreover, some of these genes are homologues of enzymes of the Calvin cycle which have been supposedly recruited for glycolysis-related functions in the glycosome (Hannaert et al., 2003) (i.e. peroxisome-derived organelle invented specifically by kinetoplastids and diplomonids; Morales et al., 2016; Rybicka, 1996). In reaction to this fact plus other occurrences of cyanobacterial-like genes in heterotrophic protists, a plastid-early hypothesis was formulated. This hypothesis placed the primary plastid acquisition much earlier in the evolution and suggested that all bikonts except the Archaeplastida underwent its secondary loss (Andersson & Roger, 2002; Maruyama, Matsuzaki, Misawa, & Nozaki, 2009). The less controversial theories focused on Euglenozoa suggested that the current secondary plastid of euglenophytes could have originated before the split of kinetoplastids and other euglenozoans and become lost in all lineages but euglenophytes (Bodył et al., 2010; Hannaert et al., 2003) resulting in the presence of “green” genes dispersed throughout the whole group. The possibility of a cryptic, euglenophyte-unrelated plastid gain and loss within the kinetoplastid lineage, was also considered (Bodył et al., 2010; Leander, 2004; Martin & Borst, 2003). The alternative hypothesis to all the previously mentioned suggests a mere lateral gene transfer from cyanobacteria or primary or secondary

plastid-containing eukaryote based solely on the “you are what you eat” notion (Doolittle, 1998) without a requirement for a previously existing stable relationship between the host and the endosymbiont/organelle (Bodyl et al., 2010; Leander, 2004; Maruyama et al., 2009).

The mosaicism in the evolutionary origins of euglenophyte genes provokes similar questions and similar approaches to their resolving. Euglenophytes contain undisputable amount of laterally transferred genes originating from rhodophytes or secondary algae containing red-algal plastids (e.g. haptophytes or stramenopiles), i.e. genes gained from other sources than the ancestor of the current plastid related to *Pyramimonas*. Most of these genes play a role in the typical plastid metabolic pathways and processes. The examples of these “nongreen” genes include several enzymes of the Calvin cycle (fructose biphosphatase, glyceraldehyde 3-phosphate dehydrogenase, phosphoribulokinase, ribose 5-phosphate isomerase, and triosephosphate isomerase), several enzymes of the tetrapyrrole biosynthesis (glutamate 1-semialdehyde 2,1-aminotransferase, uroporphyrinogen decarboxylase, coproporphyrinogen oxidase, and protoporphyrinogen oxidase), glycolytic enzyme glucokinase, and tocopherol biosynthesis enzyme homogentisate phytyltransferase (Lakey & Triemer, 2016; Markunas & Triemer, 2016; Maruyama, Suzuki, Weber, Archibald, & Nozaki, 2011). Whether similar plastid-related genes, green or nongreen, are present in heterotrophic euglenids, and to what extent, is currently unknown.

Where do these genes come from? The least controversial explanation is simple lateral gene transfer from eukaryotic prey, congruent with the “you are what you eat” hypothesis (Doolittle, 1998). Many eukaryovorous euglenids prey on algae, often in a rather generalist way. This nutritional strategy represents a clear prerequisite for the plastid acquisition as well as a logical source of the lateral gene transfer from various phototrophs. But how come that the transferred genes were retained by the euglenid predator despite having photosynthetic or other plastidal function and being virtually useless to a heterotrophic organism? The answer could be that the said euglenid predator was not a pure heterotroph but rather a mixotroph which already possessed the plastid or *Pyramimonas*-related endosymbiont not yet fully transformed into a true organelle, much like the contemporary species *R. viridis* (Yamaguchi et al., 2012). Genes obtained from the algal prey could have been immediately recruited for a function in a preexisting plastid-localized process, allowing the reductive evolution of the plastid/endosymbiont genome and its further integration into the host. This evolutionary

process was recently suggested as a major driving force of the organellogenesis of the chromatophore of *Paulinella chromatophora* (Nowack et al., 2016) (a rhizarian amoeba with a plastid-like organelle, chromatophore, acquired via second independent primary endosymbiosis; Marin, Nowack, & Melkonian, 2005).

Alternative hypotheses count with the possibility of euglenophyte predecessors harbouring other endosymbiont(s) or transient or even stable plastid(s) in their evolutionary past which were lost and replaced with the plastid from *Pyramimonas*-related alga. This model is based on the “shopping bag” hypothesis and reflects the current state of knowledge regarding the transient relationships, repeated gains and losses, and overall plastid fluidity observed in dinoflagellates (Bodył et al., 2010; Howe, Barbrook, Nisbet, Lockhart, & Larkum, 2008; Larkum, Lockhart, & Howe, 2007; O’Neill, Trick, Henrissat, & Field, 2015). If this hypothesis was correct, at least one “red” endosymbiont or preplastid was present in the ancestor of euglenophytes at some point. It is even possible that some of these plastid-like symbioses or similar ecological relationships could have taken place long time before the acquisition of the extant green plastid, explaining the presence of genes of algal origin in heterotrophic euglenids (Bodył et al., 2010; Leander, 2004; Maruyama et al., 2011) and making some of the implications of the plastid-early hypothesis true. The plastid-related genes it had left in the nuclear genome of the euglenid could have facilitated the acquisition of a brand new plastid and its “enslavement” by reductive genome evolution, quite opposite to the *Paulinella*-like model proposed earlier. Another imaginable scenario is the coexistence of red and green endosymbionts/preplastids in a single cell for a certain period of time during which the host had time to gain genes from both before ultimately “deciding” to let the red one go and keep the green one.

These hypotheses are not mutually exclusive and the truth can well lie somewhere between these proposed models—or somewhere else entirely.

---



### 3. PLASTID MORPHOLOGY: DISPLAY OF DIVERSITY

Not much can be said about the morphology of euglenophyte plastids in general because their shapes, sizes, numbers, positions, and other characteristics vary greatly among different genera and even species or strains (Ciugulea & Triemer, 2010; Leedale, 1967).

The most basic universal characteristic is perhaps that the euglenophyte plastids are enveloped by three membranes which reflect its secondary endosymbiotic origin and evolutionary history: the inner two membranes are generally believed to be homologous with the two membranes of primary plastid (i.e. originating in the two envelope membranes of cyanobacteria), the additional one is of an eukaryotic ancestry, derived either from the euglenid endomembrane system or from the cytoplasmic membrane of the green-algal ancestor of the plastid (Gibbs, 1978; Lefort-Tran, 1981). The same number of plastid membranes is described in dinoflagellates with peridinin-containing plastid, while four membranes are generally more common occurrence in secondary plastids. The outermost membrane of euglenophyte plastid is not spatially continuous with the ER as in the case of some other organisms with secondary plastids such as cryptophytes, haptophytes, and heterokonts (Bolte et al., 2009; Maier, Zauner, & Hempel, 2015). However, substantial communication via vesicles was observed to take place between the endomembrane system and plastid, and it was proposed that its outermost membrane might in fact act as a part of the secretory system (Sulli & Schwartzbach, 1995, 1996). This would be crucial for the targeting of ER-synthesized proteins and other molecules into the plastid.

Euglenophyte plastids are very variable in terms of shape: they can be spherical, oval, partially flattened, disc-shaped, watch-glass-shaped, biconcave or biconvex, with various numbers of lobes and prominences, or even ribbon- or star-shaped (Leedale, 1967). A certain plastid shape is usually specific to a certain monophyletic or paraphyletic group of euglenophytes. The amount of plastids per cell ranges from several to several dozens and it is also usually species-specific but in this case a certain range is conserved rather than a concrete number: in the model species *E. gracilis*, for example, the number of plastids varies from 6 to 12 (Leedale, 1967). This variability might be linked to the fact that plastid replication can take place either synchronously with the cell division or independently of it.

Euglenophyte plastids have well-developed thylakoids (also termed *discs* in older literature) which are organized in elongated stacks (also termed *lamellae* or *bands*) of varying number of layers (usually three) instead of grana (i.e. relatively high cylindrical stacks of thylakoids present in plant plastids) (Ben-Shaul, Schiff, & Epstein, 1966; Gibbs, 1970; Gibor & Granick, 1962).

Many euglenophyte plastids contain visible pyrenoids (i.e. dense regions where most of the carbon fixation enzymes are localized): these may be

either naked or covered by a watch glass-shaped paramylon cap, either from one side (haplopyrenoids) or from both sides (diplopyrenoids). Under transmission electron microscopy, the pyrenoids appear as conspicuously delimited regions of dense granulation (Gibbs, 1970). Pyrenoids are absent in plastids of the family Phacaceae and also in some members of the Euglenaceae (Gibbs, 1970; Leedale, 1967). The presence/absence and appearance of pyrenoids can also change during the life of a single cell as a result of changes in environmental factors—light intensity, length of the light/dark cycle, starvation, nutritional value of the medium, and growth phase of the culture.

Euglenophytes possess a single eyespot (*stigma*), light-sensing, granular, red/orange-coloured, and carotenoid-containing organelle, which enables positive or negative phototaxis. In contrast to green algae, where eyespots exist within the plastids, the euglenophyte eyespot is located in the cytoplasm near the base of the flagellum (Benedetti & Checcucci, 1975; Iseki et al., 2002; Osafune & Schiff, 1980b; Walne & Arnott, 1967). However, it was probably derived from plastid as well (Walne & Arnott, 1967).



#### 4. PLASTID GENOMES: SPED-UP EVOLUTION AND INTRONS GONE HAYWIRE

To this date, 17 euglenophyte plastid genomes (cpDNAs) have been published. The sampling covers most of the diversity of the Euglenaceae family with 14 cpDNAs (6 from the genus *Euglena* and 8 from other genera), one cpDNA of the Phacaceae family and two cpDNAs from the two genera of the Eutreptiales (Bennet, Wiegert, & Triemer, 2012; Bennett & Triemer, 2015; Bennett et al., 2014; Dabbagh & Preisfeld, 2016; Gockel & Hachtel, 2000; Hallick et al., 1993; Hrdá et al., 2012; Kasiborski, Bennett, Linton, & Lane, 2016; Pombert et al., 2012; Wiegert et al., 2012, 2013). Euglenophyte cpDNAs generally take the form of a circular chromosome. Four genomes are not complete and have not been circularized due to unknown number of repetitive sequences (Kasiborski et al., 2016) or unknown number of the ribosomal operons (Wiegert et al., 2012, 2013). Their basic characteristics including the comparison with the cpDNA of *P. parkeae* (Turmel et al., 2009) are summarized in Table 1.

The cpDNA of *P. parkeae*, the closest known relative of the euglenophyte plastid, is 101,605 bp long and contains 110 genes (Turmel et al., 2009). These values were seemingly reduced during the secondary

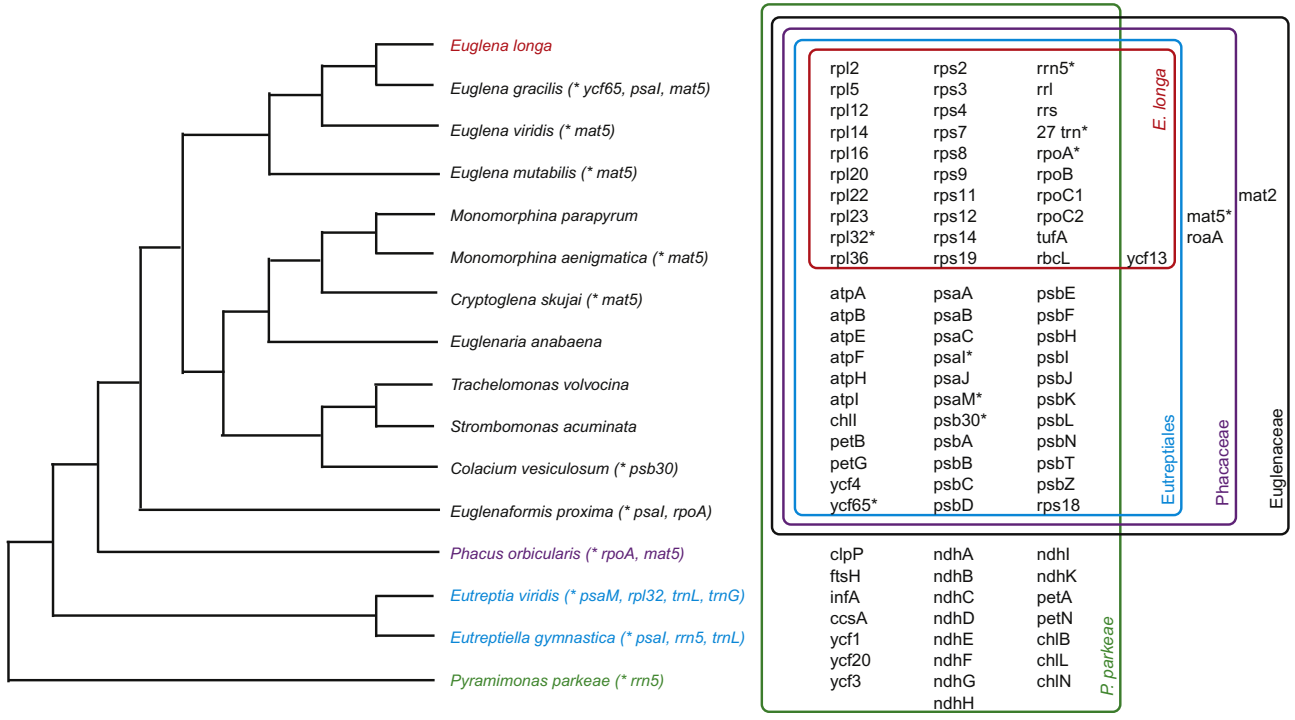


**Table 1** Characteristics of the Plastid Genomes of 17 Euglenophytes and *Pyramimonas parkeae* According to Sequences Deposited in GenBank

Taxonomy	Species/Strain	cp Genome Size (bp)	Number of Genes	Number of Introns	GC Content (%)
Euglenales	<i>Euglena gracilis</i> Z	143,171	90	145	26.1
	<i>E. gracilis</i> var. <i>bacillaris</i>	132,034	90	134	25.8
	<i>Euglena longa</i>	73,345	57	60	22.4
	<i>Euglena viridis</i> epitype	91,616	92	77	26.4
	<i>E. viridis</i> SAG 1224-17d	76,156	92	77	26.2
	<i>Euglena mutabilis</i>	86,975	92	76	26.7
	<i>Monomorphina aenigmatica</i>	74,746	92	53	29.4
	<i>Monomorphina parapyrum</i>	80,147	93	80	28.0
	<i>Cryptoglena skujai</i>	106,843	92	84	26.3
	<i>Euglenaria anabaena</i>	88,487	93	82	28.0
	<i>Trachelomonas volvocina</i>	85,392*	93	94	27.3
	<i>Strombomonas acuminata</i> *	144,166	93	110	26.6
	<i>Colacium vesiculosum</i> *	128,892	92	130	26.1
<i>Euglenaformis proxima</i>	94,185	91	113	26.9	
Phacales	<i>Phacus orbicularis</i> *	65,992	90	66	27.2
	<i>Eutreptia viridis</i> *	65,523	86	27	28.6
Eutreptiales	<i>Eutreptiella gymnastica</i>	67,623	87	8	34.3
	<i>Pyramimonas parkeae</i>	101,605	110	1	34.7

The “number of genes” indicates both protein and RNA coding but duplicates and ORFs were not included. The genome sizes marked by *asterisk* denote incomplete genomes. The first three columns marked taxonomy symbolize phylogenetic relationships between the organisms—merged rows represent defined clades; the Euglenales are coloured in *orange* and additionally divided into the Euglenaceae (*light orange*) and the Phacaceae (*light red*), the Eutreptiales are coloured in *yellow* and the prasinophytes are coloured in *purple*.

plastid establishment—22 genes (e.g. all genes of NADH-plastoquinone oxidoreductase of plastidal respiratory chain) were lost or transferred into the nucleus of the common ancestor of euglenophytes (Fig. 2). The highly conserved core content of euglenophyte cpDNAs consists of 89 genes (including *rrn5* not present in *P. parkeae*). These include 32 genes for photosynthetic proteins, 5 genes for transcription/translation proteins, 22 genes for ribosomal proteins, 3 rRNAs, and 27 tRNAs. One to four of these genes were not found in some lineages but their absence does not show any



**Fig. 2** Gene contents of the plastid genomes in 14 euglenophytes and *Pyramimonas parkeae*. Asterisks denote the genes missing in certain plastid genomes. Schematic phylogenetic tree is drawn according to Bennett and Triemer (2015).



phylogenetic pattern. Four genes were gained after the secondary endosymbiosis: the Eutreptiales have obtained *ycf13* (synonymous to the intron-encoded maturase *mat1*), the Phacaceae possess three additional genes, *ycf13*, *mat5* (found within the *psbA* gene of *Lepocinclis spirogyroides*—GenBank record), and *roaA* (ribosomal operon-associated gene), and the Euglenaceae possess *mat2* in addition to the three aforementioned genes (Fig. 2; Bennett & Triemer, 2015). The *mat5* gene was lost in several independent instances within the Euglenales (Bennett & Triemer, 2015; Kasiborski et al., 2016). Unsurprisingly, genes for photosynthetic proteins have been lost in *Euglena longa* whose plastid has no photosynthetic activity. The gene contents are summarized in Fig. 2.

The plastid genome organization differs between *P. parkeae*, which contains two inverted repeats with ribosomal operon and large and small single copy region (Turmel et al., 2009), and most euglenophytes, which do not have this quadripartite arrangement. The cpDNA of *Eutreptiella gymnastica* represents a single exception as it contains two inverted repeats (one of them discontinuous), each with rRNA operon (Hrdá et al., 2012). The rRNA operon in *E. gracilis* and *E. longa* is organized in three tandemly repeated copies (Gockel & Hachtel, 2000; Hallick et al., 1993). The replication origins of euglenophyte cpDNAs are presumably located in the VNTR (variable number of tandem repeats) region (Koller & Delius, 1982; Ravel-Chapuis, Heizmann, & Nigon, 1982). Overall, the genes are arranged into 15 conserved gene clusters whose order and orientation have been excessively rearranged and it is hard to trace and reconstruct the course of these rearrangements (Dabbagh & Preisfeld, 2016).

The most noticeable trend in the euglenophyte plastid genome evolution is undoubtedly the rapid intron gain. The numbers of recognized introns in the genomes change over time as the annotations improve and so the numbers provided in Table 1 reflect the status of genome annotations published in the GenBank database as of July 2017. Remarkably, the cpDNA of *Pyramimonas* contains only a single intron like those of other prasinophyte algae such as *Ostreococcus tauri* and *Pycnococcus provasoli* (Turmel et al., 2009). The number of introns started to grow after the plastid acquisition by the ancestral euglenid before the split of Eutreptiales and Euglenales: cpDNAs of *Eutreptiella* and *Eutreptia* which represent the deepest splits from other euglenophytes already contain 8 and 27 introns, respectively (Hrdá et al., 2012; Wiegert et al., 2012). An intron of *psbC* gene, which carries an intron-encoded maturase *ycf13*, is considered the ancestral intron because it is the only homologous intron in euglenophyte cpDNAs (Bennett &

Triemer, 2015; Pombert et al., 2012). The major intron gain and amplification took place within the Euglenales lineage resulting in several species containing over 100 introns in their cpDNAs, the most extreme case being the model *E. gracilis* strain Z with at least 145 introns (Bennett & Triemer, 2015; Bennett et al., 2014; Hallick et al., 1993; Wiegert et al., 2013). The intron accumulation seems to be an ongoing process which is afoot at this very moment and it could be potentially very interesting to review some of the “old” sequences and see if they changed during the long-term cultivation by humans during a tiny part of the period of their sped-up evolution.

Not only are the euglenophyte plastid introns numerous, they are also unique in their structure. Majority of them are classified as group II introns, which are known from prokaryotes and mitochondria and plastids of eukaryotes. They function as self-splicing ribozymes and contain six stem-loop-forming domains and a conserved 5'-border motif (GUGYG). Group II introns are mobile elements and their mobility is mediated by maturases (Bonen & Vogel, 2001; Sheveleva & Hallick, 2004). The euglenophyte plastid group II introns are often significantly shorter (the average length of euglenophyte group II intron is 463 nt—approximately 100 nt shorter than the average group II intron of liverwort; Dabbagh & Preisfeld, 2016) with some of the conserved domains being missing and/or divergent beyond recognition. In addition to group II introns, euglenophyte plastid genomes contain group III introns which are exclusive to these organisms. They seem to be extremely derived form of group II introns which is much shorter (of average length around 100 nt) and lacks almost all core structures retaining only one modified conserved domain on 3'-end and a degenerate 5'-border motif (the consensus is NUNNG) (Bonen & Vogel, 2001; Doetsch, 2000; Doetsch, Favreau, Kuscuglu, Thompson, & Hallick, 2001; Jenkins, Hong, & Hallick, 1995; Thompson, Copertino, Thompson, Favreau, & Hallick, 1995). To make matters even more complicated, euglenophyte plastid genomes also contain a number of twintrons—introns nested inside other introns (termed internal and external introns, respectively) which are spliced subsequently. Twintrons come in different types: group II, group III, and mixed or even complex ones where multiple internal introns (Copertino, Christopher, & Hallick, 1991; Copertino & Hallick, 1991, 1993) or even additional introns inside an internal intron (Drager & Hallick, 1993) can be observed in one external intron. In several cases, the recent conversion of a simple intron into a twintron is traceable on certain insertion sites. For example, six loci containing a twintron in the *E. gracilis* cpDNA contain only a single intron in *Monomorpha aenigmatica*.

In most of these cases, the introns of *M. aenigmatica* are orthologous to the external introns of their twintron counterparts in *E. gracilis* (Pombert et al., 2012). This further supports the notion that intron and twintron propagation in euglenophyte cpDNAs is a recent and probably still active process.

The relatively low GC content of the cpDNAs of euglenophytes is also linked to the intron accumulation since euglenid plastid introns are generally AT-rich and tend to bias the overall GC content of the whole genome. This is especially apparent in the case of *E. longa* whose cpDNA has the lowest number of genes while still being rich in introns, as a result its GC content is extremely low (22.4%) (Gockel & Hachtel, 2000).



## 5. PLASTID BIOGENESIS AND HOUSEKEEPING: HOW TO MAKE IT AND HOW TO CONTROL IT

The crucial part of organellogenesis is the ability of the host cell to take control over the replication and biogenesis of the formerly independent endosymbiotic hostage. The development of plastids of euglenophytes was studied in detail in the model species *E. gracilis* and the following subchapter will be based mainly on the findings regarding this particular organism.

*Euglena* plastids develop from proplastids which are nascent or dark-induced regressed stages with low or no photosynthetic activity, small number of thylakoids, and low amount of photosynthetic pigments stored mostly in their precursor forms (protochlorophyll and protochlorophyllide).

The plastid development is triggered by light and influenced by other environmental stimuli, especially availability of organic carbon sources which act as catabolic repressors that modulate the metabolic mode of the cell towards the more efficient one in the given situation. Glucose, ethanol, and acetate were shown to inhibit chlorophyll synthesis, while malate and succinate do not seem to function as catabolic repressors. Ethanol was also shown to inhibit synthesis or activity of various other proteins of carbon fixation pathway, light-harvesting antennae, and plastid housekeeping system (Horrum & Schwartzbach, 1980; Monroy & Schwartzbach, 1984; Schwartzbach, 2017). As a result, the cell performs the energetically expensive switch to the autotrophic nutritional mode and develops mature plastids only when carbon sources consumable by the relatively energetically cheap glyoxylate pathway become unavailable.

The DNA replication in plastids can take place independently of the nuclear DNA replication. However, the process of plastid greening and

development is directed by the nucleus on the transcriptional and translational level. The proteosynthetic apparatus of the developing plastids receives an external boost as its nucleus-encoded components (ribosomal proteins and initiation and elongation factors) are upregulated by light, synthesized in the cytoplasm, and transported into proplastids as well as other nucleus-encoded plastid proteins (Bingham & Schiff, 1979; Bingham & Schiff, 1979; Egan & Carell, 1972; Fox et al., 1980; Kraus & Spremulli, 1988). The amounts of rRNA, tRNA, and other nucleic acids which are fairly low in inactive proplastids increase up to threefold after illumination (Egan & Carell, 1972). The plastid genes are expressed constitutively regardless of light or dark conditions and the light increases the transcriptional activity but does not change the composition and ratio of the expressed genes (Geimer et al., 2009).

Protochlorophyll and protochlorophyllide are phototransformable molecules which are believed to act as photoreceptors on the thylakoid membranes of the developing plastids (perhaps in cooperation with a yet unspecified photoreceptor specialized to blue light sensing). These precursors are converted to chlorophyll and play a role in the initiation of de novo chlorophyll synthesis (Egan, Dorsky, & Schiff, 1975; Kirk, 1970; Stern, Epstein, & Schiff, 1964; Stern, Schiff, & Epstein, 1964).

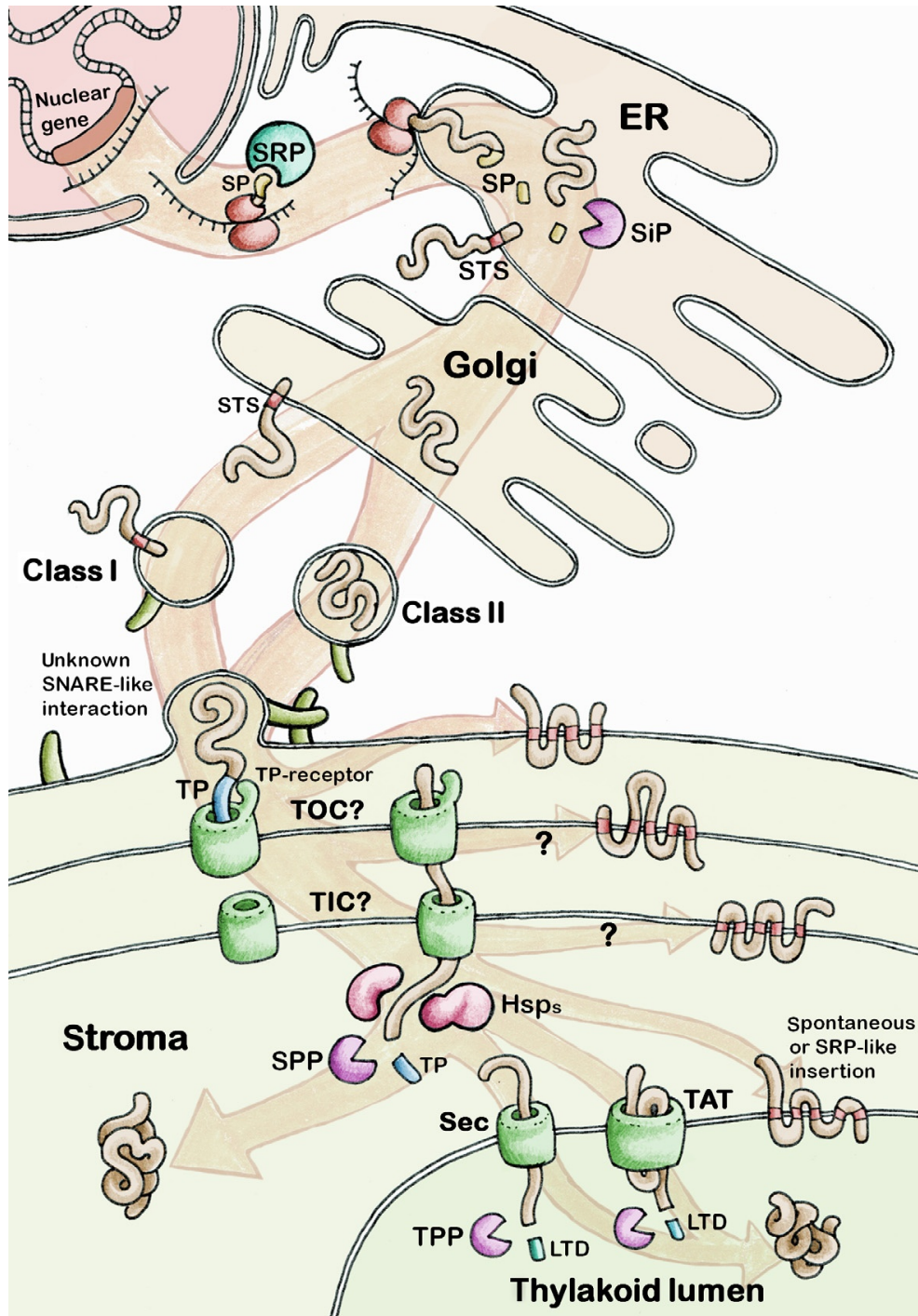
The proplastids start to grow in size and new thylakoids are formed by the invagination of the innermost membrane and eventually start fusing and stacking into lamellae (Ben-Shaul et al., 1966). Thylakoid formation and chlorophyll synthesis are mutually dependent so the time dynamics of these two processes are concurrent. This development starts soon after the induction by light but it progresses very slowly during the initial lag phase and sets off rapidly after approximately 6 h. The length of the lag can be, however, significantly modified by the adjustment of the light conditions, namely preillumination. While the concentration of chlorophyll grows steadily, the concentration of carotenoids remains more or less the same—the final ratio between these two types of pigments is ca. 2.5:1 in the mature plastids (Stern, Schiff, et al., 1964). The interconnected carbon fixation and oxygen production start several hours after the plastid differentiation induction and grow rapidly after approximately 10 h. Around this time, the plastids swell quickly gaining roughly three times the original mass and thylakoid amount. It was noticed that the number of initial proplastids in *E. gracilis* is usually around 30 per cell while the final number of mature plastids is around 10. The logical deduction is that the proplastids actually form aggregates of three and fuse around this time off the differentiation causing a significant leap in

the plastids size and metabolic capacity. The process of the plastid maturation is completed in about 72 h (Ben-Shaul et al., 1966).

A vast majority of proteins which are synthesized and inserted into thylakoids during this process are coded in the nucleus and translated on cytoplasmic ribosomes—the gene expression is almost completely under the control of the host cell (Bingham & Schiff, 1979), only a small number of proteins are synthesized in the plastid (Schwartzbach & Schiff, 1974). These gene products are imported into plastids based on their rather complex-targeting signals.

The plastid protein import mechanism is much more complicated in euglenophytes in comparison to plants and is far from clearly resolved to this date. A schematic of the protein transport into euglenophyte plastid is shown in Fig. 3. Plastid-targeted precursor proteins (preproteins) coded in nucleus generally have an N-terminal signal peptide which is very similar to the classical signal peptide of proteins destined to the secretory pathway. The signal peptide is supposedly recognized by the signal recognition particle right after its translation and protrusion from the cytoplasmic ribosome and the rest of the translation takes place on the rough ER and the nascent preprotein is cotranslationally imported into the ER lumen. Subsequently, the signal peptide is cleaved by a signal peptidase in the ER revealing the following part of the signal whose amino acid composition is similar to that of a canonical plant chloroplast-targeting signal, the so-called transit peptide: it is rich in serine, threonine, and alanine; depleted in aspartic and glutamic acid; and has a slight positive charge as a result. The length of the transit peptides is quite variable and ranges from 36 to 135 amino acid residues in *E. gracilis* (Durnford & Gray, 2006). The origin of these transit peptides could be related to *cis*- and *trans*-splicing of short introns which were identified in some of these sequences (Vesteg et al., 2010). In euglenophytes, nucleus-encoded plastid preproteins can be divided into two classes based on the presence or absence of the third, highly hydrophobic domain which follows immediately after the transit peptide (Durnford & Gray, 2006) and acts as a stop-transfer transmembrane anchor. Preproteins are transported from the ER to the Golgi apparatus and they are either packed into a transport vesicle (class II) or anchored in its membrane (class I) with the transit peptide inside and the rest of the protein protruding into the cytoplasm (Durnford & Gray, 2006; Sulli, Fang, Muchhal, & Schwartzbach, 1999; Sulli & Schwartzbach, 1996). Vesicles from Golgi fuse with the outermost plastid membrane by a yet unknown mechanism. This process is resistant to *N*-ethylmaleimide that interacts with the *N*-ethylmaleimide-sensitive factor (NSF) and inhibits





**Fig. 3** Schematic of the plastid protein import pathway in euglenophytes. Nuclear-encoded plastid-targeted proteins are cotranslationally transported into the ER by the same mechanism as proteins destined into the secretory pathway based on their signal peptide (SP). They are either completely inserted into the ER lumen or remain anchored in its membrane by a hydrophobic stop-transfer signal (STS). Their SPs are cleaved by the signal peptidase (SiP) in the ER lumen. They pass through Golgi and are loaded onto vesicles which then fuse with the outermost plastid membrane via an unknown, SNARE-independent mechanism. Proteins destined to plastid stroma or thylakoids pass the remaining two membranes via transit peptide (TP)-dependent process which is probably facilitated by plant TIC- and TOC-like complexes. In plastid

soluble NSF attachment protein receptor (SNARE)-dependent membrane fusion. This suggests that SNAREs are not utilized in the vesicular transport of plastid preproteins (Sláviková *et al.*, 2005). After reaching the inter-membrane space between the outermost and the middle plastid membrane, the transit peptide is apparently recognized by a yet unknown receptor and the preprotein is pulled through the inner two membranes, most probably by a system homologous to the TOC/TIC (translocons at the outer/inner envelope membrane of chloroplasts) machinery which is present in primary plastids of plants (Shi & Theg, 2013) and generally conserved (albeit reduced in the number of subunits) among other secondary plastids (Maier *et al.*, 2015; Sheiner & Striepen, 2013). The transit peptides of euglenid plastid proteins were repeatedly demonstrated to function in heterologous system and successfully direct the protein into plant plastid both *in vitro* and *in vivo* (Inagaki, Fujita, Hase, & Yamamoto, 2000; Shashidhara, Limsli, Shackletonii, Robinsonii, & Smith, 1992; Sláviková *et al.*, 2005). These results suggest the presence of a TOC/TIC-like machinery in euglenophyte plastids. However, no direct proof of existence of these protein complexes in euglenophytes was brought to this date.

The transit peptide is presumably cleaved by stromal processing peptidase and the protein assumes its native conformation with the help of putative translocase-associated and stromal chaperones (Maier *et al.*, 2015; Shi & Theg, 2013). Proteins destined to the thylakoid membrane or lumen are supposedly further sorted based on their additional signals (N-terminal, C-terminal, or internal) and inserted into thylakoids by specialized import pathways known from primary plastids such as the Sec-dependent pathway or the twin-arginine translocation pathway (Jarvis & Robinson, 2004).

The transcription of genes encoded in the plastid genome was observed to take place constitutively with practically all genes including several unknown ORFs and pseudogenes being expressed and present as mRNAs in the plastid, albeit in very low concentration. The transcripts sometimes form very large and relatively stable polycistronic units which are subsequently subjected to extensive posttranscriptional modifications. The plastid

---

stroma, the TPs are cleaved by the stromal processing peptidase (SPP) and the proteins are sorted into their final localization (stroma, thylakoid membrane, or thylakoid lumen) and folded to their mature conformation. Proteins destined to thylakoid lumen are transported based on their lumen-targeting domains (LTD) which are then cleaved by the thylakoid-processing peptidase (TPP). Proteins of the plastid envelope are inserted into their target membrane via an unknown mechanism or spontaneously.

gene expression does not react to environmental stimuli in terms of quality but rather on quantitative level. Moreover, some genes seem to have very short turnover and to be expressed only to be degraded very quickly. In summary, regulation of the gene expression in euglenophyte plastid might be taking place predominantly on the level of posttranscriptional RNA processing and/or translation rather than transcription (Geimer et al., 2009; Thompson et al., 1995). A small portion of plastid transcripts was shown to have polyadenylated 3'-ends in *E. gracilis*. The low ratios of polyadenylated vs nonpolyadenylated RNAs (only 1/350 to 1/100,000) suggest that the polyadenylation serves as a signal for exonucleolytic degradation as it does in the case of primary plastids (Záhonová et al., 2014).



## 6. PLASTID METABOLISM: A FACTORY WITH REDUNDANT PRODUCTION LINES

In this chapter, findings on biosynthetic pathways in euglenophyte plastid will be briefly summarized with focus on their potentially unique characteristics and differences from other groups which are interesting from evolutionary point of view. Almost all studies cited in this section were performed on the model euglenophyte *E. gracilis*, and generalizations based on them should be taken with a grain of salt.

Photosynthetic apparatus of the *E. gracilis* plastid is comparable to that of green algae and plants in regard to the function and architecture. Photosynthesis is a subject of tight regulation by light (Yoshida et al., 2016) and it reacts negatively to both extremes in light conditions as the photosynthetic activity is inhibited or even completely compromised by both darkness and the excessive illumination by visible or UV light (Richter, Helbling, Streb, & Häder, 2007). Additionally, the photosynthetic capacity of *E. gracilis* is affected by salinity (Gonzalez-Moreno, Gomez-Barrera, Perales, & Moreno-Sanchez, 1997) and other ion concentrations (Ferroni, Baldisserotto, Fasulo, Pagnoni, & Pancaldi, 2004; Krajčovič et al., 2015).

*E. gracilis* synthesizes chlorophylls *a* and *b* via the standard plastid-localized pathway from aminolevulinic acid. There are several spectroscopically distinguishable precursor forms of chlorophylls (i.e. protochlorophyll, protochlorophyllide, and phytyl-protochlorophyllide), and developing plastids especially contain these precursors to a great extent, and their ratio, dynamics, and interconversions are believed to play a major role in regulation of biosynthesis of chlorophyll and possibly other compounds (Kirk, 1970).



For example, protochlorophyllide has been shown to inhibit further steps in chlorophyll biosynthesis until its light-induced conversion (Schwartzbach, Schiff, & Goldstein, 1975).

Aminolevulinic acid can be synthesized in two ways: through the C5 pathway from glutamate or through the C4 (Shemin) pathway from glycine and succinyl-CoA; most organisms utilize only one of these pathways. In *E. gracilis*, aminolevulinic acid was proven to be synthesized via the C5 pathway and for some time the organism was thought to lack the C4 pathway (Gomez-Silva, Timko, & Schiff, 1985). However, the situation was recently shown to be more complex as two pathways for the synthesis of aminolevulinic acid followed by the synthesis of haem have been predicted in *Euglena*. One pathway is predicted to be partially localized in either the cytosol or mitochondrion to produce haem for mitochondria. This pathway utilizes aminolevulinic acid synthesized in the C4 pathway. Another pathway is localized in the plastid to produce chlorophyll and it utilizes aminolevulinic acid synthesized in the C5 pathway (Kořený & Oborník, 2011).

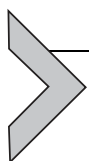
Carotenoid synthesis seems to be linked to chlorophyll production and it was even hypothesized that one could be synthesized from another due to the observed reciprocal proportion of these two types of pigments (Wolken & Mellon, 1956). The regulation of biosynthesis of carotenoids and chlorophyll indeed seems to be tuned and subjected to reciprocal feedback but the exact mechanism of it remains unclear. Major *E. gracilis* carotenoids are antheraxanthin (more than 80%),  $\beta$ -carotene (11%), and neoxanthin (7%), which represent 99% of total carotenoids. Minor carotenoids include cryptoxanthin,  $\gamma$ -carotene,  $\zeta$ -carotene, echinenone, hydroxyechinenone, and its derivate canthaxanthin (also termed euglenanone in older literature as it was believed to be unique to euglenids) (Krinsky & Goldsmith, 1960). The first metabolite in carotenoid synthesis pathway is geranylgeranyl diphosphate (GGPP) which is synthesized in putatively plastidal methylerythritol phosphate/deoxyxylulose phosphate (MEP/DOXP) pathway—as opposed to the GGPP synthesized in mitochondrial mevalonate pathway which is used for the synthesis of phytols and sterols in *E. gracilis*. Therefore, there are two distinct GGPP pools in *E. gracilis* and carotenoids and phytols are synthesized by independent pathways (Kim, Filtz, & Proteau, 2004).

A variety of terpenoids are synthesized in the *E. gracilis* plastid: plastoquinone,  $\alpha$ -tocopherol,  $\alpha$ -tocopherolquinone, phytylquinone, nonaprenyl, octaprenyl, nonaprenyl toluquinone, octaprenyl toluquinone, and phytyl

pyrophosphate (Griffiths, Threlfall, & Goodwin, 1967; Thomas & Threlfall, 1974). Their production is not directly linked to photosynthesis but it is light dependent; terpenoid formation is mostly inhibited in dark-grown cultures as opposed to ubiquinone which is synthesized in mitochondrion and whose production level does not correlate with different light conditions (Griffiths et al., 1967). Polyprenylation/phytylation or nonoxidative decarboxylation of homogentisate were proposed as a mechanism of this reaction (Thomas & Threlfall, 1974). Tocopherols (compounds collectively termed as vitamin E) are present in all *E. gracilis* cell fractions including plastids where 97% of their bulk is represented by  $\alpha$ -tocopherol which is thought to be synthesized in situ from other tocopherols which are transported from outside the plastid (Shigeoka, Onishi, Nakano, & Kitaoka, 1986). This compound functions as a protective agent against reactive oxygen species and its synthesis modulation by various substrates and conditions was studied in the view of potential biotechnological or pharmaceutical application (Fujita, Aoyagi, Ogbonna, & Tanaka, 2008; Fujita, Ogbonna, Tanaka, & Aoyagi, 2009). Ascorbic acid (vitamin C) is also an important antioxidative protectant. Remarkably, in *E. gracilis* it is synthesized by the alternative terminal enzyme, L-galactonolactone dehydrogenase, instead of L-gulonolactone oxidase which was lost multiple times and then replaced by an alternative enzyme in multiple phototrophic lineages (Wheeler, Ishikawa, Pornsaksit, & Smirnov, 2015).

*E. gracilis* also synthesizes galactolipids (monogalactosyl and digalactosyl diglycerides) and sulpholipids (mainly sulphoquinovosyl diglyceride) in its plastid (Davies, Mercer, & Goodwin, 1966; Rosenberg & Gouaux, 1967; Rosenberg, Gouaux, & Milch, 1966). Production of these lipids is linked to the plastid development and thylakoid membrane amplification and is modulated by light conditions—much higher amounts of these compounds can be found in photosynthetically active cells in comparison to dark-grown or bleached ones (Matson, Fei, & Chang, 1970). Interestingly and in contrast to plants, monogalactosyl and digalactosyl diglycerides are probably synthesized by two different enzymes in *E. gracilis* (Blee & Schantz, 1978). Additionally, the conversion from monogalactosyl to digalactosyl diglycerides is not possible, and the ratio of these galactolipids is significantly skewed towards digalactosyl diglycerides which are synthesized preferentially (Matson et al., 1970). These lipids are essential for the thylakoid assembly and growth, and some have been shown to be miscible with chlorophyll and phytol, major nonprotein components of the thylakoid membranes (Liljenberg & Selstam, 1980).

The major storage compound of euglenids is paramylon, a starch-like polysaccharide, which forms granules of various morphologies in the cytosol and can take up to 90% of the dry cell weight. The paramylon consists of an unbranched, water-insoluble  $\beta$ -1,3-glucan which is rather unusual since most eukaryotes generally synthesize  $\alpha$ -glucans (or branched  $\beta$ -1,4- or  $\beta$ -1,6-glucans in case of fungi and some grains) (Barras & Stone, 1968; Barsanti, Vismara, Passarelli, & Gualtieri, 2001; Dwyer & Smillie, 1971; Monfils, Triemer, & Bellairs, 2011; Šantek, Felski, Friehs, Lotz, & Flaschel, 2010). The paramylon biogenesis occurs in all euglenids including nonphotosynthetic species suggesting that it is not evolutionarily related to plastid acquisition. In heterotrophic euglenids, the paramylon is synthesized in mitochondrial prominences and mitochondrion-derived vesicles by the gluconeogenesis pathway (Calvayrac & Briand, 1978). In photosynthetic euglenophytes, the paramylon can be synthesized by both the original mitochondrial pathway and during carbon fixation in pyrenoids of plastids. The close association of plastid and mitochondrion and the eccentric flow of vesicles derived presumably from both organelles have been observed during the paramylon biogenesis (Calvayrac, Laval-Martin, Briand, & Farineau, 1981). This functional connection of the two organelles is remarkable and can be a potential source of unusual molecular phenomena regarding exchange of metabolites and possibly other compounds. The other major storage compounds of euglenids are wax esters which are not synthesized in plastid (Inui, Miyatake, Nakano, & Kitaoka, 1982; Koritala, 1989; Schneider & Betz, 1985) but whose production is linked to the production of paramylon, since interconversions between these two storage molecules occur regularly and their balance shifts in reaction to light and oxygen availability: in aerobic conditions, the cell preferentially accumulates paramylon; in anaerobic conditions, the wax ester accumulation is more significant (Barras & Stone, 1968; Inui et al., 1982). Both of these compounds are potentially usable in biotechnology, agriculture, nutrition, and even cancer prevention (Karaca et al., 2014; Krajčovič et al., 2015; Kuda, Enomoto, & Yano, 2009; Rodríguez-Zavala, Ortiz-Cruz, Mendoza-Hernández, & Moreno-Sánchez, 2010; Sugiyama et al., 2009; Watanabe et al., 2013).



## **7. SECONDARY OSMOTROPHY AND PLASTID BLEACHING: PLASTIDS THAT FORGOT HOW TO PLASTID**

Secondarily osmotrophic euglenids represent euglenophytes which have lost their photosynthetic pigments and the ability to perform

photosynthesis. There are at least five species of euglenophytes that have lost the photosynthesis independently (Marin, 2004; Marin, Palm, Klingberg, & Melkonian, 2003). It is a matter of a debate whether some of these euglenophytes such as *Euglena quartana* (previously *Khawkinea quartana*) or *Phacus ocellatus* (previously *Hyalophacus ocellatus*) have lost the plastid compartments completely or whether they still contain residual plastids with genomes (Marin, 2004). However, the best-studied secondary osmotroph is *E. longa*, a close relative of photosynthetic *E. gracilis* (Mullner, Angeler, Samuel, Linton, & Triemer, 2001), which contains a plastid with a genome (Gockel & Hachtel, 2000).

The circular 73 kb plastid genome of *E. longa* is about half the size of a plastid genome of *E. gracilis* and it carries 57 protein-coding genes (Gockel & Hachtel, 2000)—housekeeping genes responsible for transcription and translation (*rrn*, *rpl*, *rps*, *rpo*, and *tuf* genes), tRNA genes, and several ORFs and genes with unknown function (*orf* and *ycf* genes). Transcripts and proteins of various genes, including Rubisco large subunit gene (*rbcL*), have been found suggesting that the *E. longa* plastid possesses functional transcription and translation machinery (Gockel & Hachtel, 2000; Sheveleva & Hallick, 2004; Záhonová et al., 2016). It was recently shown that an intact plastid genome is essential for *E. longa* growth (Hadariová, Vesteg, Birčák, Schwartzbach, & Krajčovič, 2017) but it remains a mystery which gene(s) is/are indispensable for its survival (Gockel & Hachtel, 2000; Hadariová et al., 2017). The photosynthesis-related genes (photosystems I and II, cytochrome *b6f* complex, ATP-synthase) have disappeared from the *E. longa* plastid genome with the exception of *rbcL* (Gockel & Hachtel, 2000). Both *rbcL* and nucleus-encoded *RbcS* genes are translated but their abundance in cells is very low. Protein sequences of *E. longa rbcL* and *RbcS* are highly divergent compared to their homologues in the photosynthetic relatives, suggesting that the Rubisco enzyme of *E. longa* probably has an unusual function, if any (Záhonová et al., 2016). Some other nonphotosynthetic organisms retain *rbcL* genes in their plastid genomes and it is hypothesized that these *RbcL* proteins may either act as oxygenases, be involved in glycine and serine biosynthesis, be required for an alternative lipid biosynthesis pathway (Schwender, Goffman, Ohlrogge, & Shachar-Hill, 2004), or perform another unidentified function (Sanchez-Puerta, Lippmeier, Apt, & Delwiche, 2007; Wolfe & dePamphilis, 1998; Záhonová et al., 2016).

*E. longa* lacks the eyespot and the paraflagellar swelling, but apart from that, the cell is indistinguishable from bleached mutants of *E. gracilis* under light microscopy, and thus it was formerly viewed as a naturally bleached

form of *E. gracilis* (Bodył, 1996). Bleaching of *E. gracilis* is an irreversible process inducible by antibiotics, UV light, high pressure, heat, mutagens, or carcinogens (Krajčovič, Ebringer, & Schwartzbach, 2001). In *E. gracilis*, various morphological and ultrastructural changes can be observed during bleaching. During the switching from autotrophy to heterotrophy, the plastids and plastid DNA are degraded and the colour of the culture changes from green to white (or slightly orange or pink depending on the presence of carotenoids). The bleaching level depends on several factors—pH, content of phosphate in the medium, or the age of the culture. It is possible to reach 100% plastid bleaching in some cases (Cook, Harris, & Nachtwey, 1974; Krajčovič, Ebringer, & Polónyi, 1989). Spontaneous bleaching in *E. gracilis* was observed as early as in 1912 by Ternetz and streptomycin was the first defined agent for controlled bleaching of all cells in a population. In 1961, Ebringer widened a list of bleaching agents by adding erythromycin and some other macrolide antibiotics (Krajčovič et al., 2001).

There are various types of antibacterial drugs with different modes of action. Some of them act as inhibitors of bacterial DNA synthesis (e.g. mitomycin or anthramycin) while the ones with aminohexose molecular structure function as inhibitors of bacterial as well as plastidal protein synthesis (e.g. kanamycin, pactamycin, or neomycin). Their irreversible effect on *E. gracilis* was observed in the past (Ebringer, 1972). Perhaps the most effective eliminators of *E. gracilis* plastids are quinolone antibiotics (inhibitors of bacterial DNA gyrase), especially their new derivatives—fluoroquinolones (Krajčovič et al., 1989). Various ultrastructural and plastid DNA changes occur when different bleaching agents are used (Krajčovič et al., 2001).

For example, the quinolones, but also nitrosoguanidine and furylfuramide, reduce the number of thylakoids and destroy plastids (Polónyi, Ebringer, Krajcovic, & Kapeller, 1990). The plastid DNA degradation has been observed in all bleached *E. gracilis* cultures. Conkling, Thomas, and Ortiz (1993) noticed a gradual loss of cpDNA and the study of Krajčovič et al. (1999) has demonstrated a structural rearrangement of cpDNA in *E. gracilis* cells bleached by elevated temperature. Recently, it was shown that the plastid gene *rpl16* invariably persists in *E. gracilis* strain Z bleached in the long-term by ofloxacin (DNA gyrase inhibitor) or streptomycin (bacterial protein synthesis inhibitor) (Hadariová et al., 2017), which suggested that this gene was retained presumably due to its position on the cpDNA being close to the replication origin (Hadariová et al., 2017). A different gene, 16S rRNA, has been retained in *E. gracilis* FACHB47 strain



treated with the same antibiotics (Wang, Shi, & Xu, 2004). It has been reported that plastid transcripts are reduced after treating by various bleaching agents—UV light, temperature, and antibiotics (Geimer et al., 2009).

In the past decades, several stable white mutants derived from *E. gracilis* have been characterized (Polónyi, Ebringer, Dobias, & Krajčovič, 1998; Schiff, Lyman, & Russell, 1971).  $W_3BUL$  mutant was induced with UV irradiation of *E. gracilis* var. *bacillaris* strain (Schiff et al., 1971). It contains plastid remnants (Heizmann, Salvador, & Nigon, 1976; Osafune & Schiff, 1980a; Osafune, Schiff, & Hase, 1987; Parthier & Neumann, 1977) and a specific type of sulpholipid in its thylakoid membranes (Saidha & Schiff, 1989). The treatment of *E. gracilis* var. *bacillaris* with streptomycin produced another mutant strain,  $W_{10}BSmL$ . In contrast to  $W_3BUL$ ,  $W_{10}BSmL$  contains neither the eyespot (Osafune & Schiff, 1980b), carotenoids (Fong & Schiff, 1979), and sulpholipids (Saidha & Schiff, 1989), nor the plastid residues (Osafune & Schiff, 1983). The  $W_{gm}ZOflL$  mutant was derived from *E. gracilis* strain *Z* by treatment with an ofloxacin derivative (Polónyi et al., 1998). The common features of  $W_{gm}ZOflL$  and  $W_3BUL$  mutants include the presence of the eyespot and carotenoids. The giant mitochondria were described in  $W_{gm}ZOflL$  in the past (Polónyi et al., 1998) but current microscopic observations suggest that  $W_{gm}ZOflL$  mutant does not possess mitochondria of such form anymore (personal experiences). The level of the plastid genome degradation, i.e., plastid genes presence/absence of the nonphotosynthetic mutants of *E. gracilis*, is still unknown but they probably lack most of the, if not all, plastid genes (unpublished data).

The bleaching of *E. gracilis* induced by fluoroquinolone ofloxacin has been shown to be useful in the investigation of potential antimutagens such as flavonoids (Križková, Nagy, Polónyi, & Ebringer, 1998) and antioxidants as ascorbic acid, sodium selenite, and many others (Ebringer et al., 1996; Kogan et al., 2004; Križková, Ďuračková, Šandula, Sasinková, & Krajčovič, 2001; Križková, Mučaji, Nagy, & Krajčovič, 2004; Križková et al., 2006). Ciprofloxacin, besides the effective bleaching of *E. gracilis*, also causes a suppression of the rudimentary plastid (apicoplast) genome replication in a parasite *Toxoplasma gondii* (Roos & Fichera, 1997). Various other quinolones have been demonstrated as useful inhibitors of a malaria-causing apicomplexan *Plasmodium falciparum* (Mahmoudi et al., 2003). Thus, *E. gracilis* plastid bleaching may also have practical implications. The research of the plastid genome degradation in bleached *E. gracilis* may represent a suitable tool for the development of new drugs harmless to humans and

successful in fighting apicomplexan parasites that cause millions of fatal disease cases in the third-world countries annually (Krajčovič et al., 2001).



## 8. CONCLUSIONS

Euglenophytes are a group of phototrophic protists which have been discovered as early as at the end of 17th century. During the last century, their biochemistry and ultrastructure were inspected to a great detail. However, there are still many missing pieces in the puzzle of their molecular biology, genetics, and phylogenetics as some of the methods crucial for these disciplines became widely available only recently and others are still quite limited because of the missing genomic data and insufficient means for genetic transformation. These hurdles are, however, very much worth overcoming because photosynthetic euglenids are potentially very useful models for both applied and basic research. From the applied point of view, they represent easily and environment-friendly cultivated microalgae capable of synthesizing various compounds usable in the production of biofuels and other lipid-based technological materials as well as nutritional and pharmaceutical products. From the basic point of view, they are one of the key groups to understand secondary endosymbioses, establishment and reduction of organelles and evolutionary processes taking part therein. Findings concerning euglenid molecular biology, genetics, and genomics can help to illuminate the evolution of excavates and, by extension, of eukaryotes as whole.

## ACKNOWLEDGEMENTS

The authors wish to acknowledge professor Bożena Zakryś who is the author of the drawings used in Fig. 1 and who kindly approved their usage in this publication.

Support for the authors' salaries and stipend came from the project of the Ministry of Education, Youth and Sports of CR within the National Sustainability Program II (Project BIOCEV-FAR) LQ1604 and by the project "BIOCEV" (CZ.1.05/1.1.00/02.0109).

## REFERENCES

- Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., et al. (2012). The revised classification of eukaryotes. *The Journal of Eukaryotic Microbiology*, 59(5), 429–493. <http://dx.doi.org/10.1111/j.1550-7408.2012.00644.x>.
- Andersson, J. O., & Roger, A. J. (2002). A cyanobacterial gene in nonphotosynthetic protists—An early chloroplast acquisition in eukaryotes? *Current Biology: CB*, 12(2), 115–119. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11818061>.
- Barras, D. R., & Stone, B. A. (1968). Carbohydrate composition and metabolism in *Euglena*. In D. E. Buetow (Ed.), *The biology of Euglena: Vol. II. Biochemistry* (149–191). New York: Academic Press.

- Barsanti, L., Vismara, R., Passarelli, V., & Gualtieri, P. (2001). Paramylon ( $\beta$ -1,3-glucan) content in wild type and WZSL mutant of *Euglena gracilis*. Effects of growth conditions. *Journal of Applied Phycology*, 13(1), 59–65. <http://dx.doi.org/10.1023/A:1008105416065>.
- Benedetti, A. P., & Checcucci, A. (1975). Paraflagellar body (PFB) pigments studied by fluorescence microscopy in *Euglena gracilis*. *Plant Science Letters*, 4, 47–51.
- Bennet, M., Wiegert, K., & Triemer, R. (2012). Comparative chloroplast genomics between *Euglena viridis* and *Euglena gracilis* (Euglenophyta). *Phycologia*, 51(6), 711–718. <http://dx.doi.org/10.2216/12-017.1>.
- Bennett, M. S., & Triemer, R. E. (2015). Chloroplast genome evolution in the Euglenaceae. *Journal of Eukaryotic Microbiology*, 62(6), 773–785. <http://dx.doi.org/10.1111/jeu.12235>.
- Bennett, M. S., Wiegert, K. E., & Triemer, R. E. (2014). Characterization of *Euglenaformis* gen. nov. and the chloroplast genome of *Euglenaformis* [*Euglena*] *proxima* (Euglenophyta). *Phycologia*, 53(1), 66–73. <http://dx.doi.org/10.2216/13-198.1>.
- Ben-Shaul, Y., Schiff, J. A., & Epstein, H. T. (1966). Studies of chloroplast development in euglena: VII. Fine structure of the developing plastid 1. *Biophysical Journal*, 6(4), 373–383. <http://dx.doi.org/10.1038/185825a0>.
- Bingham, S., & Schiff, J. A. (1979). Events surrounding the early development of *Euglena* chloroplasts. 15. Origin of plastid thylakoid polypeptides in wild-type and mutant cells. *Biochimica et Biophysica Acta (BBA)—Bioenergetics*, 547(3), 512–530.
- Bingham, S., & Schiff, J. A. (1979). Events surrounding the early development of *Euglena* chloroplasts. 16. Plastid thylakoid polypeptides during greening. *Biochimica et Biophysica Acta (BBA)—Bioenergetics*, 547(3), 531–543.
- Blee, E., & Schantz, R. (1978). Biosynthesis of galactolipids in *Euglena gracilis*: I, incorporation of UDP galactose into galactosyldiglycerides. *Plant Science Letters*, 13, 247–255.
- Bodył, A. (1996). Is the origin of *Astasia longa* an example of the inheritance of acquired characteristics? *Acta Protozoologica*, 35, 87–94, Nencki Institute of Experimental Biology.
- Bodył, A., Mackiewicz, P., & Milanowski, R. (2010). Did trypanosomatid parasites contain a eukaryotic alga-derived plastid in their evolutionary past? *Journal of Parasitology*, 96(2), 465–475. <http://dx.doi.org/10.1645/Ge-1810.1>.
- Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S., & Maier, U.-G. (2009). Protein targeting into secondary plastids. *The Journal of Eukaryotic Microbiology*, 56(1), 9–15. <http://dx.doi.org/10.1111/j.1550-7408.2008.00370.x>.
- Bonen, L., & Vogel, J. (2001). The ins and outs of group II introns. *Trends in Genetics*, 17(6), 322–331. [http://dx.doi.org/10.1016/S0168-9525\(01\)02324-1](http://dx.doi.org/10.1016/S0168-9525(01)02324-1).
- Calvayrac, R., & Briand, J. (1978). Paramylon synthesis and the chondriome of *Euglena gracilis* Z. In G. Ducet & C. Lance (Eds.), *Plant mitochondria* (pp. 435–443). Amsterdam: Elsevier/North Holland.
- Calvayrac, R., Laval-Martin, D., Briand, J., & Farineau, J. (1981). Paramylon synthesis by *Euglena gracilis* photoheterotrophically grown under low O<sub>2</sub> pressure. *Planta*, 153(1), 6–13. <http://dx.doi.org/10.1007/BF00385311>.
- Cavalier-Smith, T. (1981). Eukaryote kingdoms: Seven or nine? *Biosystems*, 14(3–4), 461–481. [http://dx.doi.org/10.1016/0303-2647\(81\)90050-2](http://dx.doi.org/10.1016/0303-2647(81)90050-2).
- Cavalier-Smith, T. (2016). Higher classification and phylogeny of Euglenozoa. *European Journal of Protistology*, 56, 250–276. <http://dx.doi.org/10.1016/j.ejop.2016.09.003>.
- Ciugulea, I., & Triemer, R. E. (2010). *A color atlas of photosynthetic euglenoids*. East Lansing, MI: Michigan State University Press. Retrieved from <http://library.wur.nl/WebQuery/clc/1953618>.
- Conkling, B. A., Thomas, E. J., & Ortiz, W. (1993). Delayed but complete loss of chloroplast DNA in heat-bleaching cultures of *Euglena gracilis*. *Journal of Plant Physiology*, 142(3), 307–311. [http://dx.doi.org/10.1016/S0176-1617\(11\)80427-X](http://dx.doi.org/10.1016/S0176-1617(11)80427-X).
- Cook, J. R., Harris, P., & Nachtwey, D. S. (1974). Irreversible plastid loss in *Euglena gracilis* under physiological conditions. *Plant Physiology*, 53(2), 284–290. <http://dx.doi.org/10.1104/PP.53.2.284>.



- Copertino, D. W., Christopher, D. A., & Hallick, R. B. (1991). A mixed group-II/group-III twintron in the *Euglena gracilis* chloroplast ribosomal protein S3-gene—Evidence for intron insertion during gene evolution. *Nucleic Acids Research*, *19*(23), 6491–6497.
- Copertino, D. W., & Hallick, R. B. (1991). Group II twintron: An intron within an intron in a chloroplast cytochrome b-559 gene. *The EMBO Journal*, *10*(2), 433–442. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452664&tool=pmcentrez&rendertype=abstract>.
- Copertino, D. W., & Hallick, R. B. (1993). Group II and group III introns of twintrons: Potential relationships with nuclear pre-mRNA introns. *Trends in Biochemical Sciences*, *18*(12), 467–471. [http://dx.doi.org/10.1016/0968-0004\(93\)90008-B](http://dx.doi.org/10.1016/0968-0004(93)90008-B).
- Dabbagh, N., & Preisfeld, A. (2016). The chloroplast genome of *Euglena mutabilis*—cluster arrangement, intron analysis, and intrageneric trends. *Journal of Eukaryotic Microbiology*, *1993*, 31–44. <http://dx.doi.org/10.1111/jeu.12334>.
- Davies, W. H., Mercer, E. I., & Goodwin, T. W. (1966). Some observations on the biosynthesis of the plant sulpholipid by *Euglena gracilis*. *The Biochemical Journal*, *98*(2), 369–373. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1264853&tool=pmcentrez&rendertype=abstract>.
- de M. Bicudo, C. E., & Menezes, M. (2016). Phylogeny and classification of Euglenophyceae: A brief review. *Frontiers in Ecology and Evolution*, *4*(March), 1–15. <http://dx.doi.org/10.3389/fevo.2016.00017>.
- Doetsch, N. A. (2000). *Group III intron structure and evolutionary analysis in euglenoid chloroplast genomes (doctoral dissertation)*. The University of Arizona. <http://hdl.handle.net/10150/284139>.
- Doetsch, N. A., Favreau, M. R., Kuscuglu, N., Thompson, M. D., & Hallick, R. B. (2001). Chloroplast transformation in *Euglena gracilis*: Splicing of a group III twintron transcribed from a transgenic psbK operon. *Current Genetics*, *39*(1), 49–60. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11318107>.
- Doolittle, W. F. (1998). You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics: TIG*, *14*(8), 307–311. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9724962>.
- Douglas, S. E., & Turner, S. (1991). Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. *Journal of Molecular Evolution*, *33*(3), 267–273. <http://dx.doi.org/10.1007/BF02100678>.
- Drager, R. G., & Hallick, R. B. (1993). A complex twintron is excised as four individual introns. *Nucleic Acids Research*, *21*(10), 2389–2394. <http://dx.doi.org/10.1093/nar/21.10.2389>.
- Durnford, D. G., & Gray, M. W. (2006). Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryotic Cell*, *5*(12), 2079–2091. <http://dx.doi.org/10.1128/EC.00222-06>.
- Dwyer, M. R., & Smillie, R. M. (1971).  $\beta$ -1,3 glucan: A source of carbon and energy for chloroplast development in *Euglena Gracilis*. *Australian Journal of Biological Sciences*, *24*(1), 15–22.
- Ebringer, L. (1972). Are plastids derived from prokaryotic micro-organisms? Action of antibiotics on chloroplasts of *Euglena gracilis*. *Journal of General Microbiology*, *71*(1), 35–52. <http://dx.doi.org/10.1099/00221287-71-1-35>.
- Ebringer, L., Dobias, J., Krajčovič, J., Polónyi, J., Križková, L., & Lahitová, N. (1996). Antimutagens reduce ofloxacin-induced bleaching in *Euglena gracilis*. *Mutation Research/Environmental Mutagenesis and Related Subjects*, *359*(2), 85–93. [http://dx.doi.org/10.1016/S0165-1161\(96\)90255-1](http://dx.doi.org/10.1016/S0165-1161(96)90255-1).
- Egan, J. M., & Carell, E. F. (1972). Studies on chloroplast development and replication in *euglena*: III. A study of the site of synthesis of alkaline deoxyribonuclease induced during chloroplast development in *Euglena gracilis*. *Plant Physiology*, *50*(3), 391–395.
- Egan, J. M., Dorsky, D., & Schiff, J. A. (1975). Events surrounding the early development of *euglena* chloroplasts: VI. Action spectra for the formation of chlorophyll, lag elimination

- in chlorophyll synthesis, and appearance of TPN-dependent triose phosphate dehydrogenase and alkaline DNase activities. *Plant Physiology*, 56(2), 318–323.
- Ehrenberg, C. G. (1830). *Organisation, Systematik und geographisches Verhältniss der Infusionsthierchen*. Berlin: Druckerei der Königlichen Akademie der Wissenschaften. Retrieved from <http://www.biodiversitylibrary.org/item/18341>.
- Ferroni, L., Baldisserotto, C., Fasulo, M. P., Pagnoni, A., & Pancaldi, S. (2004). Adaptive modifications of the photosynthetic apparatus in *Euglena gracilis* Klebs exposed to manganese excess. *Protoplasma*, 224(3–4), 167–177. <http://dx.doi.org/10.1007/s00709-004-0072-4>.
- Fong, F., & Schiff, J. A. (1979). Blue-light-induced absorbance changes associated with carotenoids in *Euglena*. *Planta*, 146(2), 119–127. <http://dx.doi.org/10.1007/BF00388221>.
- Fox, L., Erion, J., Tarnowski, J., Spemulli, L., Brot, N., & Weissbach, H. (1980). Communication *Euglena gracilis* Chloroplast EF-Ts. *The Journal of Biological Chemistry*, 255(13), 6018–6019. Retrieved from <http://www.jbc.org/content/255/13/6018.full.pdf>.
- Fujita, T., Aoyagi, H., Ogbonna, J. C., & Tanaka, H. (2008). Effect of mixed organic substrate on  $\alpha$ -tocopherol production by *Euglena gracilis* in photoheterotrophic culture. *Applied Microbiology and Biotechnology*, 79(3), 371–378. <http://dx.doi.org/10.1007/s00253-008-1443-0>.
- Fujita, T., Ogbonna, J. C., Tanaka, H., & Aoyagi, H. (2009). Effects of reactive oxygen species on  $\alpha$ -tocopherol production in mitochondria and chloroplasts of *Euglena gracilis*. *Journal of Applied Phycology*, 21(2), 185–191. <http://dx.doi.org/10.1007/s10811-008-9349-x>.
- Geimer, S., Belicová, A., Legen, J., Sláviková, S., Herrmann, R. G., & Krajcovic, J. (2009). Transcriptome analysis of the *Euglena gracilis* plastid chromosome. *Current Genetics*, 55(4), 425–438. <http://dx.doi.org/10.1007/s00294-009-0256-8>.
- Gibbs, S. P. (1970). The comparative ultrastructure of the algal chloroplast. *Annals of the New York Academy of Sciences*, 175(1), 454–473. <http://dx.doi.org/10.1111/j.1749-6632.1970.tb45167.x>.
- Gibbs, S. P. (1978). The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Canadian Journal of Botany*, 56(22), 2883–2889. <http://dx.doi.org/10.1139/b78-345>.
- Gibbs, S. P. (1981). The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Annals of the New York Academy of Sciences*, 361, 193–208. [1 Origins and E] <http://dx.doi.org/10.1111/j.1749-6632.1981.tb54365.x>.
- Gibor, A., & Granick, S. (1962). The plastid system of normal and bleached *Euglena gracilis*. *The Journal of Protozoology*, 9(3), 327–334.
- Gockel, G., & Hachtel, W. (2000). Complete gene map of the plastid genome of the non-photosynthetic Euglenoid flagellate *Astasia longa*. *Protist*, 151(4), 347–351. [http://dx.doi.org/10.1078/S1434-4610\(04\)70033-4](http://dx.doi.org/10.1078/S1434-4610(04)70033-4).
- Gomez-Silva, B., Timko, M. P., & Schiff, J. A. (1985). Chlorophyll biosynthesis from glutamate or 5-aminolevulinate in intact *Euglena* chloroplasts. *Planta*, 165(1), 12–22. <http://dx.doi.org/10.1007/BF00392206>.
- Gonzalez-Moreno, S., Gomez-Barrera, J., Perales, H., & Moreno-Sanchez, R. (1997). Multiple effects of salinity on photosynthesis of the protist *Euglena gracilis*. *Physiologia Plantarum*, 101(4), 777–786. <http://dx.doi.org/10.1111/j.1399-3054.1997.tb01063.x>.
- Griffiths, W. T., Threlfall, D. R., & Goodwin, T. W. (1967). Nature, intracellular distribution and formation of terpenoid quinones in maize and barley shoots. *The Biochemical Journal*, 103(2), 589–600. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1270445&tool=pmcentrez&rendertype=abstract>.
- Guiry, M. D., & Guiry, G. M. (2017). *AlgaeBase*. National University of Ireland, Galway: World-wide Electronic Publication.

- Hadariová, L., Vesteg, M., Birčák, E., Schwartzbach, S. D., & Krajčovič, J. (2017). An intact plastid genome is essential for the survival of colorless *Euglena longa* but not *Euglena gracilis*. *Current Genetics*, *63*(2), 331–334. <http://dx.doi.org/10.1007/s00294-016-0641-z>.
- Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., et al. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research*, *21*(15), 3537–3544. <http://dx.doi.org/10.1093/nar/21.15.3537>.
- Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J.-P., Rigden, D. J., Michels, P. A. M., et al. (2003). Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 1067–1071. <http://dx.doi.org/10.1073/pnas.0335769100>.
- Harris, J. (1695). Some microscopical observations of vast numbers of animalcula seen in water by John Harris, M. A. Rector of Winchelsea in Sussex, and F.R.S. *Philosophical Transactions of the Royal Society of London*, *19*(215–235), 254–259. <http://dx.doi.org/10.1098/rstl.1695.0036>.
- Heizmann, P., Salvador, G. F., & Nigon, V. (1976). Occurrence of plastidial rRNAs and plastidial structures in bleached mutants of *Euglena gracilis*. *Experimental Cell Research*, *99*(2), 253–260. [http://dx.doi.org/10.1016/0014-4827\(76\)90581-4](http://dx.doi.org/10.1016/0014-4827(76)90581-4).
- Horum, M. A., & Schwartzbach, S. D. (1980). Nutritional regulation of organelle biogenesis in euglena: Repression of chlorophyll and NADP-glyceraldehyde-3-phosphate dehydrogenase synthesis. *Plant Physiology*, *65*(2), 382–386. <http://dx.doi.org/10.1104/PP.65.2.382>.
- Howe, C. J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J., & Larkum, A. W. D. (2008). The origin of plastids. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1504), 2675–2685. <http://dx.doi.org/10.1098/rstb.2008.0050>.
- Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012). The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PloS One*, *7*(3), e33746. <http://dx.doi.org/10.1371/journal.pone.0033746>.
- Inagaki, J., Fujita, Y., Hase, T., & Yamamoto, Y. (2000). Protein translocation within chloroplast is similar in *Euglena* and higher plants. *Biochemical and Biophysical Research Communications*, *277*(2), 436–442. <http://dx.doi.org/10.1006/bbrc.2000.3702>.
- Inui, H., Miyatake, K., Nakano, Y., & Kitaoka, S. (1982). Wax ester fermentation in *Euglena gracilis*. *FEBS Letters*, *150*(1), 89–93. [http://dx.doi.org/10.1016/0014-5793\(82\)81310-0](http://dx.doi.org/10.1016/0014-5793(82)81310-0).
- Iseki, M., Matsunaga, S., Murakami, A., Ohno, K., Shiga, K., Yoshida, K., et al. (2002). A blue-light-activated adenyl cyclase mediates photoavoidance in *Euglena gracilis*. *Nature*, *415*(6875), 1047–1051.
- Jarvis, P., & Robinson, C. (2004). Mechanisms of protein import and routing in chloroplasts. *Current Biology: CB*, *14*(24), R1064–R1077. <http://dx.doi.org/10.1016/j.cub.2004.11.049>.
- Jenkins, K. P., Hong, L., & Hallick, R. B. (1995). Alternative splicing of the *Euglena gracilis* chloroplast *roaA* transcript. *RNA (New York, N.Y.)*, *1*(6), 624–633.
- Karaca, H., Bozkurt, O., Ozaslan, E., Baldane, S., Berk, V., Inanc, M., et al. (2014). Positive effects of oral  $\beta$ -glucan on mucositis and leukopenia in colorectal cancer patients receiving adjuvant FOLFOX-4 combination chemotherapy. *Asian Pacific Journal of Cancer Prevention*, *15*, 3641–3644. <http://dx.doi.org/10.7314/APJCP.2014.15.8.3641>.
- Karnkowska, A., Bennett, M. S., Watzka, D., Kim, J. I., Zakryś, B., & Triemer, R. E. (2015). Phylogenetic relationships and morphological character evolution of photosynthetic Euglenids (Excavata) inferred from taxon-rich analyses of five genes. *Journal of Eukaryotic Microbiology*, *62*(3), 362–373. <http://dx.doi.org/10.1111/jeu.12192>.

- Kasiborski, B. A., Bennett, M. S., Linton, E. W., & Lane, C. (2016). The chloroplast genome of *Phacus orbicularis* (Euglenophyceae): An initial datum point for the phacaceae. *Journal of Phycology*, 52(3), 404–411. <http://dx.doi.org/10.1111/jpy.12403>.
- Kim, D., Filtz, M. R., & Proteau, P. J. (2004). The methylerythritol phosphate pathway contributes to carotenoid but not phytol biosynthesis in *Euglena gracilis*. *Journal of Natural Products*, 67(6), 1067–1069. <http://dx.doi.org/10.1021/np049892x>.
- Kim, J. I., Linton, E. W., & Shin, W. (2015). Taxon-rich multigene phylogeny of the photosynthetic euglenoids (Euglenophyceae). *Frontiers in Ecology and Evolution*, 3(August), 1–11. <http://dx.doi.org/10.3389/fevo.2015.00098>.
- Kim, J. I., & Shin, W. (2008). Phylogeny of the euglenales inferred from plastid LSU rDNA sequences. *Journal of Phycology*, 44(4), 994–1000. <http://dx.doi.org/10.1111/j.1529-8817.2008.00536.x>.
- Kirk, J. T. O. (1970). Biochemical aspects of chloroplast development. *Annual Review of Plant Physiology*, 21(1), 11–42. <http://dx.doi.org/10.1146/annurev.pp.21.060170.000303>.
- Kogan, G., Skorik, Y., Zitnanová, I., Krizkova, L., Durackova, Z., Gomes, C., et al. (2004). Antioxidant and antimutagenic activity of N-(2-carboxyethyl)chitosan. *Toxicology and Applied Pharmacology*, 201(3), 303–310. <http://dx.doi.org/10.1016/j.taap.2004.05.009>.
- Koller, B., & Delius, H. (1982). Origin of replication in chloroplast DNA of *Euglena gracilis* located close to the region of variable size. *The EMBO Journal*, 1(8), 995–998.
- Kořený, L., & Oborník, M. (2011). Sequence evidence for the presence of two tetrapyrrole pathways in *Euglena gracilis*. *Genome Biology and Evolution*, 3(1), 359–364. <http://dx.doi.org/10.1093/gbe/evr029>.
- Koritala, S. (1989). Microbiological synthesis of wax esters by *Euglena gracilis*. *Journal of the American Oil Chemists Society*, 66(1), 133–134. <http://dx.doi.org/10.1007/BF02661801>.
- Krajčovič, J., Ebringer, L., & Polónyi, J. (1989). Quinolones and coumarins eliminate chloroplasts from *Euglena gracilis*. *Antimicrobial Agents and Chemotherapy*, 33(11), 1883–1889. <http://dx.doi.org/10.1128/AAC.33.11.1883>.
- Krajčovič, J., Ebringer, L., & Schwartzbach, S. D. (2001). Reversion of endosymbiosis? In *Symbiosis* (pp. 185–206). Dordrecht: Kluwer Academic Publishers. [http://dx.doi.org/10.1007/0-306-48173-1\\_11](http://dx.doi.org/10.1007/0-306-48173-1_11).
- Krajčovič, J., Vacula, R., Steiner, J. M., Löffelhardt, W., Belicová, A., Sláviková, S., et al. (1999). Molecular effects of some stress factors on the chloroplast genetic apparatus of the flagellate *Euglena gracilis*. In *The chloroplast: From molecular biology to biotechnology* (pp. 121–128). Dordrecht: Springer Netherlands. [http://dx.doi.org/10.1007/978-94-011-4788-0\\_19](http://dx.doi.org/10.1007/978-94-011-4788-0_19).
- Krajčovič, J., Vesteg, M., & Schwartzbach, S. D. (2015). Euglenoid flagellates: A multifaceted biotechnology platform. *Journal of Biotechnology*, 202, 135–145. <http://dx.doi.org/10.1016/j.jbiotec.2014.11.035>.
- Kraus, B. L., & Spremulli, L. L. (1988). Evidence for the nuclear location of the genes for chloroplast IF-2 and IF-3 in *Euglena*. *Plant Physiology*, 88(4), 993–995. <http://dx.doi.org/10.1104/PP.88.4.993>.
- Krinsky, N. I., & Goldsmith, T. H. (1960). The carotenoids of the flagellated alga, *Euglena gracilis*. *Archives of Biochemistry and Biophysics*, 91(12), 271–279. [http://dx.doi.org/10.1016/0003-9861\(60\)90501-4](http://dx.doi.org/10.1016/0003-9861(60)90501-4).
- Križková, L., Ďuračková, Z., Šandula, J., Sasinková, V., & Krajčovič, J. (2001). Antioxidative and antimutagenic activity of yeast cell wall mannans in vitro. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 497(1–2), 213–222. [http://dx.doi.org/10.1016/S1383-5718\(01\)00257-1](http://dx.doi.org/10.1016/S1383-5718(01)00257-1).
- Križková, L., Mučaji, P., Nagy, M., & Krajčovič, J. (2004). Triterpenoid cynarasaponins from *Cynara cardunculus* L. reduce chemically induced mutagenesis in vitro. *Phytomedicine*, 11(7–8), 673–678. <http://dx.doi.org/10.1016/j.phymed.2003.09.001>.
- Križková, L., Nagy, M., Polónyi, J., & Ebringer, L. (1998). The effect of flavonoids on ofloxacin-induced mutagenicity in *Euglena gracilis*. *Mutation Research/Genetic Toxicology*



- and *Environmental Mutagenesis*, 416(1–2), 85–92. [http://dx.doi.org/10.1016/S1383-5718\(98\)00080-1](http://dx.doi.org/10.1016/S1383-5718(98)00080-1).
- Križková, L., Žitňanová, I., Mislovičová, D., Masárová, J., Sasinková, V., Ďuračková, Z., et al. (2006). Antioxidant and antimutagenic activity of mannan neoglycoconjugates: Mannan–human serum albumin and mannan–penicillin G acylase. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 606(1–2), 72–79. <http://dx.doi.org/10.1016/j.mrgentox.2006.03.003>.
- Kuda, T., Enomoto, T., & Yano, T. (2009). Effects of two storage  $\beta$ -1,3-glucans, laminaran from *Eicenia bicyclis* and paramylon from *Euglena gracili*, on cecal environment and plasma lipid levels in rats. *Journal of Functional Foods*, 1, 399–404. <http://dx.doi.org/10.1016/j.jff.2009.08.003>.
- Lakey, B., & Triemer, R. (2016). The tetrapyrrole synthesis pathway as a model of horizontal gene transfer in euglenoids. *Journal of Phycology*, 53, 198–217. <http://dx.doi.org/10.1111/jpy.12491>.
- Larkum, A. W. D., Lockhart, P. J., & Howe, C. J. (2007). Shopping for plastids. *Trends in Plant Science*, 12(5), 189–195. <http://dx.doi.org/10.1016/j.tplants.2007.03.011>.
- Lax, G., & Simpson, A. G. B. (2013). Combining molecular data with classical morphology for uncultured phagotrophic euglenids (Excavata): A single-cell approach. *Journal of Eukaryotic Microbiology*, 60(6), 615–625. <http://dx.doi.org/10.1111/jeu.12068>.
- Leander, B. S. (2004). Did trypanosomatid parasites have photosynthetic ancestors? *Trends in Microbiology*, 12(6), 251–258. <http://dx.doi.org/10.1016/j.tim.2004.04.001>.
- Leander, B. S., Esson, H. J., & Breglia, S. A. (2007). Macroevolution of complex cytoskeletal systems in euglenids. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 29(10), 987–1000. <http://dx.doi.org/10.1002/bies.20645>.
- Leander, B. S., Triemer, R. E., & Farmer, M. a. (2001). Character evolution in heterotrophic euglenids. *European Journal of Protistology*, 37, 337–356.
- Leedale, G. F. (1967). *Euglenoid flagellates*. Englewood Cliffs, NJ: Prentice-Hall.
- Lefort-Tran, M. (1981). The triple layered organization of the *Euglena* chloroplast envelope (signification and functions). *Plant Biology*, 94(1), 463–476.
- Liljenberg, C., & Selstam, E. (1980). Interactions of chlorophyll a and terpenoid alcohols with chloroplast acyl lipids in monomolecular films. *Physiologia Plantarum*, 48(1966), 428–434.
- Linton, E. W., Karnkowska-Ishikawa, A., Kim, J. I., Shin, W., Bennett, M. S., Kwiatowski, J., et al. (2010). Reconstructing Euglenoid evolutionary relationships using three genes: Nuclear SSU and LSU, and chloroplast SSU rDNA sequences and the description of *Euglenaria* gen.nov. (Euglenophyta). *Protist*, 161(4), 603–619. <http://dx.doi.org/10.1016/j.protis.2010.02.002>.
- Mahmoudi, N., Ciceron, L., Franetich, J.-F., Farhati, K., Silvie, O., Eling, W., et al. (2003). In vitro activities of 25 quinolones and fluoroquinolones against liver and blood stage *Plasmodium* spp. *Antimicrobial Agents and Chemotherapy*, 47(8), 2636–2639. <http://dx.doi.org/10.1128/AAC.47.8.2636-2639.2003>.
- Maier, U. G., Zauner, S., & Hempel, F. (2015). Protein import into complex plastids: Cellular organization of higher complexity. *European Journal of Cell Biology*, 94(7–9), 340–348. <http://dx.doi.org/10.1016/j.ejcb.2015.05.008>.
- Marin, B. (2004). Origin and fate of chloroplasts in the Euglenoida. *Protist*, 155(March), 13–14. <http://dx.doi.org/10.1078/1434461000159>.
- Marin, B., Nowack, E. C. M., & Melkonian, M. (2005). A plastid in the making: Evidence for a second primary endosymbiosis. *Protist*, 156(4), 425–432. <http://dx.doi.org/10.1016/j.protis.2005.09.001>.
- Marin, B., Palm, A., Klingberg, M., & Melkonian, M. (2003). Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist*, 154(1), 99–145. <http://dx.doi.org/10.1078/143446103764928521>.

- Markunas, C. M., & Triemer, R. E. (2016). Evolutionary history of the enzymes involved in the Calvin-Benson cycle in Euglenids. *Journal of Eukaryotic Microbiology*, 63(3), 326–339. <http://dx.doi.org/10.1111/jeu.12282>.
- Martin, W., & Borst, P. (2003). Secondary loss of chloroplasts in trypanosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 765–767. <http://dx.doi.org/10.1073/pnas.0437776100>.
- Maruyama, S., Matsuzaki, M., Misawa, K., & Nozaki, H. (2009). Cyanobacterial contribution to the genomes of the plastid-lacking protists. *BMC Evolutionary Biology*, 9(1), 197. <http://dx.doi.org/10.1093/bioinformatics/17.8.754>.
- Maruyama, S., Suzaki, T., Weber, A. P. M., Archibald, J. M., & Nozaki, H. (2011). Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology*, 11(1), 105. <http://dx.doi.org/10.1186/1471-2148-11-105>.
- Matson, R. S., Fei, M., & Chang, S. B. (1970). Comparative studies of biosynthesis of galactolipids in *Euglena-gracilis* strain-Z. *Plant Physiology*, 45(4), 531.
- Monfils, A. K., Triemer, R. E., & Bellairs, E. F. (2011). Characterization of paramylon morphological diversity in photosynthetic euglenoids (Euglenales, Euglenophyta). *Phycologia*, 50(2), 156–169. <http://dx.doi.org/10.2216/09-112.1>.
- Monroy, A. F., & Schwartzbach, S. D. (1984). Catabolite repression of chloroplast development in *Euglena*. *Cell Biology*, 81, 2786–2790. Retrieved from <http://www.pnas.org/content/81/9/2786.full.pdf>.
- Morales, J., Hashimoto, M., Williams, T. A., Hirawake-Mogi, H., Makiuchi, T., Tsubouchi, A., et al. (2016). Differential remodelling of peroxisome function underpins the environmental and metabolic adaptability of diplomonads and kinetoplastids. *Proceedings of the Royal Society of London B: Biological Sciences*, 283(1830). Retrieved from <http://rspb.royalsocietypublishing.org/content/283/1830/20160520.long>.
- Morden, C. W., & Golden, S. S. (1991). Sequence analysis and phylogenetic reconstruction of the genes encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyllb-containing prokaryote *Prochlorothrix hollandica*. *Journal of Molecular Evolution*, 32(5), 379–395. <http://dx.doi.org/10.1007/BF02101278>.
- Mullner, A. N., Angeler, D. G., Samuel, R., Linton, E. W., & Triemer, R. E. (2001). Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence. *International Journal of Systematic and Evolutionary Microbiology*, 51(3), 783–791. <http://dx.doi.org/10.1099/00207713-51-3-783>.
- Nowack, E. C. M., Price, D. C., Bhattacharya, D., Singer, A., Melkonian, M., & Grossman, A. R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), 12214–12219. <http://dx.doi.org/10.1073/pnas.1608016113>.
- O'Neill, E. C., Trick, M., Henrissat, B., & Field, R. A. (2015). *Euglena* in time: Evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspectives in Science*, 6, 84–93. <http://dx.doi.org/10.1016/j.pisc.2015.07.002>.
- Ogawa, T., Tamoi, M., Kimura, A., Mine, A., Sakuyama, H., Yoshida, E., et al. (2015). Enhancement of photosynthetic capacity in *Euglena gracilis* by expression of cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase leads to increases in biomass and wax ester production. *Biotechnology for Biofuels*, 8(1), 80. <http://dx.doi.org/10.1186/s13068-015-0264-5>.
- Osafune, T., & Schiff, J. A. (1980a). Events surrounding the early development of *Euglena* chloroplasts. *Journal of Ultrastructure Research*, 73(1), 64–76. [http://dx.doi.org/10.1016/0022-5320\(80\)90116-1](http://dx.doi.org/10.1016/0022-5320(80)90116-1).

- Osafune, T., & Schiff, J. A. (1980b). Stigma and flagellar swelling in relation to light and carotenoids in *Euglena gracilis* var. *bacillaris*. *Journal of Ultrastructure Research*, 73(3), 336–349. [http://dx.doi.org/10.1016/S0022-5320\(80\)90093-3](http://dx.doi.org/10.1016/S0022-5320(80)90093-3).
- Osafune, T., & Schiff, J. A. (1983). W10BSmL, a mutant of *Euglena gracilis* var. *bacillaris* lacking plastids. *Experimental Cell Research*, 148(2), 530–535. [http://dx.doi.org/10.1016/0014-4827\(83\)90176-3](http://dx.doi.org/10.1016/0014-4827(83)90176-3).
- Osafune, T., Schiff, J. A., & Hase, E. (1987). Light-independent and dependent phases of proplastid development in *Euglena gracilis* W3BUL. *Cell Structure and Function*, 12(5), 453–461. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3119230>.
- Parthier, B., & Neumann, D. (1977). Structural and functional analysis of some plastid mutants of *Euglena gracilis*. *Biochemie und Physiologie der Pflanzen*, 171(6), 547–562. [http://dx.doi.org/10.1016/S0015-3796\(17\)30349-9](http://dx.doi.org/10.1016/S0015-3796(17)30349-9).
- Polónyi, J., Ebringer, L., Dobias, J., & Krajčovič, J. (1998). Giant mitochondria in chloroplast-deprived *Euglena gracilis* late after N-succinimidylfloxacin treatment. *Folia Microbiologica*, 43(6), 661–666. <http://dx.doi.org/10.1007/BF02816386>.
- Polónyi, J., Ebringer, L., Krajcovic, J., & Kapeller, K. (1990). Injured mitochondria in cells of *Euglena gracilis* after DNA gyrase inhibitors treatment. *Zeitschrift für Mikroskopisch-Anatomische Forschung*, 104(1), 61–78. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2161591>.
- Pombert, J. F., James, E. R., Janouškovec, J., & Keeling, P. J. (2012). Evidence for transitional stages in the evolution of Euglenid group II introns and twintrons in the *Monomorpha aenigmatica* plastid genome. *PloS One*, 7(12). <http://dx.doi.org/10.1371/journal.pone.0053433>.
- Ravel-Chapuis, P., Heizmann, P., & Nigon, V. (1982). Electron microscopic localization of the replication origin of *Euglena gracilis* chloroplast DNA. *Nature*, 300(5887), 78–81. <http://dx.doi.org/10.1038/300078a0>.
- Richter, P., Helbling, W., Streb, C., & Häder, D.-P. (2007). PAR and UV effects on vertical migration and photosynthesis in *Euglena gracilis*? *Photochemistry and Photobiology*, 83(4), 818–823. <http://dx.doi.org/10.1111/j.1751-1097.2007.00134.x>.
- Rodríguez-Zavala, J. S., Ortiz-Cruz, M. A., Mendoza-Hernández, G., & Moreno-Sánchez, R. (2010). Increased synthesis of  $\alpha$ -tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *Journal of Applied Microbiology*, 109(6), 2160–2172. <http://dx.doi.org/10.1111/j.1365-2672.2010.04848.x>.
- Roos, D. S., & Fichera, M. E. (1997). A plastid organelle as a drug target in apicomplexan parasites. *Nature*, 390(6658), 407–409. <http://dx.doi.org/10.1038/37132>.
- Rosenberg, A., & Gouaux, J. (1967). Quantitative and compositional changes in monogalactosyl and digalactosyl diglycerides during light-induced formation of chloroplasts in *Euglena gracilis*. *Journal of Lipid Research*, 8, 80–83.
- Rosenberg, A., Gouaux, J., & Milch, P. (1966). Monogalactosyl and digalactosyl diglycerides from heterotrophic, hetero-autotrophic, and photobiotic *Euglena gracilis*. 7, 733–738.
- Rybicka, K. K. (1996). Glycosomes—The organelles of glycogen metabolism. *Tissue and Cell*, 28(3), 253–265. [http://dx.doi.org/10.1016/S0040-8166\(96\)80013-9](http://dx.doi.org/10.1016/S0040-8166(96)80013-9).
- Sagan, L. (1993). On the origin of mitosing cells. 1967. *The Journal of NIH Research: Life Sciences Research and News about the National Institutes of Health and the Alcohol, Drug Abuse, and Mental Health Administration*, 5(3), 65–72. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11541390>.
- Saidha, T., & Schiff, J. A. (1989). The role of mitochondria in sulfolipid biosynthesis by *Euglena* chloroplasts. *Biochimica et Biophysica Acta (BBA)—Lipids and Lipid Metabolism*, 1001(3), 268–273. [http://dx.doi.org/10.1016/0005-2760\(89\)90110-0](http://dx.doi.org/10.1016/0005-2760(89)90110-0).
- Sanchez-Puerta, M. V., Lippmeier, J. C., Apt, K. E., & Delwiche, C. F. (2007). Plastid genes in a non-photosynthetic dinoflagellate. *Protist*, 158(1), 105–117. <http://dx.doi.org/10.1016/j.protis.2006.09.004>.

- Šantek, B., Felski, M., Friehs, K., Lotz, M., & Flaschel, E. (2010). Production of paramylon, a  $\beta$ -1,3-glucan, by heterotrophic cultivation of *Euglena gracilis* on potato liquor. *Engineering in Life Sciences*, 10(2), 165–170. <http://dx.doi.org/10.1002/elsc.200900077>.
- Schiff, J. A., Lyman, H., & Russell, G. K. (1971). Isolation of mutants from *Euglena gracilis*. *Methods in Enzymology*, 23, 143–162. [http://dx.doi.org/10.1016/S0076-6879\(71\)23088-3](http://dx.doi.org/10.1016/S0076-6879(71)23088-3).
- Schneider, T., & Betz, A. (1985). Waxmonoester fermentation in *Euglena gracilis* T. Factors favouring the synthesis of odd-numbered fatty acids and alcohols. *Planta*, 166(1), 67–73. <http://dx.doi.org/10.1007/BF00397387>.
- Schwartzbach, S. D. (2017). Photo and nutritional regulation of *Euglena* organelle development. In S. D. Schwartzbach & S. Shigeoka (Eds.), *Euglena: Biochemistry, cell and molecular biology* (pp. 159–182). Cham, Switzerland: Springer International Publishing. [http://dx.doi.org/10.1007/978-3-319-54910-1\\_9](http://dx.doi.org/10.1007/978-3-319-54910-1_9).
- Schwartzbach, S. D., & Schiff, J. A. (1974). Chloroplast and cytoplasmic ribosomes of *Euglena*: Selective binding of dihydrostreptomycin to chloroplast ribosomes. *Journal of Bacteriology*, 120(1), 334–341.
- Schwartzbach, S. D., Schiff, J. A., & Goldstein, N. H. (1975). Events surrounding the early development of *euglena* chloroplasts: V. Control of paramylum degradation. *Plant Physiology*, 56(2), 313–317. <http://dx.doi.org/10.1111/j.1751-1097.1976.tb06873.x>.
- Schwender, J., Goffman, F., Ohlrogge, J. B., & Shachar-Hill, Y. (2004). Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, 432(7018), 779–782. <http://dx.doi.org/10.1038/nature03145>.
- Shashidhara, L. S. S., Limsli, S. H., Shackletonii, J. B., Robinsonii, C., & Smith, A. G. (1992). Protein targeting across the three membranes of the *Euglena* chloroplast envelope. *The Journal of Biological Chemistry*, 267, 12885–12891.
- Sheiner, L., & Striepen, B. (2013). Protein sorting in complex plastids. *Biochimica et Biophysica Acta—Molecular Cell Research*, 1833(2), 352–359. <http://dx.doi.org/10.1016/j.bbamcr.2012.05.030>.
- Sheveleva, E. V., & Hallick, R. B. (2004). Recent horizontal intron transfer to a chloroplast genome. *Nucleic Acids Research*, 32(2), 803–810. <http://dx.doi.org/10.1093/nar/gkh225>.
- Shi, L.-X., & Theg, S. M. (2013). The chloroplast protein import system: From algae to trees. *Biochimica et Biophysica Acta*, 1833(2), 314–331. <http://dx.doi.org/10.1016/j.bbamcr.2012.10.002>.
- Shigeoka, S., Onishi, T., Nakano, Y., & Kitaoka, S. (1986). The contents and subcellular distribution of tocopherols in *Euglena gracilis*. *Agricultural and Biological Chemistry*, 50(4), 1063–1065. <http://dx.doi.org/10.1271/abb1961.50.1063>.
- Sláviková, S., Vacula, R., Fang, Z., Ehara, T., Osafune, T., & Schwartzbach, S. D. (2005). Homologous and heterologous reconstitution of Golgi to chloroplast transport and protein import into the complex chloroplasts of *Euglena*. *Journal of Cell Science*, 118(Pt. 8), 1651–1661. <http://dx.doi.org/10.1242/jcs.02277>.
- Stern, A. I., Epstein, H. T., & Schiff, J. A. (1964a). Studies of chloroplast development in *Euglena*. VI. Light intensity as a controlling factor in development. *Plant Physiology*, 39(2), 226–231.
- Stern, A. I., Schiff, J. A., & Epstein, H. T. (1964b). Studies of chloroplast development in *Euglena*. V. Pigment biosynthesis, photosynthetic oxygen evolution and carbon dioxide fixation during chloroplast development. *Plant Physiology*, 39(2), 220–226. <http://dx.doi.org/10.1104/pp.39.2.220>.
- Sugiyama, A., Suzuki, K., Mitra, S., Arashida, R., Yoshida, E., Nakano, R., et al. (2009). Hepatoprotective effects of paramylon, a beta-1, 3-D-glucan isolated from *Euglena gracilis* Z, on acute liver injury induced by carbon tetrachloride in rats. *The Journal of Veterinary Medical Science the Japanese Society of Veterinary Science*, 71(7), 885–890. <http://dx.doi.org/10.1292/jvms.71.885>.



- Sulli, C., Fang, Z. W., Muchhal, U., & Schwartzbach, S. D. (1999). Topology of Euglena chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *Journal of Biological Chemistry*, 274(1), 457–463. <http://dx.doi.org/10.1074/jbc.274.1.457>.
- Sulli, C., & Schwartzbach, S. D. (1995). The polyprotein precursor to the Euglena light-harvesting chlorophyll a/b-binding protein is transported to the Golgi apparatus prior to chloroplast import and polyprotein processing. *Journal of Biological Chemistry*, 270, 13084–13090. <http://dx.doi.org/10.1074/jbc.270.22.13084>.
- Sulli, C., & Schwartzbach, S. D. (1996). A soluble protein is imported into Euglena chloroplasts as a membrane-bound precursor. *The Plant Cell*, 8(1), 43–53.
- Teerawanichpan, P., & Qiu, X. (2010). Fatty acyl-CoA reductase and wax synthase from *Euglena gracilis* in the biosynthesis of medium-chain wax esters. *Lipids*, 45(3), 263–273. <http://dx.doi.org/10.1007/s11745-010-3395-2>.
- Thomas, B. G., & Threlfall, D. R. (1974). Synthesis of polyprenyltoluquinols from homogentisate and polyprenyl pyrophosphates in particulate fractions of *Euglena* and sugar beet. *Biochemical Journal*, 142, 437–440.
- Thompson, M. D., Copertino, D. W., Thompson, E., Favreau, M. R., & Hallick, R. B. (1995). Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Research*, 23(23), 4745–4752. <http://dx.doi.org/10.1093/nar/23.23.4745>.
- Triemer, R. E., Linton, E., Shin, W., Nudelman, A., Monfils, A., Bennett, M., et al. (2006). Phylogeny of the euglenales based upon combined SSU and LSU rDNA sequence comparisons and description of *Discoplastis* gen. nov. (Euglenophyta). *Journal of Phycology*, 42(3), 731–740. <http://dx.doi.org/10.1111/j.1529-8817.2006.00219.x>.
- Turmel, M., Gagnon, M.-C., O’Kelly, C. J., Otis, C., & Lemieux, C. (2009). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Molecular Biology and Evolution*, 26(3), 631–648. <http://dx.doi.org/10.1093/molbev/msn285>.
- Vesteg, M., Vacula, R., Steiner, J. M., Mateásiková, B., Löffelhardt, W., Brejová, B., et al. (2010). A possible role for short introns in the acquisition of stroma-targeting peptides in the flagellate *Euglena gracilis*. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 17(4), 223–231. <http://dx.doi.org/10.1093/dnares/dsq015>.
- Walne, P. L., & Arnott, H. J. (1967). The comparative ultrastructure and possible function of eyespots: *Euglena granulata* and *Chlamydomonas eugametos*. *Planta*, 77(4), 325–353. <http://dx.doi.org/10.1007/BF00389319>.
- Wang, J., Shi, Z., & Xu, X. (2004). Residual plastids of bleached mutants of *Euglena gracilis* and their effects on the expression of nucleus-encoded genes. *Progress in Natural Science*, 14(3), 213–217. <http://dx.doi.org/10.1080/10020070412331343371>.
- Watanabe, T., Shimada, R., Matsuyama, A., Yuasa, M., Sawamura, H., Yoshida, E., et al. (2013). Antitumor activity of the  $\beta$ -glucan paramylon from *Euglena* against preneoplastic colonic aberrant crypt foci in mice. *Food & Function*, 4(11), 1685. <http://dx.doi.org/10.1039/c3fo60256g>.
- Wheeler, G., Ishikawa, T., Pornsaksit, V., & Smirnoff, N. (2015). Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*, 4, 1–25. <http://dx.doi.org/10.7554/eLife.06369>.
- Wiegert, K. E., Bennett, M. S., & Triemer, R. E. (2012). Evolution of the chloroplast genome in photosynthetic euglenoids: A comparison of *Eutreptia viridis* and *Euglena gracilis* (Euglenophyta). *Protist*, 163(6), 832–843. <http://dx.doi.org/10.1016/j.protis.2012.01.002>.

- Wiegert, K. E., Bennett, M. S., & Triemer, R. E. (2013). Tracing patterns of chloroplast evolution in euglenoids: Contributions from *Colacium vesiculosum* and *Strombomonas acuminata* (Euglenophyta). *Journal of Eukaryotic Microbiology*, 60(2), 214–221. <http://dx.doi.org/10.1111/jeu.12025>.
- Wolfe, A. D., & dePamphilis, C. W. (1998). The effect of relaxed functional constraints on the photosynthetic gene *rbcL* in photosynthetic and nonphotosynthetic parasitic plants. *Molecular Biology and Evolution*, 15(10), 1243–1258. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a025853>.
- Wolken, J. J., & Mellon, A. D. (1956). The relationship between chlorophyll and the carotenoids in the algal flagellate, *Euglena*. *The Journal of General Physiology*, 39(5), 675–685.
- Yamaguchi, A., Yubuki, N., & Leander, B. S. (2012). Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: Description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evolutionary Biology*, 12(1), 29. <http://dx.doi.org/10.1186/1471-2148-12-29>.
- Yoshida, Y., Tomiyama, T., Maruta, T., Tomita, M., Ishikawa, T., & Arakawa, K. (2016). De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*, 17(1), 182. <http://dx.doi.org/10.1186/s12864-016-2540-6>.
- Záhonová, K., Füßy, Z., Oborník, M., Eliáš, M., Yurchenko, V., & Stern, D. (2016). RuBisCO in non-photosynthetic alga *Euglena Longa*: Divergent features, transcriptomic analysis and regulation of complex formation. *PloS One*, 11(7), e0158790. <http://dx.doi.org/10.1371/journal.pone.0158790>.
- Záhonová, K., Hadariová, L., Vacula, R., Yurchenko, V., Eliáš, M., Krajčovič, J., et al. (2014). A small portion of plastid transcripts is polyadenylated in the flagellate *Euglena gracilis*. *FEBS Letters*, 588(5), 783–788. <http://dx.doi.org/10.1016/j.febslet.2014.01.034>.
- Zakryś, B., Milanowski, R., & Karnkowska, A. (2017). Evolutionary origin of *Euglena*. In S. D. Schwartzbach & S. Shigeoka (Eds.), *Euglena: Biochemistry, cell and molecular biology* (pp. 3–17). Cham, Switzerland: Springer International Publishing. [http://dx.doi.org/10.1007/978-3-319-54910-1\\_1](http://dx.doi.org/10.1007/978-3-319-54910-1_1).