

**Univerzita Karlova v Praze**  
**Přírodovědecká fakulta**

Studijní program: Biologie

Studijní obor: Protistologie



Bc. Soňa Lukešová

**Diverzita prasinofytních řas příbuzných plastidu euglen**  
Diversity of prasinophyte algae related to the euglenid plastid

DIPLOMOVÁ PRÁCE

Vedoucí závěrečné práce: Doc. Mgr. Vladimír Hampl, Ph.D.

Praha, 2016

**Prohlášení:**

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

V Praze, 2. 5. 2016

Soňa Lukešová

## **Poděkování:**

Na tomto místě bych chtěla poděkovat vedoucímu mé diplomové práce Doc. Vladimíru Hamplovi za cenné rady, ochotu, trpělivost a čas, který mi po celou dobu magisterského studia věnoval. Velké poděkování patří rovněž Dr. Janě Szabové za počáteční pomoc s laboratorními metodami. Ráda bych také poděkovala Dr. Anně Karnkowské za pomoc při skládání plastidového genomu řasy *Pterosperma cristatum* a s vyhledáváním TARA sekvencí, Mgr. Petru Soukalovi za pomoc s fylogenetickými analýzami a Dr. Štěpánce Hrdé za introdukci k programu Geneious. Děkuji také všem ostatním členům naší laboratoře Evoluční protistologie za vstřícnost a vytvoření příjemného prostředí. V neposlední řadě bych chtěla poděkovat své rodině za podporu a motivaci, kterou mi po celou dobu vzniku této práce dodávala.

## Abstrakt

Euglenophyceae představují skupinu jednobuněčných eukaryotických organismů, které v evoluci získaly schopnost fotosyntézy. Jejich plastidy vznikly sekundární endosymbiózou s primární zelenou řasou. Fylogenetické studie pátrající po původu této organely ukázaly na zelenou řasu *Pyramimonas parkeae* ze skupiny Prasinophyta jako nejbližšího známého příbuzného plastidům euglen. Plastidové genomy *Pyramimonas parkeae* a *Euglena* sdílí unikátní pořadí určitých genů sdružených do genových klastrů, což také ukazuje na skupinu Pyramimonadales jako donora plastidu. Není nicméně vyloučené, že se v prostředí vyskytují organismy, které by byly plastidům euglen ještě příbuznější než *P. parkeae*. Ve své diplomové práci jsem se proto zaměřila na zkoumání diverzity Pyramimonadales a Euglenophyceae v environmentálních vzorcích. Použila jsem k tomu několik přístupů. Pomocí specifické PCR jsem amplifikovala části plastidového genomu v environmentálních vzorcích a určila jsem jejich pozici na fylogenetickém stromu. Provedla jsem také několik rozsáhlých fylogenetických analýz založených na 16S rRNA a 18S rRNA sekvencích z dostupných zástupců Euglenophyceae, Prasinophyta a environmentálních vzorků. Výsledky odhalily přítomnost velkého množství environmentálních sekvencí spadajících mezi Pyramimonadales, organismy příbuznější plastidu než ty dosud známé však odhaleny nebyly. Analýzy odkryly rovněž dříve nepozorované environmentální mořské sekvence větící se uvnitř Euglenophyceae, a to i v rámci Euglenales považované dosud za čistě sladkovodní. Součástí práce byla také sekvenace a anotace plastidového genomu *Pterosperma cristatum*, řasy ze skupiny Pyramimonadales, která představuje další bod pro rekonstrukci evoluce endosymbiotické události. Analýza plastidových genů Euglenophyceae, *Pterosperma* a zelené linie řas a rostlin potvrdila původ plastidů euglen ve skupině Pyramimonadales. Předpokládané stáří této endosymbiózy bylo molekulárním datováním odhadnuto na 372,8–280,6 milionů let.

**Klíčová slova:** Prasinophyta, euglenida, plastid, environmentální sekvenování, fylogeneze

## Abstract

Euglenophyceae represent a group of unicellular eukaryotic organisms that have gained during their evolution the ability to photosynthesize. They acquired plastids via secondary endosymbiosis with a green alga as the plastid donor. Phylogenetic studies searching for the origin of this organelle revealed the green alga *Pyramimonas parkeae* from Prasinophytes as the closest known relative to euglenid plastids. *Pyramimonas parkeae* and *Euglena* share several genes clusters with unique order of genes in their plastid genomes, which also point to the Pyramimonadales as the donor of the plastids. However, it is possible, that organisms more closely related to euglenid plastids than *P. parkeae*, occur in the environment. In my diploma thesis I focused on the exploration of diversity of Pyramimonadales and Euglenophyceae in environmental samples. I used several approaches to perform this task. I amplified parts of the plastid genomes in environmental samples by using specific PCR and determined their position in the phylogenetic tree. I also made large-scale phylogenetic analyses based on 16S rRNA and 18S rRNA sequences from representatives of the groups Euglenophyceae, Prasinophytes and environmental samples. The results revealed the presence of a large number of environmental sequences relative to the Pyramimonadales and also previously undetected amount of environmental marine sequences branching within Euglenophyceae, even within Euglenales considered so far as strictly freshwater group. No lineage, which would be more closely related to euglenid plastid than already known lineages, has been detected. The other part of my work was focussed on sequencing and annotation of the plastid genome of Pyramimonadales species *Pterosperma cristatum*, which represent another node for the reconstruction of the evolution of plastids. Analyses of plastid genes from Euglenophyceae, *Pterosperma* and other green algae confirmed the origin of euglenid plastids in Pyramimonadales. Expected age of the endosymbiotic event was estimated using molecular dating to 372.8–280.6 million years.

**Keywords:** Prasinophytes, euglenids, plastid, environmental sequencing, phylogeny

## Obsah

Abstrakt.....	4
Abstract.....	5
1. Úvod.....	8
2. Přehled literatury.....	9
2.1. Evoluce fototrofie u euglenidů.....	9
2.2. Charakteristika a diverzita Euglenophyceae.....	9
2.3. Euglenophyceae a sekundární endosymbióza.....	10
2.3.1. Hledání donora chloroplastu.....	11
2.4. Prasinofytní řasy.....	12
2.4.1. Morfologické znaky a taxonomie.....	12
2.4.2. Fylogeneze.....	16
2.4.3. Původ euglenidího chloroplastu v Pyramimonadales.....	16
3. Cíle práce.....	18
4. Materiál a metodika.....	19
4.1. Použitý materiál.....	19
4.2. Izolace DNA.....	21
4.3. Amplifikace částí plastidového genomu prasinofyt.....	21
4.3.1. Navržení primerů a PCR amplifikace.....	21
4.3.2. Elektroforéza a purifikace PCR produktu.....	22
4.3.3. Klonování.....	23
4.3.4. Sekvence DNA.....	24
4.4. Analýza plastidového genomu <i>Pterosperma cristatum</i> .....	24
4.4.1. Sekvence a assembly.....	25
4.4.2. Kompletace genomu.....	25
4.4.3. Anotace plastidových genů.....	27
4.5. Fylogenetické analýzy.....	28
4.5.1. Datasetsy.....	28
4.5.2. Tvorba alignmentů.....	29
4.5.3. Tvorba fylogenetických stromů.....	30
4.5.4. Molekulární datování.....	31
5. Výsledky.....	33

5.1.	Analýza genových klastrů .....	33
5.2.	Fylogeneze na základě 18S rRNA.....	35
5.3.	Fylogeneze na základě 16S rRNA.....	37
5.4.	Plastidový genom <i>Pterosperma cristatum</i> .....	40
5.5.	Fylogenomická analýza.....	46
5.6.	Molekulární datování .....	47
6.	Diskuse .....	49
6.1.	Fylogeneze Pyramimonadales a Euglenophyceae na základě 16S a 18 rRNA .....	49
6.2.	Analýzy environmentálních sekvencí .....	50
6.2.1.	Amplifikace částí plastidového genomu .....	50
6.2.2.	Sladkovodní skupiny Pyramimonadales .....	52
6.2.3.	Mořské linie Euglenales .....	52
6.2.4.	Úskalí environmentálního sekvenování .....	53
6.3.	Plastidový genom <i>Pterosperma cristatum</i> .....	53
6.3.1.	Rozdíly v genovém obsahu prasinofytních řas a Euglenophyceae.....	55
6.4.	Fylogenomická analýza.....	56
6.5.	Datování endosymbiotické události .....	57
7.	Závěr .....	58
8.	Seznam použitých zkratk.....	59
9.	Seznam použité literatury.....	60

## 1. Úvod

Endosymbióza, úzký vztah mezi dvěma organismy, kde jeden žije v těle druhého, zásadně ovlivnila formování života na Zemi. Předcházela vzniku některých důležitých buněčných organel eukaryot, jako jsou mitochondrie a plastidy. Plastidy zajišťují fotosyntézu, a mají tak velký význam nejen pro organismy samotné, ale také pro fungování globálních ekosystémů. Endosymbiotická teorie (Sagan, 1967) říká, že vznikly pohlcením řasy jiným eukaryotickým organismem. První z účastníků měl již nějakou podobu fungujícího fotosyntetického aparátu a stal se donorem této organely. Druhým aktérem byl heterotrofní organismus se schopností fagocytózy. Rozlišujeme několik typů plastidů podle endosymbiotické události, při které vznikly. Polapení sinice dalo zrod primárním plastidům a celé říši Archaeplastida zahrnující Glaucophyta, červenou (Rhodophyta) a zelenou (Viridiplantae) linii řas a rostlin. Chromatofor, plastidům podobná organela nacházející se u améb rodu *Paulinella* z kmene Cercozoa, vznikla také primární endosymbiózou (Marin et al., 2005). Sekundární a terciární endosymbiózy se potom účastnila řasa s již etablovaným plastidem a další heterotrofní eukaryotický organismus. Získání sekundárního plastidu se odehrálo několikrát nezávisle na sobě v různých liniích eukaryot. Jednu z takovýchto skupin představují Euglenida, skupina známá také pod českým názvem krásnoočka.

Ve své diplomové práci jsem se zaměřila na řešení otázky původu plastidů fototrofních euglen. Nejblíže známým příbuzným plastidům euglen je prasinofytní řasa *Pyramimonas parkeae* (Turmel et al., 2009). Pyramimonadales nicméně představují poměrně diverzifikovanou skupinu s množstvím popsáných zástupců. Environmentální studie zkoumající skrytou diverzitu moří a oceánů kromě toho odhalily značnou rozmanitost prasinofytních řas v environmentálních vzorcích (Guillou et al., 2004; Viprey et al., 2008; de Vargas et al., 2015). V prostředí by se proto mohl nacházet organismus, který by byl bližší plastidům euglen než zmiňovaný *P. parkeae*. Diverzitu prasinofytních a euglenofytních řas jsem posuzovala fylogenetickými analýzami plastidových genů dostupných sekvencí z kultivovatelných zástupců i environmentálních sekvencí. Snažila jsem se v prostředí nalézt další linie blízké příbuzné plastidům Euglenophyceae. Za tímto účelem jsem navrhla specifické primery, kterými jsem amplifikovala z environmentálních vzorků části plastidových genomů sdílených řasami *Pyramimonas* a *Euglena*. Zaměřila jsem se také na sekvenaci a anotaci plastidového genomu *Perosperma cristatum*, řasy ze skupiny Pyramimonadales, která představuje další bod pro rekonstrukci fylogeneze Euglenophyceae. A pokusila jsem se rovněž datovat endosymbiotickou událost, jež vedla ke vzniku sekundárních plastidů euglen.



## 2. Přehled literatury

### 2.1. Evoluce fototrofie u euglenidů

Euglenida představují významnou skupinu eukaryotických organismů s rozsáhlou škálou potravních strategií. Někteří zástupci se živí osmotrofně absorbováním živin skrze membránu (např. rody *Rhodomonas*, *Distigma*). U jiných se vyvinula buněčná ústa, která jim umožňují fagocytovat kořist. Podle toho, jaký typ kořisti pohlcují, je můžeme dále rozdělit na bakterivorní (např. rody *Petalomonas*, *Ploetia*, *Entosiphon*) a eukaryvorní druhy (např. rody *Peranema*, *Heteronema*, *Urceolus*). Velká skupina euglenidů získává energii ze světelného záření za pomoci fotosyntézy. Zvláštním typem potravní strategie je mixotrofie. Mixotrofie kombinuje dva druhy výživy, heterotrofní a autotrofní způsob (např. *Rapaza viridis*) (Yamaguchi et al., 2012).

Životní strategie se odráží v morfologii jednotlivých taxonů a odpovídá také evoluci euglenidů (Leander et al., 2001; Leander, 2004). Kladistická analýza Leandra et al. (2001) ukazuje, že první euglenida byla pravděpodobně bakterivorní. Bazální skupina kinetoplasteí (Bodonida), sesterské třídy euglenidů, se živí také převážně pohlcováním bakterií. Společný předek obou skupin byl tudíž s největší pravděpodobností rovněž volně žijící bakterivor. Vytvoření specializovaných buněčných úst a zmnožení počtu pelikulárních pruhů vyztužujících buňku umožnilo pohlcování větší kořisti a vznik eukaryvorních linií. Z nich se nezávisle na sobě vyvinuli primární osmotrofové a fototrofové s redukovanými buněčnými ústy (Leander et al., 2001). Fagocytující aparát je nicméně zásadní předpoklad pro polapení endosymbionta a vznik fotosyntetické organely (Leander, 2004). Přejídným článkem mezi eukaryvorním a fototrofním způsobem výživy by proto mohla být mixotrofie (Yamaguchi et al., 2012). Mixotrofní *Rapaza viridis* má již ustálený vlastní chloroplast, zůstává však stále závislá na lovu kořisti. Důležitost tohoto stupně v evoluci euglenidů potvrzuje také postavení *Rapaza viridis* v SSU rRNA analýze, kde zaujímá pozici oddělující heterotrofní od fototrofních euglen (Yamaguchi et al., 2012).

V průběhu evoluce došlo u některých fototrofních zástupců ke ztrátě plastidu, a stali se tak opět sekundárně osmotrofními (např. *Euglena longa*). Fotosyntetizující euglenida a všichni jejich heterotrofní potomci vytvářejí monofyletický klad, což ukazuje na společný zisk endosymbionta (Marin, 2004). Tento klad je klasifikován jako třída Euglenophyceae.

### 2.2. Charakteristika a diverzita Euglenophyceae

Euglenophyceae zahrnují volně žijící organismy obývající sladkovodní, ale i mořské prostředí. Typicky vřetenovitá buňka nese jeden až dva bičíky vycházející z flagelární kapsy. Vždy jsou ale přítomny oba bičíky, jeden však může být velmi zkrácený a skrytý uvnitř flagelární kapsy. V cytoplazmě vedle flagelární kapsy se nachází stigma složená z lipidových kapének obsahující

karotenoidy, jež spolu s fotoreceptory na bázi dlouhého bičíku slouží k fotorecepci. Povrch buňky vyztužuje pelikula složená z proteinových pásů podložených mikrotubuly (Dragos et al., 1997). Jednotlivé pásy k sobě těsně přiléhají a vytvářejí souvislou podporu cytoplazmatické membrány. Pevnost pelikuly určuje míru metabolie, tedy to, jak je buňka pohyblivá, či naopak rigidní. Pod pelikulou se nacházejí speciální tělíska mukocysty, které vylučují na povrch slizovitou hmotu. Ta může být mineralizována (Pereira et al., 2003) za tvorby hnědě zbarvené loriky charakteristické pro rody *Trachelomonas* a *Strombomonas* či v případě rodu *Colacium* sloužit jako mucilózní stopka, kterou se její nositel přichytává k substrátu nebo na povrch svých hostitelů (Al-Dhaheri a Willey, 1996). Mimoto je vylučovaný materiál důležitý pro formování odolných palmelových stádií (Triemer, 1980), jež jsou klíčové pro přežití nepříznivých podmínek. V buňce se nachází klasická eukaryotická výbava. Důležitými organelami fototrofních euglen jsou chloroplasty obalené třemi membránami. Obsahují chlorofyl a+b,  $\beta$  karoteny a malé množství xantofylů. Tvar, velikost a počet chloroplastů se různí druh od druhu. Zásobní látkou je paramylon ( $\alpha$ -1,3 glukán), který se hromadí v zrnech rozmanitých tvarů (Monfils et al., 2011).

Fototrofní euglenida můžeme rozdělit do dvou řádů: Euglenales a Eutreptiales. Skupina Eutreptiales se v molekulárních analýzách větví na bázi fylogenetického stromu Euglenophyceae (Marin et al., 2003; Kim et al., 2010; Linton et al., 2010; Karnkowska et al., 2015). Zahrnuje jediné dva mořské rody *Eutreptia* a *Eutreptiella*, které pro pohyb v prostředí používají oba bičíky. Známí zástupci Euglenales jsou oproti tomu výhradně sladkovodní organismy s jedním dlouhým a druhým zkráceným bičíkem. Zahrnují celkem 11 rodů. Molekulární data ukazují na jejich rozdělení do dvou monofyletických skupin Euglenaceae a Phacaceae (Kim et al., 2010; Karnkowska et al., 2015). Morfologickými synapomorfiemi, které jasně vyznačují čeleď Phacaceae, jsou malé parietálně situované chloroplasty bez pyrenoidů (Kim et al., 2010). Patří sem rody *Phacus*, *Lepocinclis* a *Discoplastis*. Skupina Euglenaceae zahrnuje tradiční rody *Trachelomonas*, *Strombomonas*, *Colacium*, *Monomorphina*, *Cryptoglana*, a *Euglena*. Rod *Euglena* představuje nejpočetnější skupinu v rámci druhové rozmanitosti. Ačkoliv jsou si jednotliví členové morfologicky hodně podobní, molekulární analýzy odhalily další dva samostatné rody: *Euglenaria* (Linton et al., 2010) a *Euglenaformis* (Bennett et al., 2014), jež zahrnuje jediný druh *Euglenaformis proxima* (dříve známý jako *Euglena proxima*).

### 2.3. Euglenophyceae a sekundární endosymbióza

Chloroplasty fototrofních euglen mají původ v sekundární endosymbióze s primární zelenou řasou (Gibbs, 1981). Hlavním důkazem pro tento původ je přítomnost tří obalových membrán chloroplastů, výskyt totožných fotosyntetických pigmentů jako u zelených řas a fylogeneze plastidových genů (Turmel et al., 2009; Bennett et al., 2014). Vzájemná podobnost v genovém obsahu mezi plastidovými

genomy jednotlivých zástupců Euglenophyceae (Wiegert et al., 2012) navíc ukazuje, že v evoluci vznikly pravděpodobně pouze jedinou endosymbiotickou událostí.

Během procesu endosymbiózy docházelo k postupným změnám polapené primární řasy. Symbiont prošel obrovskou redukcí buněčných organel a ostatních kompartmentů. Mitochondrie, bičík, Golgiho aparát, endoplazmatické retikulum a v případě euglenofyt dokonce i jádro zanikly (Keeling, 2010). Výsledný sekundární chloroplast ve výsledku tvoří pouze fotosyntetizující plastid a systém membrán, které ho obklopují. Oproti jiným sekundárním chloroplastům, které jsou obklopené čtyřmi membránami, obalují ty euglenofytní pouze tři membrány. Předpokládá se, že vnitřní dvě membrány pochází z primárního chloroplastu. Třetí membrána může být odvozená buď od hostitelského endoplazmatického retikula, nebo od plasmatické membrány donora (Keeling, 2010). K původům jednotlivých membrán však neexistuje přímý důkaz.

Genom získaného endosymbionta prošel navíc rozsáhlými změnami. Většina genů zodpovědných za správnou funkci chloroplastu byla přesunuta z jaderného genomu donorové buňky do jádra hostitelského organismu. Velmi brzy došlo rovněž k velkým ztrátám jaderných, ale i plastidových genů endosymbionta, jak bylo prokázáno srovnáním plastidových genomů *Eutreptiella gymnastica* a *Euglena gracilis* (Hrdá et al., 2012). Výsledný chloroplast si zachoval pouze některé geny nezbytné pro schopnost samotné fotosyntézy. Pořadí genů v genomech se však mezi jednotlivými zástupci značně liší, za což zodpovídá přeskupování genů v rámci genomu (Hrdá et al., 2012; Wiegert et al., 2012).

### 2.3.1. Hledání donora chloroplastu

Spolu se skupinou Chlorarachniophyta a obrněnkami rodu *Lepidodinium* představují Euglenophyceae jediné fototrofní řasy, jejichž chloroplasty pocházejí ze zelených řas. Cavalier-Smith (1999) v minulosti postuloval hypotézu Cabozoa, která tvrdí, že chloroplasty skupiny Chlorarachniophyta a Euglenophyceae vznikly společnou endosymbiotickou událostí. Hypotéza nebyla nicméně nikdy fylogeneticky podpořena, neboť členové skupiny Chlorarachniophyta se na základě analýzy plastidových genů (Rogers et al., 2007; Takahashi et al., 2007; Turmel et al., 2009) větví s odlišnými skupinami zelených řas, než je tomu v případě Euglenophyceae. Skutečnost, že jsou chloroplastové genomy euglenidů a chlorarachniofyt více příbuzné jiným řasám než sobě navzájem (Rogers et al., 2007), poukazuje spíše na nezávislý získ plastidů. Podobná situace platí rovněž pro plastidy obrněnek rodu *Lepidodinium*, které se ve fylogenetických analýzách větví se zcela odlišnou skupinou zelených řas (Kamikawa et al., 2015).

Analýza plastid-kódujícího rRNA operonu (Marin, 2004) odhalila příbuznost plastidů Euglenophyceae s prasinofytní linií zelených řas. Prasinophyta představují převážně mořské jednobuněčné organismy. To souhlasí s představou, že k endosymbióze došlo v moři (Marin, 2004),

protože bazální větve skupiny zaujímají právě moššší zástupci rodu *Eutreptia* a *Eutreptiella* (Marin et al., 2003; Kim et al., 2010; Linton et al., 2010; Karnkowska et al., 2015).

## 2.4. Prasinofytní řasy

Prasinofytní řasy spadají do linie zelených řas, označované také jako Viridiplantae či Chlorobionta. Charakteristickými znaky Viridiplantae je přítomnost primárních chloroplastů se dvěma obalovými membránami, tylakoidy uspořádané do lamel a fotosyntetické pigmenty ve složení chlorofyl a+b a přídatné xantofyly a karotenoidy. Dělí se na dvě vývojové linie Chlorophyta a Streptophyta, která mimo jiné obsahují také vyšší rostliny (Embryophyta). Prasinofytní řasy vytvářejí parafyletickou třídu Prasinophyceae a zaujímají bazální pozice v chlorofytní linii (Leliaert et al., 2011). Zahrnují převážně mořské planktonní druhy, najdou se však mezi nimi i sladkovodní zástupci (Swale a Belcher, 1968; Marin a Melkonian, 2010; Yamaguchi et al., 2011).

### 2.4.1. Morfologické znaky a taxonomie

Hlavním znakem skupiny Prasinophyta je povrch krytý vrstvami organických šupin, které vznikají v Golgiho aparátu (Becker et al., 1994). Někteří příslušníci této skupiny nicméně postrádají šupiny, bičíky nebo dokonce obojí (Miyashita et al., 1993; Marin a Melkonian, 2010), a neexistuje proto morfologický nebo biochemický znak, který by jasně celou skupinu sjednotil. Většina prasinofytních řas dorůstá velmi malých rozměrů a jsou rozšířené součástí mořského pikoplanktonu (Guillou et al., 2004; Vaulot et al., 2008; Viprey et al., 2008; Shi et al., 2009). Prasinofytní skupiny se liší především tvarem a velikostí buňky, morfologií šupin, počtem a stylem pohybu bičíků, mechanismem dělení a biochemickými vlastnostmi, jako je například přítomnost přídatných pigmentů a zásobních látek. Fylogenetické studie využívající molekulární data jako SSU rRNA (Nakayama et al., 1998) a SSU rDNA (Fawley et al., 2000; Guillou et al., 2004; Viprey et al., 2008) odhalily celkem devět samostatných prasinofytních kladů. Taxonomické rozdělení skupiny Prasinophyta do jednotlivých kladů mimo jiné koresponduje také s obsahem přídatných fotosyntetických pigmentů v buňce (Latasa et al., 2004). V následujícím textu se budu věnovat krátkému popisu jednotlivých prasinofytních kladů.

#### Klad I - Pyramimonadales

Do kladu I spadají především členové řádu Pyramimonadales rozdělení do dvou linií (Viprey et al., 2008). První z nich obsahuje mimo jiné rody *Cymbomonas* a *Halospharea*. Do druhé linie patří kromě rodu *Pyramimonas* také *Pterosperma*. Členové tohoto kladu mají v pohyblivém stádiu minimálně čtyři bičíky (u některých zástupců 8 nebo 16 bičíků) vycházející z anteriorní části buňky. Typický je také systém uspořádání bičíkových kořenů. Buňka i bičík jsou kryté organickými šupinami uspořádanými

v několika vrstvách. Někteří zástupci (*Pterosperma*, *Halospharea*) vytváří ve svém životním cyklu odolné cysty hojně ve fosilním záznamu (Dotzler et al., 2007; Kustatscher et al., 2014). U jiných příslušníků kladu (*Cymbomonas*) byla prokázána schopnost fagocytózy, která by mohla být pozůstatkem heterotrofního předka zelených řas (Maruyama a Kim, 2013).

Nejpočetnější rod tohoto kladu *Pyramimonas* zahrnuje mořské, brakické i sladkovodní druhy. Typická buňka má více či méně pyramidový tvar. Obsahuje jeden pohárovitě tvarovaný chloroplast, který se v přední části rozděluje v několik laloků. Součástí chloroplastu je také pyrenoid a světločivná skvrna. Šupiny pokrývající buňku a bičíky vznikají v Golgiho aparátu (Becker et al., 1994) a jsou shromažďovány ve speciálním rezervoáru. Odtud jsou později uvolňovány kanálkem do flagelární kapsy a dále na povrch. Na základě ultrastrukturálních a biochemických vlastností byl rod *Pyramimonas* dále rozdělen do několika podrodů (McFadden et al., 1986; Hori et al., 1995). Molekulární studie založené na analýze velké podjednotky genu pro RuBisCO (*rbcL*) potvrdily existenci oddělených skupin, avšak jejich vzájemné postavení na fylogenetickém stromu se v různých studiích liší (Daugbjerg et al., 1994; Suda, 2004; Harðardóttir et al., 2014).

Velikost a tvar buňky, vzájemné uspořádání bazálních tělísek bičíků, struktura šupinového pokryvu, umístění světločivné skvrny a pyrenoidu představují důležité znaky pro odlišování jednotlivých podrodů. Podrod *Pyramimonas* zahrnuje bentické a epifytické druhy s variabilním počtem bičíků (*P. tetra-rhynchus* - 4, *P. propulsa* - 8, *P. cyrptoptera* - 16). Zmnožení počtu bičíků přitom může být adaptací na bentický způsob života (Daugbjerg et al., 1994). Do podrodu *Vestigifera* spadají druhy vyznačující se nepřítomností podkladové vrstvy drobných šupin. Patří sem například *P. mantoniae* či *P. orientalis*. *Trichocystis* a *Punctatae* představují podrody s podobnými morfologickými znaky. Hlavním rozdílem je přítomnost trichocyst u *Trichocystis*, zatímco členové podrodu *Punctatae* obsahují zvláštní mucilózní tělíska (Hori et al., 1995). Dobře známým zástupcem skupiny *Trichocystis* je *P. parkeae*. Na rozdíl od ostatních vychází podrod *Punctatae* ve fylogenetických analýzách jako polyfyletický taxon (Suda, 2004; Harðardóttir et al., 2014). Jeho členové se totiž rozpadají do několika oddělených linií. Poslední dva podrody jsou zastoupené pouze jedním druhem: *P. virginica* reprezentuje *Hexactis* a *P. longicauda* byla přiřazená do skupiny *Macrura* (Hori et al., 1995).

Oproti rodu *Pyramimonas* zahrnuje rod *Pterosperma* pouze mořské organismy. Ve svém životním cyklu vytváří dvě odlišná stádia. První z nich je kokoidní cysta phycoma, která se pomocí bočních křidélek okružujících celou buňku vznáší ve vodním sloupci. Phycoma obsahuje množství chloroplastů s pyrenoidem. Jednotlivé druhy se liší právě především v morfologii této cysty, jež může nabývat rozmanitých tvarů. Buněčné dělení probíhá ve zralé cystě za vzniku pohyblivých buněk se čtyřmi velmi dlouhými bičíky s unikátním flagelárním aparátem (Inouye et al., 1990). Motilní stádia obsahují pouze jeden chloroplast s pyrenoidem, extrusomy a povrch mají krytý organickými šupinami.

## **Klad II - Mamiellophyceae**

Nejdominantnější prasinofytní klad II neboli Mamiellophyceae s množstvím popsáných zástupců se rozděluje do tří oddělených linií: Mamiellales, Dolichomastigales a Monomastigales (Marin a Melkonian, 2010). Skupina Mamiellales zahrnuje dva bezbičíkaté rody *Bathycoccus* a *Ostreococcus*, ale také rody *Micromonas*, *Mamiella* a *Mantoniella* s jedním nebo dvěma bičíky. Do linie Dolichomastigales spadají dvoubičíkaté rody *Crustomastix* a *Dolichomastix*. Skupinu Monomastigales reprezentuje jediný rod *Monomastix*, který se vyznačuje přítomností pouze jednoho bičíku. Charakteristickými znaky skupiny jsou: nepřítomnost čtvercově tvarovaných šupin v podkladové vrstvě, jak je tomu u ostatních Prasinophyceae, a výskyt jedné až dvou vrstev pavučinově tvarovaných šupin na buňce, případně bičíku (Nakayama et al., 1998). Jeden, výjimečně dva chloroplasty mohou obsahovat pyrenoid a světločivnou skvrnu. Dominantním přídatným fotosyntetickým pigmentem je prasinoxanthin (Latasa et al., 2004). Příslušníci tohoto kladu obývají převážně mořské prostředí, najdou se mezi nimi však i ryze sladkovodní skupiny (Monomastigales).

*Bathycoccus*, *Ostreococcus* a *Micromonas* jsou skupiny zahrnující velmi drobné organismy hojně zastoupené v mořském pikoplanktonu (Guillou et al., 2004; Viprey et al., 2008; Shi et al., 2009). *Ostreococcus tauri* představuje vůbec nejmenší volně žijící eukaryotický organismus. Kvůli své malé velikosti má značně redukovaný genom a buněčnou strukturu, která se skládá z jádra, jednoho chloroplastu a mitochondrie. *Bathycoccus* vytváří homogenní klad, neboť se jeho sekvence z různých oblastí liší pouze nepatrně (Guillou et al., 2004). *Ostreococcus* a *Micromonas* jsou oproti tomu více diverzifikované a dělí se do několika ekotypových podskupin (Viprey et al., 2008).

## **Klady III (Nephroselmidaceae) a V (Pycnococcaceae)**

Příbuzenské vztahy mezi členy původní skupiny Pseudoscorfieldiales zahrnující rody *Nephroselmis* a *Pseudoscourfieldia* byly v minulosti velmi sporné. Později se ukázalo, že si tyto dvě skupiny nejsou vůbec příbuzné (Fawley et al., 1999) a vytváří dva oddělené klady III (Nephroselmidaceae) a V (Pycnococcaceae), kam kromě rodu *Pseudoscourfieldia* spadá také rod *Pycnococcus* (Fawley et al., 2000).

Skupina Nephroselmidaceae sestává pouze z jediného popsáného rodu *Nephroselmis*. Poměrně velká buňka je laterálně zploštělá se dvěma ventrálně umístěnými heterodynamickými bičíky. Bičíky i povrch buňky pokrývají šupiny. Tvar a rozmístění šupin na těle buňky charakterizuje jednotlivé druhy (Nakayama et al., 2007; Yamaguchi et al., 2011). Většina známých zástupců obývá mořské prostředí. Najdeme mezi nimi však i organismy hojné ve sladkovodních rybnících a jezerech (*N. olivacea*). Jelikož tento sladkovodní zástupce vytváří ve fylogenetických analýzách jedinou linii mezi mořskými organismy, usuzuje se, že společný předek skupiny žil s největší pravděpodobností rovněž v moři (Yamaguchi et al., 2011).

*Pycnococcus provasolii* je velmi drobná kulovitá řasa postrádající šupiny i bičík. Větší *Pseudoscourfieldia marina* má oproti tomu dva bičíky pokryté šupinami. Ačkoliv se tyto dva stěžejně

druhy Pycnococcaceae morfologicky liší, jejich sekvence jsou si velmi podobné, a mohly by proto představovat odlišná životní stádia téhož organismu (Fawley et al., 1999). Jejich hlavním přídatným fotosyntetickým pigmentem je prasinoxanthin (Latasa et al., 2004).

#### **Klad IV - Chlorodendrophyceae**

Tradičně byla mezi Prasinophyta zařazována také skupina Chlorodendrophyceae se dvěma rody *Tetraselmis* a *Sherffelia* (Nakayama et al., 1998). Hlavním důvodem byla jejich morfologie podobná ostatním prasinofytním skupinám. Z anteriorní části oválné buňky vychází čtyři bičíky kryté šupinami. Povrch buněk členů rodu *Tetraselmis* nekryjí typické prasinofytní šupiny, nýbrž théka vzniklá splynutím jednotlivých šupin (Becker et al., 1994). Jediný chloroplast s pyrenoidem se v přední části dělí na laloky. Molekulárních data (Nakayama et al., 2007) a později i srovnávací analýzy plastidových genomů (Turmel et al., 2016) nicméně prokázaly příbuznost Chlorodendrophyceae s korunními skupinami chlorofyt zahrnující bazální linii Pedinophyceae a UTC klad (Ulvoephyceae, Trebouxiophyceae a Chlorophyceae). Tomu rovněž odpovídá sdílení některých znaků mezi Chlorodendrophyceae a chlorofytními skupinami (Trebouxiophyceae a Chlorophyceae), jako jsou uzavřená mitóza a fykoplást (Leliaert et al., 2011).

#### **Klad VI - Prasinococcales**

Klad VI (Prasinococcales) zahrnuje dva mořské rody *Prasinococcus* a *Prasinoderma*. Zástupci této skupiny mají kulovité buňky bez bičíku a bazálního tělíska. Ačkoliv postrádají šupinový povrch, buňky obklopuje silná buněčná stěna. *P. capsulatus* představuje jediný popsáný druh rodu *Prasinococcus*. Kromě buněčné stěny jeho povrch kryje také tlustá želatinózní matrix elipsoidního nebo kapkovitého tvaru zvaná capsula (Miyashita et al., 1993). Členové rodu *Prasinoderma* oproti tomu obsahují několikvrstevnou buněčnou stěnu a mohou se vyskytovat buď samostatně (*P. singularis*) nebo tvoří volné lepidivé kolonie (*P. coloniale*) (Jouenne et al., 2011). Hlavní charakteristikou skupiny je přítomnost jednoho miskovitého chloroplastu obsahujícího pyrenoid obklopený škrobovou pochvou, do kterého proniká mitochondrie ležící v blízkosti chloroplastu (Miyashita et al., 1993; Jouenne et al., 2011). Nejdůležitějšími přídatnými fotosyntetickými pigmenty jsou prasinoxanthin a uriolid (Latasa et al., 2004).

#### **Klady VII, VIII a IX**

Drobný (2-3  $\mu\text{m}$ ) kulovitý organismus *Picocystis salinarum* (Lewin et al., 2000) spolu s několika environmentálními sekvencemi náleží do kladu VII. *Picocystis salinarum* je charakteristický přítomností trilaminární buněčné stěny. Jeho buňka obsahuje jeden prostorný chloroplast typicky miskovitého tvaru bez pyrenoidu. Vyskytuje se ve slaných jezerech, odkud byl také poprvé izolován (Lewin et al., 2000). Příbuzné environmentální sekvence byly kromě toho dokumentovány jako součást mořského pikoplanktonu (Vaulot et al., 2008; Shi et al., 2009).

Klady VIII a IX zahrnují pouze environmentální sekvence izolované ze Středozemního moře (Viprey et al., 2008) a Tichého oceánu (Shi et al., 2009). Zatímco klad VIII reprezentují pouze tři sekvence získané z povrchových vod Středozemního moře, klad IX je mnohem více diverzifikován a vytváří dvě oddělené podskupiny (Viprey et al., 2008).

#### **2.4.2. Fylogeneze**

Fylogenetické studie zkoumající vztahy uvnitř Prasinophyceae jsou založené především na malé ribozomální podjednotce SSU rDNA (Guillou et al., 2004; Nakayama et al., 2007), chloroplastových genech (Turmel et al., 2009) a proteinech (Lemieux et al., 2014). Nejhlubší větev všech chlorofyt dle těchto analýz reprezentuje klad VI, Prasinococcales. Tuto pozici utvrzuje především přítomnost šesti chloroplastových genů, které se nenachází u žádného jiného chlorofyta (Lemieux et al., 2014). Ke ztrátě těchto genů pravděpodobně došlo u jejich společného předka těsně po oddělení od Prasinococcales. Dalšími odštěpujícími se liniemi jsou Pyramimonadales (klad I) a Mamiellophyceae (klad II). Fylogenetické studie se nepatrně liší v řešení příbuzenského vztahu mezi těmito dvěma skupinami. Pyramimonadales a Mamiellophyceae buď vytváří sesterské skupiny (Nakayama et al., 2007; Turmel et al., 2009; Lemieux et al., 2014), nebo se skupina Pyramimonadales odvětvuje ještě před Mamiellophyceae (Guillou et al., 2004). Podpora větvení druhé topologie je nicméně poměrně slabá. Podobně se rozcházejí studie také ve vysvětlení dalších oddělujících se linií Nephroselmidaceae (klad III) a Pycnococcaceae (klad V). V některých fylogenetických analýzách vytváří obě skupiny s nízkou podporou společný klad (Guillou et al., 2004). V jiných studiích se jako první odvětvuje skupina Pycnococcaceae, až potom následuje Nephroselmidophyceae (Nakayama et al., 2007), nebo naopak (Lemieux et al., 2014). Do poslední prasinofytní skupiny, která je zároveň linií ležící nejbliže korunním skupinám chlorofyt (Pedinophyceae, Trebouxiophyceae, Ulvophyceae a Chlorophyceae), přísluší zástupci a environmentální vzorky kladu VII. Zatímco v SSU rDNA analýzách vytvářejí souhrnnou skupinu (Guillou et al., 2004; Nakayama et al., 2007), ve fylogenezi na základě chloroplastových proteinů tvoří několik nezávislých linií (Lemieux et al., 2014).

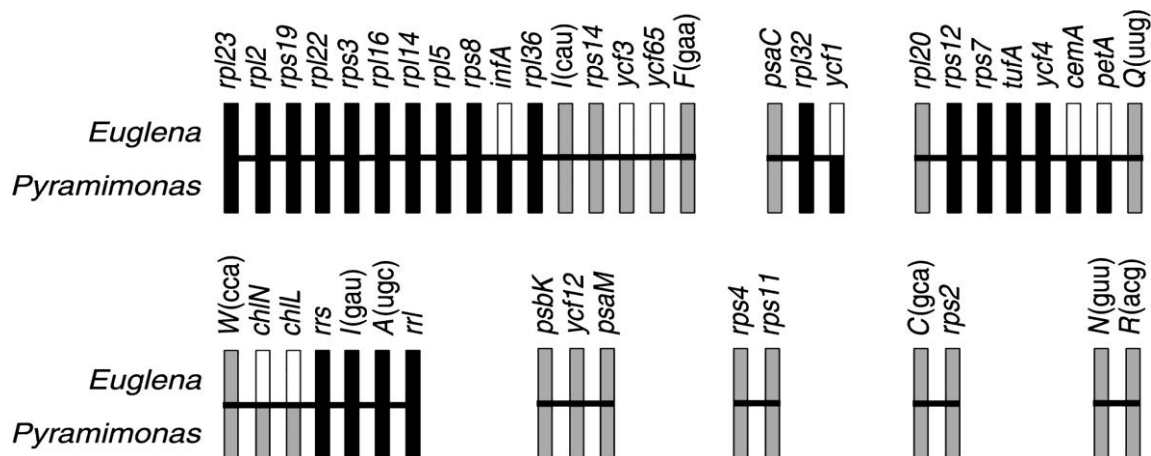
Fylogenetické analýzy spojují Chlorodendrophyceae, skupinu v minulosti zařazovanou mezi Prasinophyta, s korunními skupinami chlorofyt (Guillou et al., 2004; Nakayama et al., 2007; Turmel et al., 2016), ale jejich přesné postavení není zatím úplně dořešeno. Rovněž příbuzenský vztah mezi ryze environmentálními klady VIII a IX a ostatními prasinofytními liniemi zůstává kvůli nedostatečné statistické podpoře prozatím nejasný (Viprey et al., 2008).

#### **2.4.3. Původ euglenidího chloroplastu v Pyramimonadales**

S přibývajícím počtem molekulárních dat získávala hypotéza původu chloroplastu euglen v prasinofytních řasách stále větší podporu. Otázkou nicméně zůstávalo, která z dnešních prasinofytních linií byla donorem plastidu nebo byla tomuto donoru nejvíce příbuzná. Srovnávací



genomová analýza Turmelové et al. (2009) porovnávající cpDNA (chloroplastové DNA) vybraných zástupců prasinofytních řas a Euglenophyceae ukázala na skupinu Pyramimonadales. Ze studie vyplynulo zajímavé zjištění, že *Pyramimonas* a *Euglena* sdílí unikátní pořadí určitých genů sdružených do genových klastrů (Obr. 1), které se nenachází u žádných jiných zelených řas. Spolu s přečtením a porovnáním cpDNA dalších fototrofních euglenidů (Hallick et al., 1993, Hrdá et al., 2012; Wiegert et al., 2012; Wiegert et al., 2013; Bennett et al., 2014; Bennett a Triemer, 2015) představuje tato skutečnost silný důkaz, že chloroplasty euglen pochází ze skupiny Pyramimonadales.



**Obr. 1. Genové klastry unikátně sdílené cpDNA organismů *Euglena* a *Pyramimonas* podle Turmelové et al. (2009).** Černě vybarvená políčka naznačují ancestrální uspořádání těchto klastrů (vyskytující se i u jiných zelených řas), šedá políčka odkazují na derivované oblasti (specifické pro *Euglena* a *Pyramimonas*) a prázdná políčka označují chybějící geny.

Dnes tedy víme, že si Euglenophyta podmanila symbionta nezávisle na jiných fototrofních řasách a že donorem tohoto plastidu byl nejspíš člen skupiny Pyramimonadales (Turmel et al., 2009). Neexistují nicméně žádné další studie zkoumající v tomto ohledu, kromě jediného druhu *Pyramimonas parkeae*, ještě jiné zástupce ze skupiny Pyramimonadales či environmentální vzorky. Naše vědomosti o dané skupině organismů jsou často omezené pouze na ty zástupce, které umíme kultivovat. Celková diverzita může být však ve skutečnosti mnohem větší. Environmentální studie prokázaly existenci nových prasinofytních linií (Fawley et al., 2001; Viprey et al., 2008), které rozšiřují diverzitu skupiny. Není proto zcela vyloučeno, že se v prostředí nachází také organismy bližší chloroplastu euglen než *Pyramimonas parkeae*, pomocí nichž by bylo možné ještě lépe osvětlit evoluci chloroplastu u Euglenophyceae.

### 3. Cíle práce

(1) Hledání DNA stop prasinofytních řas blízce příbuzných plastidům euglen v environmentálních vzorcích pomocí specifických primerů.

(2) Anotace genomu *Pterosperma cristatum*, prasinofytní řasy ze skupiny Pyramimonadales.

(3) Rozsáhlé fylogenetické analýzy 16S rRNA, 18S rRNA a sekvencí plastidových proteinů od dostupných zástupců Prasinophyceae, Euglenophyceae a environmentálních sekvencí s cílem lépe zmapovat diverzitu prasinofytních linií řas blízkých plastidům euglen.

(4) Datování endosymbiotické události, která dala vzniku sekundárním plastidům Euglenophyceae, pomocí molekulárních hodin.

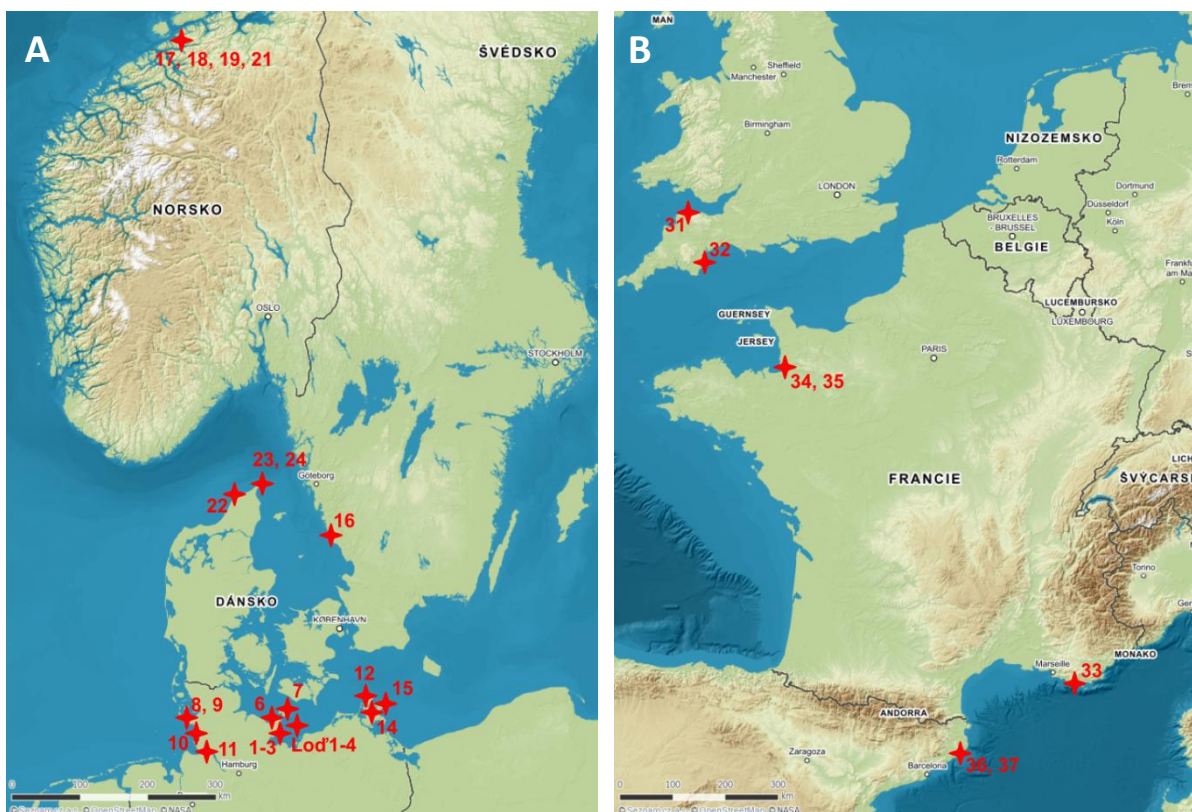
## 4. Materiál a metodika

### 4.1. Použitý materiál

Kultury druhů *Pyramimonas* pocházejí ze Scandinavian Culture Collection of Algae & Protozoa (SCCAP): *P. orientalis* (K-0003), *P. olivacea* (K-0258), *P. mantoniae* (K-0254), *P. disomata* (K-0285), *P. parkeae* (K-0007) a *P. tetrahynechus* (K-0002). Kultury byly pěstovány v LT30 médiu o složení: 1 ml zásobního roztoku  $\text{NaNO}_3$ , 1 ml zásobního roztoku  $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ , 3 ml půdního extraktu a 1 ml roztoku s příměsí sloučenin kovů. Více o přípravě a složení média je shrnuto na webových stránkách SCCAP (<http://www.sccap.dk/media/marine/1.asp>). Každé tři týdny byly kultury sterilní pipetou převedeny do čerstvého média. Vzhledem k vysoké náchylnosti na převoz jsme bohužel nezískali kulturu organismu *Pterosperma cristatum*. Japonský National Institute for Environmental Studies (NIES) nám nicméně poskytl již vyextrahovanou DNA.

Environmentální vzorky byly odebírány mými vědeckými spolupracovníky při zahraničních cestách. Vzorky pocházejí především z mořských nebo brakických vod. K jejich zachycení byla použita planktonka (velikost ok 20  $\mu\text{m}$ ) házená ze břehu či z paluby lodi. Po odebrání byly vzorky zakonzervovány etanolem v konečné koncentraci 70 %. Odběrová místa jsou vyznačena v mapách A-D, tabulka 1 shrnuje podrobnější informace o jednotlivých vzorcích.

**Mapa A-D. Místa odběrů environmentálních vzorků.** (A) severní Evropa a Německo, (B) jihozápadní Evropa, (C) Kanada, (D) Japonsko. Označení vzorků odpovídá tabulce 1.





**Tab. 1. Informace o odběrových místech environmentálních vzorků.**

Název vzorku	Místo odběru	Stanoviště	Souřadnice
Lod'1	Baltské moře, Meklenburský záliv	moře	N 54° 23,110' E 11° 9,960'
Lod'2	Baltské moře, Meklenburský záliv	moře	N 54° 24,050' E 11° 12,300'
Lod'3	Baltské moře, Meklenburský záliv	moře	N 54° 23,860' E 11° 11,612'
Lod'4	Baltské moře, Meklenburský záliv	moře	N 54° 23,550' E 11° 9,300'
Německo1	Německo, kemp Walkyrien	sladké bahno	N 54° 7,068' E 10° 55,097'
Německo2	Německo, kemp Walkyrien	sladké bahno	N 54° 7,072' E 10° 55,041'
Německo3	Německo, Meklenburský záliv	moře	N 54° 6,858' E 10° 55,158'
Německo6	Německo, Kielský záliv	moře	N 54° 22,751' E 11° 4,683'
Německo7	Německo, Fehmarn	moře	N 54° 31,541' E 11° 7,277'
Německo8	Německo, Sankt Peter-Ording	potok	N 54° 18,456' E 8° 36,517'
Německo9	Německo, Sankt Peter-Ording	moře	N 54° 17,682' E 8° 37,148'
Německo10	Německo, Büsum	slaná nádrž	N 54° 8,273' E 8° 50,509'
Německo11	Německo, Brokdorf	ústí řeky	N 53° 52,460' E 9° 17,470'
Německo12	Německo, Putgarten	moře	N 54° 39,944' E 13° 25,960'
Německo14	Německo, Rujána	moře	N 54° 30,892' E 13° 24,767'
Německo15	Německo, Rujána	moře	N 54° 34,189' E 13° 40,177'
Švédsko16	Švédsko, Träslövsläge	moře	N 57° 3,671' E 12° 16,627'
Norsko17-21	Norsko, Eidestrand, Aure	moře	
Dánsko22	Dánsko, Hirtshals	moře	N 57° 35,245' E 9° 56,543'
Dánsko23	Dánsko, Skagen	moře	N 57° 44,352' E 10° 37,938'
Dánsko24	Dánsko, Skagen	moře	N 57° 44,782' E 10° 39,052'
Kanada25-30	Kanada, Vancouver Island	moře	
Halifax	Kanada, Halifax	moře	
Anglie31	Anglie, Lynmouth	moře	N 51° 13,963' W 3° 50,089'
Anglie32	Anglie, Paington	moře	N 50° 25,951' W 3° 33,472'
Francie33	Francie, Hyères	moře	N 43° 3,820' E 6° 8,848'
Francie34-35	Francie, Le Mont Saint-Michel	moře	
Španělsko36-37	Španělsko, Saint Feliu de Guixols	moře	
Tokyo1-2	Japonsko, Tokyo	moře	

## 4.2. Izolace DNA

Kulturu či environmentální vzorek (12-15 ml) jsem centrifugovala po dobu pěti minut při 3000 G. Poté jsem opatrně odsála téměř všechny supernatant. Pelet obsahující buňky jsem rozsuspendovala v malém množství zbylého roztoku. Izolaci DNA jsem prováděla pomocí kitu DNeasy® Blood & Tissue Kit (Qiagen) dle přiloženého protokolu *Cultured Cells*. Koncentraci izolované DNA jsem poté měřila na spektrofotometru (NanoDrop® ND-1000). Extrahovaná DNA byla do použití skladována při teplotě -20 °C.

## 4.3. Amplifikace částí plastidového genomu prasinofyt

Již výše jsem zmínila, že *Pyramimonas* a *Euglena* sdílí v plastidovém genomu unikátní pořadí některých genů sdružených do genových klastrů (Turmel et al., 2009). Jim příbuzné organismy by měly rovněž v genomu tyto klastry obsahovat. Pomocí specifických primerů jsem se snažila jednotlivé genové klastry amplifikovat v environmentálních vzorcích, a tak získat sekvence blízké chloroplastu euglen. Pro analýzu byly zvoleny následující genové klastry: rps4-rps11 a trnC(gca)-rps2 (Obr. 1).

### 4.3.1. Navržení primerů a PCR amplifikace

Pro každý genový klastr jsem manuálně navrhla sadu specifických primerů (Tab. 2). Jako templát, podle kterého jsem primery navrhovala, byl použit alignment vytvořený z DNA sekvencí zástupců Euglenophyceae (*Colacium vesiculosum*, *Euglena gracilis*, *Euglena viridis*, *Eutreptia viridis*, *Eutreptiella gymnastica*, *Monomorphina aenigmatica* a *Strombomonas acuminata*) a prasinofytní řasy *Pyramimonas parkeae* dostupných v NCBI (<http://www.ncbi.nlm.nih.gov/>). Všechny primery byly před použitím naředěny na koncentraci 10 pmol/μl.

PCR amplifikaci jsem prováděla s polymerázou Emerald Amp® GT PCR Master Mix (Clontech). Aby bylo množství DNA z environmentálních vzorků dostatečné k detekci na gelové elektroforéze, bylo zapotřebí provést ještě sekundární PCR. Do sekundární reakce jsem použila 1,5 μl produktu z reakce předchozí. Jako pozitivní kontrola sloužila DNA z organismu *Pyramimonas parkeae*. Negativní kontrola obsahovala namísto DNA 1,5 μl H<sub>2</sub>O. Složení reakcí je shrnuto v tabulce 3 a průběh cyklu v tabulce 4.

**Tab. 2. Použité primery pro amplifikaci genových klastrů.**

Název použitého primeru	Genový klastr	Sekvence primeru 5'→3'
rps_4-F	rps4-rps11	GGACCWCGWTTTRMGAATWGTTTCGT
rps_4_nested-F	rps4-rps11	GAATWGTTTCGTMGWMTWGGT
rps_4_novy-F	rps4-rps11	GGAATATGTCWMGDTATMGMGGWCC
rps_4_nes_novy-F	rps4-rps11	MWTTNTTAGARATGMGWYTWGATAATAT
rps_11-R	rps4-rps11	GGTCGACAWCCATTATGYGGTA
rps_11_nested-R	rps4-rps11	GGTAMGGGWGTDAYATCTCT
rps_11_novy-R	rps4-rps11	TTNSKWGGHCKRCAWCCATTATG
rps_11_nes_novy-R	rps4-rps11	CCRCAWVMWCCWGMAGAHGMCCA
trnC-F	trnC(gca)-rps2	GGCGRCATVGCCAAGYGGTAA
trnC_nest-F	trnC(gca)-rps2	GGCARWRGATTGCAAATCYTT
trnC_nest-F2	trnC(gca)-rps2	TTTATYCCCCAGTTCRAATCTGGGTG
rps2-R	trnC(gca)-rps2	GAATCATCATNGCHGGWAT
rps2_nest-R	trnC(gca)-rps2	GGATCACARTTTGTATCWA
rps2_nest-R2	trnC(gca)-rps2	TTDWWRCATTCTYTHACMGCRTT

**Tab. 3. Složení PCR reakce pro Emerald Amp® GT PCR Master Mix.**

Složka	Použité množství
Master Mix	12,5 µl
primer F	1,5 µl
primer R	1,5 µl
DNA	1,5 µl
H <sub>2</sub> O	8 µl

**Tab. 4. Cyklus PCR amplifikace s Emerald Amp® GT PCR Master Mixem.**

Část cyklu	Teplota	Čas	Opakování
Počáteční denaturace	95 °C	4 min	1x
Denaturace	95 °C	0:30 min	↑ 40x
Annealing	50 °C	1 min	
Elongace	72 °C	1:30 min	
Závěrečná elongace	72 °C	15 min	1x
Uchování vzorku	12 °C	∞	

#### 4.3.2. Elektroforéza a purifikace PCR produktu

Přítomnost amplifikované DNA ve vzorku byla detekována horizontální gelovou elektroforézou. Připravila jsem 1% agarózový gel smícháním 0,4 g agarózy se 40 ml pufru TAE (1x). Směs jsem zahřívala, dokud se agaróza úplně nerozpustila. Po následném zchlazení jsem do gelu přidala 40 µl barviva Sybr pro vizualizaci DNA. Následně jsem nechala hotovou směs ztuhnout v předem připravené komůrce s hřebínkem s odpovídajícím počtem a velikostí zubů. Poté jsem do první jamky

v řadě nanasla pipetou 4 µl DNA markeru Gene Ruler DNA Ladder Mix (Thermo Scientific), který slouží pro odečítání velikostí DNA fragmentů. Do zbylých jamek jsem nanasla celý obsah (25 µl) PCR produktu environmentálních vzorků. Pokud jsem analyzovala DNA z kultur, použila jsem pouze 5 µl z PCR produktů, neboť to bylo dostačující pro vizualizaci přítomné DNA. Elektroforéza běžela při 95 V zhruba 30-40 min. Následně byl gel vyfocen pod UV světlem.

PCR produkty pocházející z environmentálních sekvencí byly získávány vyříznutím z gelu a následně přečištěny pomocí kitu Zymoclean™ Gel DNA Recovery Kit (Zymo Research) nebo QIAquick® Gel Extraction Kit (Qiagen). Vzhledem k dostatečnému množství DNA v PCR produktech amplifikovaných z kultur nebylo zapotřebí použít celý objem vzorku k detekci na elektroforéze. V takovém případě bylo možné izolovat PCR produkty ze zbylého množství vzorku za použití DNA Clean & Concentrator™-5 (Zymo Research). PCR produkty jsem eluovala do 20 µl EB pufru. Koncentraci získané DNA jsem změřila na spektrofotometru (NanoDrop® ND-1000).

### 4.3.3. Klonování

PCR produkty pocházející z environmentálních vzorků, které při sekvenaci vykazovaly dvojitý signál nebo nebylo jejich množství dostatečné pro sekvenaci, bylo zapotřebí zaklonovat do plasmidu. Klonování jsem prováděla s kitem pGEM®-T Easy Vector (Promega) za použití kompetentních buněk *Escherichia coli* XL1. Postupovala jsem dle přiloženého protokolu. Namíchala jsem ligační směs (Tab. 5), kterou jsem nechala přes noc inkubovat v lednici při teplotě 4°C.

**Tab. 5. Složení ligační směsi používané při klonování.**

Činidlo	Použité množství
2x ligační pufr	5 µl
pGEM vektor (50ng)	1 µl
PCR produkt	0,5-1 µl
T4 DNA ligáza	1 µl
sterilní H <sub>2</sub> O	zbytek do 10 µl

Druhý den jsem transformovala připravené plazmidy s vloženým inzertem DNA do bakteriálních buněk. Kompetentní buňky jsem nechala rozmrazit na ledu. Poté jsem k nim přidala krátce centrifugovanou ligační směs (Tab. 5), lehce jsem celý obsah mikrozkušavky promíchala a inkubovala na ledu po dobu 20 minut. Následně jsem provedla teplotní šok umístěním mikrozkušavky na 45 s do 42°C a ihned zpátky na led. Po dvou minutách jsem do vzorku přidala 250 µl SOC média (Tab. 6) a inkubovala na třepačce (150 rpm) při 37°C po dobu 1,5 h. Na LB plotny s agarem jsem nanasla směs skládající se z: 20 µl X-Galu, 25 µl antibiotika ampicilinu a 100 µl IPTG. Po zaschnutí jsem na plotny sterilní kličkou rozetřela směs s narostlými bakteriemi. Na jednu plotnu jsem přitom použila 200 µl a na druhou zbytek směsi. Plotny byly přes noc inkubovány v termostatu při 37°C a později skladovány v lednici při 4°C.

**Tab. 6. Složení SOC média.**

Složka média	Množství
bacto trypton	2 g
bacto yeast extract	0,5 g
1M NaCl	1 ml
1M KCl	250 µl
1M MgCl <sub>2</sub>	1 ml
1M MgSO <sub>4</sub>	1 ml
1M sterilní glukóza	2 ml
sterilní H <sub>2</sub> O	zbytek do 100 ml

Narostlé bakterie s požadovaným inzertem vytváří oproti bakteriím bez inzertu bílé kolonie. Pro kontrolu slouží kolony PCR. Do mikrozkušavky s 9 µl sterilní H<sub>2</sub>O jsem špičkou přenesla část vybrané kolonie. Namíchala jsem směs pro PCR a amplifikovala daný úsek se specifickými primery. Následně jsem PCR produkty detekovala na horizontální gelové elektroforéze. Na elektroforézu jsem použila jenom část vzorku, zbytek jsem přečistila kitem DNA Clean & Concentrator<sup>TM</sup>-5 (Zymo Research) a použila na sekvenaci.

#### 4.3.4. Sekvence DNA

Materiálem pro sekvenci byla přečištěná DNA izolovaná z PCR produktů. Sekvenační reakce jsem míchala do celkového objemu 8 µl. Do odměřené PCR vody jsem přidala 3,2 pmol primeru a požadované množství přečištěné DNA (vztaženo na délku fragmentu). Samotná sekvenace byla provedena Laboratoří sekvenace DNA na Přírodovědecké fakultě Univerzity Karlovy v Praze. Přečtené sekvence jsem skládala za pomoci nástroje *De novo Assemble* v programu Geneious v9.0.5 (Kearse et al., 2012). Začátky a konce složených kontigů jsem manuálně ořízla tak, aby sekvence neobsahovaly samotné sekvence primerů. Výsledný konsensus sekvencí jsem exportovala ve formátu FASTA, který jsem následně použila ve fylogenetické analýze.

#### 4.4. Analýza plastidového genomu *Pterosperma cristatum*

Prasinofytní řasa *Pterosperma cristatum* představuje svou pozicí na bázi kladu *Pyramimonas* další bod pro rekonstrukci původu chloroplastu u euglen. Dosud bylo v NCBI publikováno několik chloroplastových genů pocházejících z tohoto organismu. Pro účely multiproteinových fylogenetických analýz je však potřeba většího množství dat. Proto jsem se zaměřila ve své práci také na sekvenaci a anotaci plastidového genomu tohoto organismu.



#### 4.4.1. Sekvence a assembly

Sekvence DNA pocházející z řasy *Pterosperma cristatum* byla provedena metodou Illumina (Illumina MiSeq) v laboratoři Oddělení genomiky a bioinformatiky na Ústavu molekulární genetiky AV ČR v Praze. Rovněž prvotní poskládání (dále assembly) jednotlivých „paired end“ čtení vykonala tato laboratoř za použití nástrojů SOAP2 (Li et al., 2009) a SPAdes (Bankevich et al., 2012). Pro posouzení kvality jednotlivých assembly jsem porovnávala parametry výsledných assembly (délky sekvencí scaffoldů) a provedla lokální blastovací analýzy v programech BioEdit (Hall, 1999) a Standalone BLAST (Tao, 2014). Vyhledávala jsem scaffoldy (překrývající se kontigy oddělené gapy známé délky) obsahující sekvence plastidových genů *Pterosperma cristatum* dostupných v NCBI. Ukázalo se, že assembly vytvořené nástrojem SPAdes je vhodnější pro samotné poskládání genomu, vykazovalo totiž mnohem delší sekvence scaffoldů než ostatní.

Následně už bylo možné vyhledávat a případně spojovat scaffoldy pocházející z plastidového genomu. S touto prací mi velmi pomohla Dr. Anna Karnkowská. Nejprve jsme našly scaffoldy obsahující sekvence plastidových genů *Pterosperma cristatum* již dříve publikovaných v NCBI (atpA, atpB, petD, psaA, psaB, psbA, psbB, psbC, psbD, rbcL, tufA, SSU rRNA, LSU rRNA). Plastidový genom nicméně obsahuje mnohem větší množství genů. Za účelem jejich vyhledání jsme mapovaly scaffoldy genomu *P. cristatum* na plastidové geny příbuzného organismu *Pyramimonas parkeae* dostupné v NCBI. Přítomnost konzervovaných genů (Turmel et al., 2009; tabulka S1) jsme zjišťovaly pomocí rychlého nástroje vyvinutého pro anotaci plastidových a mitochondriálních genomů DOGMA (Wyman et al., 2004), který je volně dostupný na internetovém serveru (<http://dogma.cccb.utexas.edu/>). Nakonec jsme nalezené scaffoldy podrobily několika kolům mapování v programu Geneious v9.0.5 (Kearse et al., 2012). Během toho se veškerá čtení mapují proti již vybraným scaffoldům. Poté se mapující čtení znovu přeskládají do scaffoldů. Celý proces jsme několikrát opakovaly a výsledkem byly delší kusy plastidového genomu. Mapování rovněž poskytuje dobrý přehled o správnosti složení scaffoldů a kvalitě pokrytí čteními. V průběhu celého procesu jsme jednotlivé kusy manuálně spojovaly do větších celků. K tomu nám opět sloužila anotace nástrojem DOGMA (Wyman et al., 2004), která nám ukázala, jaké části scaffoldů se shodovaly a bylo možné je v daném místě spojit.

#### 4.4.2. Kompletace genomu

Bohužel se nám nepodařilo vytvořit jeden, nýbrž dva výsledné plastidové scaffoldy. Za účelem osekvenování chybějících částí genomu jsem navrhla primery specifické ke koncům sekvencí scaffoldů (Tab. 7). Primery jsem navrhovala v místě, ve kterém byly anotovány chloroplastové geny zjištěné rychlou analýzou na serveru DOGMA (Wyman et al., 2004). Na každém konci sekvencí scaffoldů nasedal jeden primer tak, aby syntéza polymerázou směřovala ven, a byl tak amplifikován úsek mezi nimi. PCR amplifikaci jsem prováděla s polymerázovým mixem Advantage® 2 PCR

Enzyme System (Clontech). Celkový objem reakce byl 25  $\mu$ l (Tab. 8). Pozitivní a negativní kontrolu jsem použila dle protokolu. Cykly jednotlivých reakcí jsem nastavila dle přiloženého protokolu s přidáním fáze annealingu (Tab. 9). Postupně jsem vyzkoušela všechny možné kombinace navržených primerů.

**Tab. 7. Primery použité pro kompletaci plastidového genomu *P. cristatum*.**

Název použitého primeru	Scaffold	Sekvence primeru 5'→3'
Ptcr12a	1	GAGTGTCAACCAGATGCGGCGT
Ptcr12b	1	GCTTTGACCTAGAACTGTTGC
Ptcr3a	2	CTCCTGACCTGCCGTCACAAT
Ptcr3b	2	CCATGACGGTGACATATCTTG

**Tab. 8. Složení PCR reakce pro Advantage® 2 PCR Enzyme System.**

Složka	Použité množství
10x Advantage 2 PCR pufr	2,5 $\mu$ l
DNA templát	0,5 $\mu$ l
směs primerů	1,0 $\mu$ l
50x DNTP Mix (10 mM)	0,5 $\mu$ l
50x Advantage 2 polymeráza	0,5 $\mu$ l
H <sub>2</sub> O	20 $\mu$ l

**Tab. 9. Cyklus PCR amplifikace s Advantage® 2 PCR Enzyme System.** \* - hodnota teploty se měnila dle optimálních hybridizačních teplot primerů.

Část cyklu	Teplota	Čas	Opakování
Počáteční denaturace	95 °C	1 min	1x
Denaturace	95 °C	0:30 min	↑ 35x
Annealing	55/60 °C*	0:30 min	
Elongace	68 °C	12 min	
Závěrečná elongace	68 °C	12 min	1x
Uchování vzorku	12 °C	$\infty$	

Přítomnost DNA v PCR produktech jsem detekovala pomocí gelové elektroforézy. Vzhledem k poměrně dlouhé délce amplifikovaných oblastí (nejdelší až 15 kbp) jsem se rozhodla pro užití velké elektroforézy s 0,8% gelem připravené smícháním 0,96 g agarózy se 120 ml pufru TAE (1x). Kromě běžného DNA markeru GeneRuler DNA Ladder Mix (Thermo Scientific) jsem navíc použila ještě GeneRuler High Range DNA Ladder (Thermo Scientific) navržený pro odečítání dlouhých DNA fragmentů.

#### 4.4.3. Anotace plastidových genů

Téměř veškerou práci jsem dělala v programu Geneious v9.0.5 (Kearse et al., 2012). Nejprve jsem užitím nástrojů *Find ORFs* a *Annotate from* hledala otevřené čtecí rámce (ORFs) a jim odpovídající geny. Podkladem pro nalezení odpovídajících genů mi byla lokální databáze, kterou jsem vytvořila z plastidových genomů prasinofytních řas dostupných v NCBI: *Micromonas* sp. (FJ858267), *Monomastix* sp. (FJ493497), *Nephroselmis astigmatica* (NC\_024829), *Nephroselmis olivacea* (AF137379), *Ostreococcus tauri* (NC\_008289), *Picocystis salinarum* (NC\_024828), *Prasinococcus* sp. (KJ746597), *Prasinoderma coloniale* (NC\_024817), *Pycnococcus provasolii* (FJ493498) a *Pyramimonas parkeae* (NC\_012099). Postupně jsem snižovala úroveň similarity při hledání ORF, aby mi neunikl žádný z potenciálních genů. Kontrolou pro nalezení všech genů mi byl výsledek anotace nástrojem DOGMA (Wyman et al., 2004) a tabulka S1 shrnující konzervované geny cpDNA (chloroplastové DNA) u chlorofyt (Turmel et al., 2009). Plastidové geny nezjištěné automatickou anotací jsem následně dohledávala užitím tBLASTn proti nukleotidovým sekvencím obou scaffoldů. Jako query mi sloužila proteinová sekvence daného genu stažená z NCBI. Na závěr jsem predikovala tRNA (translační RNA). Pro tento účel mi posloužil internetový server tRNAscan-SE 1.21 (Schattner et al., 2005; <http://lowelab.ucsc.edu/tRNAscan-SE/>).

Nalezené geny a jejich hranice bylo nutné ještě ověřit. Postupovala jsem následujícím způsobem. Nejdříve jsem ze scaffoldů extrahovala sekvence daných genů, které jsem následně přeložila do proteinů podle translační tabulky 11 využívané kromě bakterií také pro plastidové genomy. Z NCBI jsem si stáhla proteinové, v případě RNA nukleotidové, sekvence jednotlivých genů prasinofytních řas (výše zmíněných). Spolu s odpovídajícím predikovaným genem řasy *Pterosperma cristatum* jsem z nich vytvořila alignment nástrojem *Pairwise/Multiple Align*. Tímto způsobem jsem kontrolovala, jestli sekvence nalezených genů souhlasí se sekvencemi příbuzných organismů. Případné nesrovnalosti v sekvencích jsem se snažila individuálně řešit, například hledáním alternativních start kodonů, posunováním čtecího rámce, zjišťováním přítomnosti intronů. Veškeré predikované geny jsem zanesla přímo do konsenzuálních sekvencí jednotlivých scaffoldů užitím nástroje *Add annotation*. Nakonec jsem ke každé anotované části plastidového genomu připojila informační tabulku obsahující podrobnější údaje o konkrétních oblastech. Na závěr jsem analyzovala frekvence užívání jednotlivých kodonů. Použila jsem k tomu proteinové sekvence všech anotovaných genů (kromě nedořešených částečných a hypotetických genů) a nástroj *Codon Usage* dostupný na internetovém serveru ([http://www.bioinformatics.org/sms2/codon\\_usage.html](http://www.bioinformatics.org/sms2/codon_usage.html)).

## 4.5. Fylogenetické analýzy

Nedílnou součástí práce představovaly fylogenetické analýzy prasinofytních řas a Euglenophyceae. Kromě analýzy genových klastrů jsem se zaměřila také na další molekulární markery: 16S rRNA, 18S rRNA a plastidové proteiny. Sekvence jsem v databázích vyhledávala užitím algoritmu BLASTn. Jako query posloužily sekvence známých zástupců.

### 4.5.1. Datasetsy

#### Chloroplastové genové klastry

Základem datasetů pro genové klastry rps4-rps11 a trnC(gca)-rps2 byly sekvence zástupců Euglenophyceae (*Colacium vesiculosum*, *Euglena gracilis*, *Euglena viridis*, *Eutreptia viridis*, *Eutreptiella gymnastica*, *Monomorphina aenigmatica* a *Strombomonas acuminata*) a prasinofytní řasy *Pyramimonas parkeae* stažené z NCBI. K těmto sekvencím jsem přidala všechny sekvence získané amplifikací jednotlivých genových klastrů v environmentálních vzorcích.

#### Chloroplastová 16S rRNA

16S rRNA dataset obsahoval především nukleotidové sekvence členů skupin Prasinophyta, Euglenophyceae a environmentální sekvence stažené z NCBI, dále environmentální sekvence získané od Dr. Aleše Horáka z analýzy Janouškovec et al. (2012) a také krátké sekvence z expedice TARA (Sunagawa et al., 2015). Hlavním cílem expedice TARA je prozkoumání skryté diverzity světových oceánů. Svým konáním již přispěla vědecké komunitě obrovským množstvím environmentálních sekvencí. Za účelem vyhledávání TARA sekvencí jsem použila Standalone BLAST (Tao, 2014). Z příslušné 16S rRNA databáze (Sunagawa et al., 2015) jsem nejprve vytvořila lokální databázi, proti které jsem následně blastovala 16S rRNA sekvence z prasinofytních (*Ostreococcus tauri* AY702141.1, *Prasinoderma coloniale* LN735212.2, *Picocystis salinarum* FN563078.1, *Pycnococcus provasolii* LN735480.3, *Pyramimonas parkeae* LN735435.3) a euglenofytních řas (*Colacium vesiculosum* AF289238.1, *Euglena gracilis* X70810.2, *Euglena viridis* AF289248.1, *Eutreptia viridis* AF289247.1, *Eutreptiella gymnastica* FJ719672.1, *Monomorphina aenigmatica* EU221498.1, *Strombomonas acuminata* EU221506.1). Hity s největší shodou jsem poté vybrala do analýzy.

#### 18S rRNA – Euglenophyceae

Analýzu založenou na 18S rRNA jsem rozdělila do několika datasetů zvlášť pro Euglenophyceae a Prasinophyta. Hlavním důvodem bylo, že si Euglenophyceae a Prasinophyta na 18S rRNA nejsou příbuzné. Euglenofytní 18S rRNA dataset zahrnoval kromě sekvencí pocházejících z NCBI také sekvence stažené z databáze iMicrobe (<http://data.imicrobe.us/>).

## 18S rRNA – Prasinophyta

Tento dataset obsahoval 18S rRNA sekvence z databází GenBank (NCBI) a iMicrobe a environmentální 18S rRNA sekvence z expedice TARA (databáze W5, de Vargas et al., 2015), které jsem v databázi vyhledávala podle taxonomického přiřazení. Jako outgroup jsem zvolila několik zástupců sesterské skupiny Mamiellophyceae.

## Fylogenomická analýza

Multiproteinová analýza zahrnovala proteinové sekvence 71 plastidových genů (accD, atpA, B, E, F, H, I, ccsA, cemA, chlB, I, L, N, clpP, ftsH, infA, petA, B, D, G, L, psaA, B, C, I, J, M, psb A, B, C, D, E, F, H, I, J, K, L, M, N, T, Z, rbcL, rpl2, 5, 14, 16, 20, 23, 32, 36, rpoA, B, C1, C2, rps2, 3, 4, 7, 8, 9, 11, 12, 14, 18, 19, tufA, ycf1, 3, 4, 12) ze 47 taxonů linie zelených řas a rostlin podle Lemieux et al. (2014). K nim jsem navíc přidala proteinové sekvence daných plastidových genů z anotovaného genomu *Pterosperma cristatum* a z publikovaných cpDNA Euglenophyceae (*Colacium vesiculosum*, *Cryptoglana skujai*, *Euglena gracilis*, *Euglena viridis*, *Euglenaformis proxima*, *Euglenaria anabaena*, *Eutreitia viridis*, *Eutreptiella gymnastica*, *Monomorphina aenigmatica*, *Monomorphina parapyrum*, *Strombomonas acuminata*, *Trachelomonas volvocina*).

### 4.5.2. Tvorba alignmentů

Všechny připravené datové soubory byly alignovány programem MAFFT (Katoh et al., 2002). Výsledné alignmenty jsem manuálně zkontrolovala a případně poopravila v programu BioEdit (Hall, 1999) tak, aby obsahovaly jen věrohodně alignované oblasti. Alignmenty s nízkým počtem gapů jsem manuálně prořezala v programu BioEdit (Hall, 1999). Na prořezání alignmentů s vysokým obsahem gapů jsem použila software BMGE-1.12 (Criscuolo a Gribaldo, 2010) pracující v příkazovém řádku. Parametr frekvence gapů, nad kterou byly pozice odstraněny, jsem optimalizovala v rozmezí 50-70 % tak, aby zůstaly alignmenty co nejdelší a zároveň neobsahovaly příliš velké množství gapů. Někdy bylo nutné výsledné upravené alignmenty ještě manuálně doopravit, především zarovnat začátky a konce sekvencí. Speciální přístup vyžadovaly alignmenty s TARA sekvencemi a alignment pro fylogenomickou analýzu popsané níže.

### Alignmenty s TARA sekvencemi

Částečné 16S a 18S rRNA sekvence pocházející z expedice TARA jsou velmi krátké, některé se nepřekrývaly a způsobovaly při rekonstrukci fylogeneze problémy. Rozhodla jsem se je proto analyzovat každou zvlášť. Pro všechny TARA sekvence byly automaticky pomocí skriptu vytvořeny samostatné alignmenty spojením alignmentu příslušného datasetu (16S či 18S rRNA) a jedné TARA sekvence. Vzniklo tak celkem 76 alignmentů 16S rRNA a 79 alignmentů 18S rRNA sekvencí.

## **Fylogenomická analýza**

Při tvorbě alignmentu pro multiproteinovou analýzu jsem vycházela z alignmentu použitého ve fylogenetické analýze publikované v článku Lemieux et al. (2014), ke kterému jsem přidala alignment vzniklý konkatenací prořezaných a upravených alignmentů jednotlivých plastidových proteinů Euglenophyceae a *Pterosperma cristatum*. Konkatenace byla provedena nástrojem SequenceMatrix (Vaidya et al., 2011). Takto připravený soubor jsem nechala znovu zaligovat.

### **4.5.3. Tvorba fylogenetických stromů**

Fylogenetické analýzy byly prováděny programem RAxML (Stamatakis, 2014) na internetovém portálu CIPRES ([http://www.phylo.org/sub\\_sections/portal/](http://www.phylo.org/sub_sections/portal/)). Jedná se o nástroj, který statistickou metodou maximální věrohodnosti (maximum likelihood) vyhodnotí nejlepší fylogenetický strom. Pro DNA alignmenty jsem použila model GTR+ $\Gamma$  se čtyřmi rychlostními kategoriemi. V případě fylogenomické analýzy jsem použila model cpREV+ $\Gamma$  se čtyřmi rychlostními kategoriemi s tím, že parametry modelu byly rozvolněny pro jednotlivé části alignmentu („partitions“). Vždy jsem nejdříve nechala spočítat stromy s rychlým bootstrappingem, které sloužily hlavně k prvnímu posouzení fylogeneze a především ke kontrole vstupních dat. Finální alignmenty jsem počítala s deseti běhy RAxML z náhodného počátečního stromu a regulérním bootstrappingem. Výsledné fylogenetické stromy jsem převedla do formátu nexus a poté jsem je graficky upravila v programech FigTree v1.4.2 (Rambaut, 2014) a Inkscape. Speciální přístup vyžadovaly alignmenty s TARA sekvencemi a alignment pro fylogenomickou analýzu popsané níže.

## **16S rRNA**

Při rekonstruování 16S rRNA stromu (bez TARA sekvencí) jsem postupovala hodně obezřetně. Na počátku tento dataset obsahoval velké množství sekvencí z odlišných organismů, z nichž byla velká část popsána jako nekultivovatelné bakterie. Fylogenetický strom byl hodně nepřehledný a velký. Postupně jsem proto v sedmi kolech fylogenetické analýzy vyřadila celé skupiny sekvencí, které byly od skupiny Pyramimonadales a Euglenophyceae odděleny s bootstrapovou podporou přesahující 60 %. Takto vytvořený dataset byl použit pro konečnou analýzu i pro dílčí analýzy s jednotlivými TARA sekvencemi.

## **Stromy s TARA sekvencemi**

S výpočtem fylogenetických stromů obsahujících TARA sekvence mi pomohl Mgr. Petr Soukal. Výpočet jednotlivých stromů byl proveden automaticky pomocí skriptu metodou RAxML (verze RAxML-8.1.17) na výpočetním klastru MetaCentrum (<https://metavo.metacentrum.cz/>) ve frontě backfill. Pozice jednotlivých TARA sekvencí v jejich stromech jsem graficky vyznačila přímo do základních fylogenetických stromů 16S rRNA a prasinofytního 18S rRNA vytvořených bez TARA

sekvencí. Všechny stromy s TARA sekvencemi jsou dostupné na následující webové stránce: <http://www.protistologie.cz/hampllab/data.html>.

#### 4.5.4. Molekulární datování

K odhadu stáří společného předka Euglenophyceae a odvětvení Euglenophyceae od Pyramimonadales byla použita analýza molekulárních hodin. Molekulární datování bylo prováděno Bayéskou metodou s použitím alignmentu plastidových genů pro fylogenomickou analýzu (viz kapitola 4.5.2.) a relaxovaných hodin s lognormálním rozložením substituční rychlosti. Běh Markov Chain Monte Carlo (MCMC) jsem začala z počátečního stromu vytvořeného fylogenomickou analýzou v RAxML (Obr. 9) a analýzu jsem omezila (constrain) tak, aby byla skupina Chlorophyta monofyletická. Pro alignment jsem nastavila hranice jednotlivých genů („partitions“), pro které byl zvláště volen substituční model. To umožňuje, aby byla rychlost běhu molekulárních hodin pro každý gen jiná. Substituční model byl pro jednotlivé genové části alignmentu „partitions“ nastaven cpREV+ $\Gamma$  se čtyřmi rychlostními kategoriemi. Pro kalibraci stromu jsem použila několik kalibračních bodů s normálním rozložením jejich hodnot. Průměr a standardní odchylka byly nastaveny tak, aby rozložení hodnot na křivce odpovídalo odhadu stáří jednotlivých kalibračních bodů uvedených v literatuře (Tab. 10).

Dělala jsem současně dvě různé analýzy. Jedna obsahovala všechny kalibrační body. Ve druhé jsem vynechala kalibrační bod UTC (Tab. 10). Obávali jsme se, že tento kalibrační bod může kvůli jeho značnému stáří (1550-950 mil. let) způsobovat v některých případech neočekávanou finální topologii stromu, ve které se Streptophyta odvětvovala uvnitř linie Chlorophyta (vyřazená analýza). Navíc jsme se obávali, že interpretace takto starých fosilií nemusí být správná. Analýzy byly prováděny nástrojem BEAST 1.8 (Drummond a Rambaut, 2007) na portálu CIPRES. Markov Chain Monte Carlo (MCMC) běžel 168 hodin, což je maximální povolená doba. Za tuto dobu prošel MCMC 4 105 000 generací v analýze s UTC a 9 569 000 generací v analýze bez UTC. Ukládala jsem stromy z každé tisícáté generace. Burn-in (počet počátečních stromů, které byly z analýzy vyřazeny) byl stanoven pomocí programu Tracer v1.6 (Rambaut et al., 2014) pro analýzu s UTC na 3600, pro analýzu bez UTC na 4000 a 9000. V posledním případě jsem zvolila dva různé burn-in kvůli malému skoku MCMC k vyššímu likelihoodu na konci procesu (9 000 000 generací), který ale zahrnoval jen poslední 569 000 generací. Výsledné stromy byly sumarizovány a graficky upraveny v programech FigTree v1.4.2 (Rambaut, 2014) a Inkscape.

**Tab. 10. Kalibrační body pro molekulární datování.** Stdev značí standardní odchylku. Mil. je zkratka pro miliony let.

Kalibrační bod	Nejstarší nález	Odhad stáří	Průměr	Stdev	Reference
UTC	<i>Vandalosphaeridium</i>	1550-950 mil.	1250	180	Moczydłowska et al. (2011)
Ulvophyceae	<i>Archaeoclada</i> , <i>Variacлада</i>	950-700 mil.	825	75	Moczydłowska et al. (2011)
Chlorophyceae	<i>Appendisphaera</i>	600-560 mil.	580	12	Moczydłowska et al. (2011)
Pyramimonadales	<i>Pterospermella</i> , <i>Cymatiospharea</i>	413-405 mil.	405,85	2,4	Kustatscher et al. (2014)
Embryophyta	výtrusy	473-471 mil.	472	0,6	Parfrey et al. (2011)
cévnaté r.	nejstarší cévnaté r.	432-418 mil.	425	4,5	Parfrey et al. (2011)
krytosemenné r.	pyl krytosemenných	137-130 mil.	133,5	2,1	Parfrey et al. (2011)
dvouděložné r.	pyl dvouděložných	127-125 mil.	126	0,6	Parfrey et al. (2011)



## 5. Výsledky

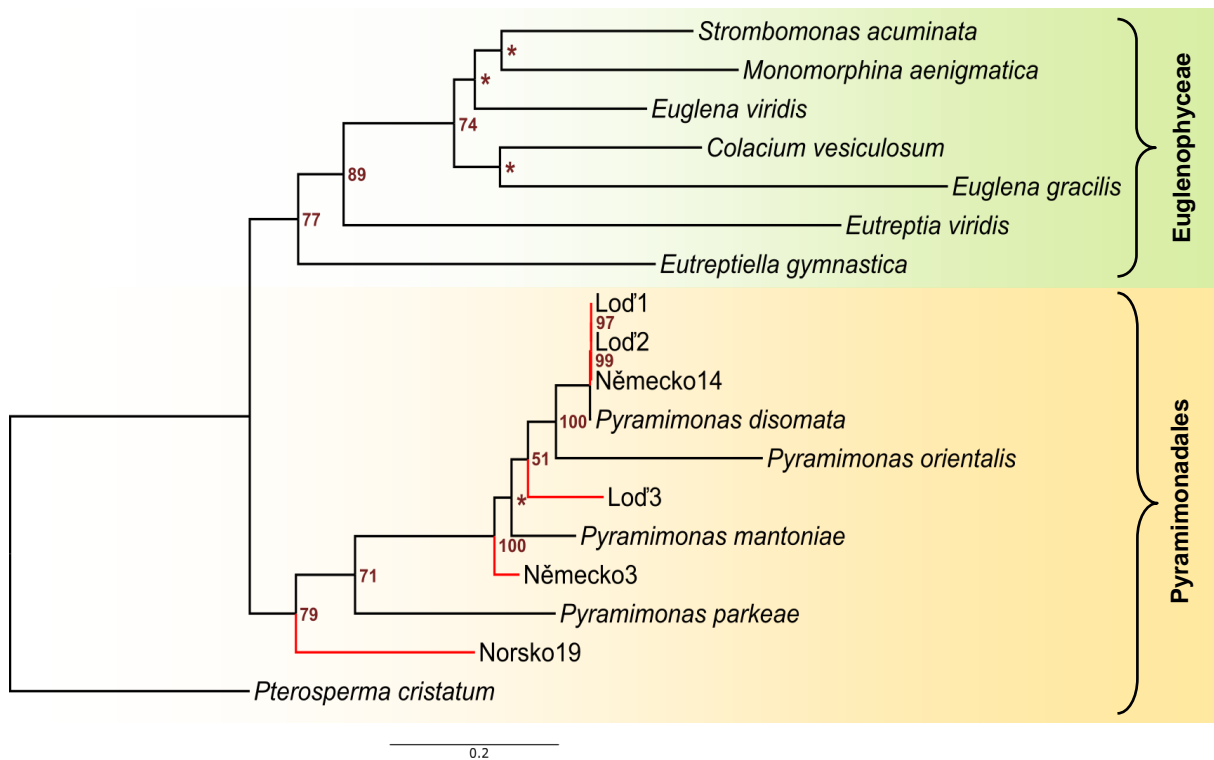
### 5.1. Analýza genových klastrů

Za pomoci specifických primerů jsem amplifikovala genové klastry z DNA získaných z environmentálních vzorků a z kultur *Pyramimonas*. Kromě dvou hlavních genových oblastí rps4-rps11 a trnC(gca)-rps2 jsem se snažila analyzovat také klastry ohrazené geny rps14-trnF(gaa) a trnN(guu)-trnR(acg) (Obr. 1). Získaná data byla však nedostatečná pro tvorbu dobře rozlišených fylogenetických stromů. V prvním případě způsobovaly problémy geny chybějící u rodu *Euglena* i ostatních Euglenophyceae. Druhý genový klaster byl příliš krátký. V úvahu přicházela rovněž oblast s geny psbK-ycf12-psaM (Obr. 1). Tento genový klaster však není sdílený pouze rody *Pyramimonas* a *Euglena*, ale vyskytuje se také u jiných chlorofytních řas (Turmel et al., 2009). Pro další zkoumání jsem proto použila pouze výsledky z dlouhých genových klastrů rps4-rps11 a trnC(gca)-rps2.

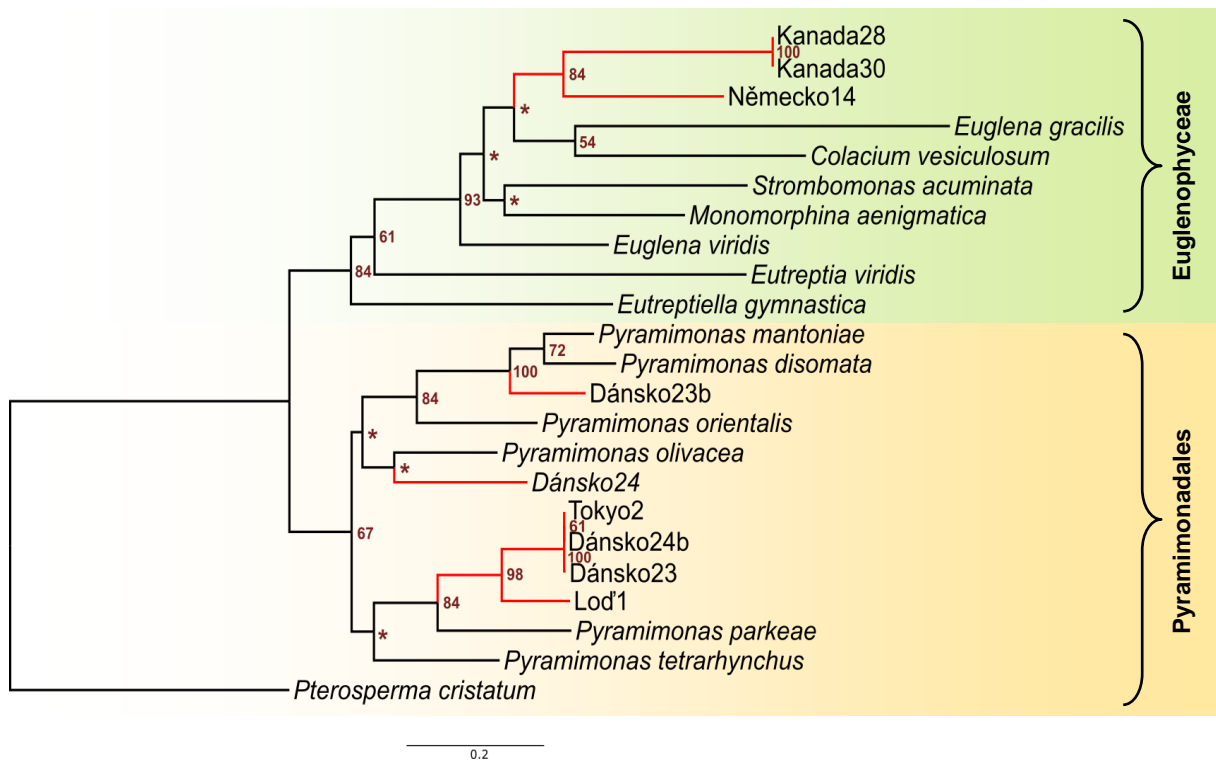
Amplifikace genových klastrů z kultur byla až na dvě výjimky úspěšná. Nepodařilo se mi amplifikovat pouze sekvence klastru rps4-rps11 z kultury *P. olivacea* a *P. tetraarhynchus*. Sekvence alespoň jednoho klastru se mi podařilo získat z 11 environmentálních DNA (Tab. 11). Fylogenetická analýza odhalila řadu environmentálních linií větvcích se buď uvnitř skupiny Pyramimonadales, nebo Euglenophyceae. Fylogenetický strom oblasti rps4-rps11 (Obr. 2) zahrnuje celkem šest nových environmentálních sekvencí spadajících mezi Pyramimonadales. Oproti tomu analýza genového klastru trnC(gca)-rps2 (Obr. 3) odhalila kromě šesti prasinofytních také tři euglenofytní linie. Z některých vzorků se mi nepodařilo amplifikovat žádnou sekvenci, jiné byly naopak zdrojem více různých sekvencí (Dánsko23, 24) či jsem z této DNA získala sekvence obou klastrů (Lod'1, Německo14). Sekvence klastru rps4-rps11 získaná z environmentálního vzorku Kanada25 byla z alignmentu vyloučena, neboť se jednalo o bakteriální kontaminaci zjištěnou nástrojem BLASTn.

**Tab. 11. Amplifikované genové klastry z environmentálních vzorků.**

Vzorek	Genový klaster	
	rps4-rps11	trnC(gca)-rps2
Lod'1		
Lod'2		
Lod'3		
Německo3		
Německo14		
Norsko19		
Dánsko23		
Dánsko24		
Kanada28		
Kanada30		
Tokyo2		



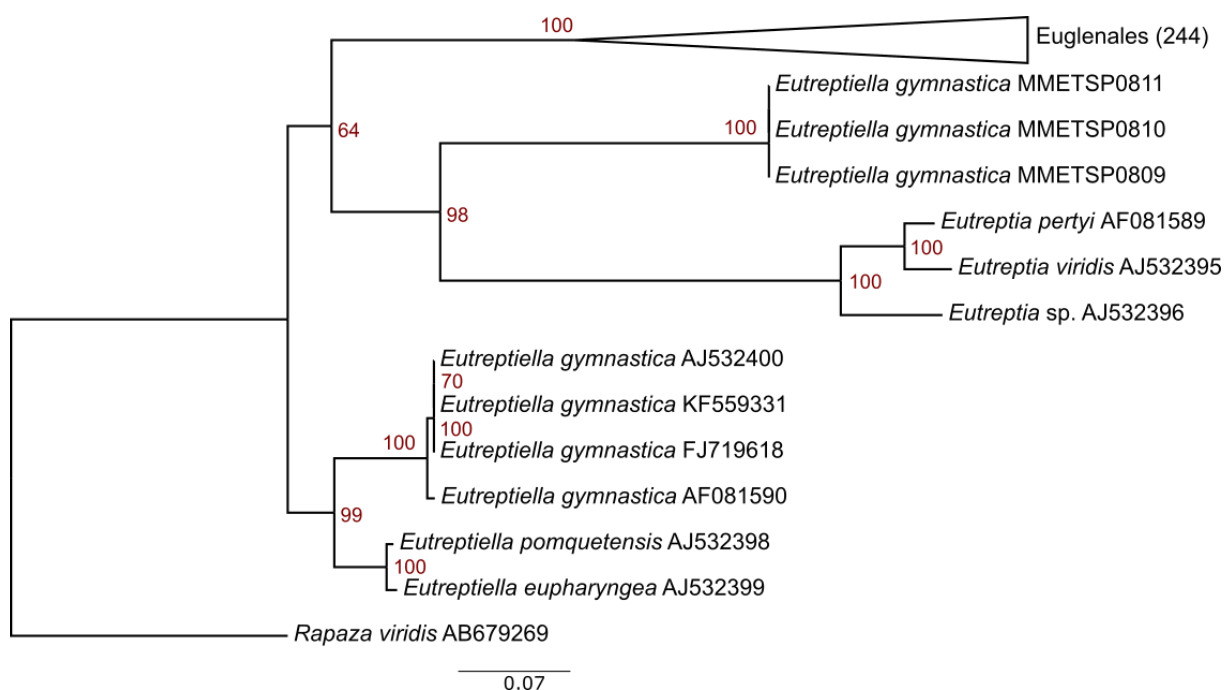
**Obr. 2. Fylogenetický strom genového klastru rps4-rps11.** Topologie byla vytvořena metodou maximum likelihood (ML) v RAxML. Environmentální sekvence jsou zvýrazněny červeně. Číselné hodnoty na uzlech vyjadřují bootstrapovou podporu větvení získanou metodou ML v RAxML s 500 opakováními. \* označuje uzle s hodnotou méně než 50 %.



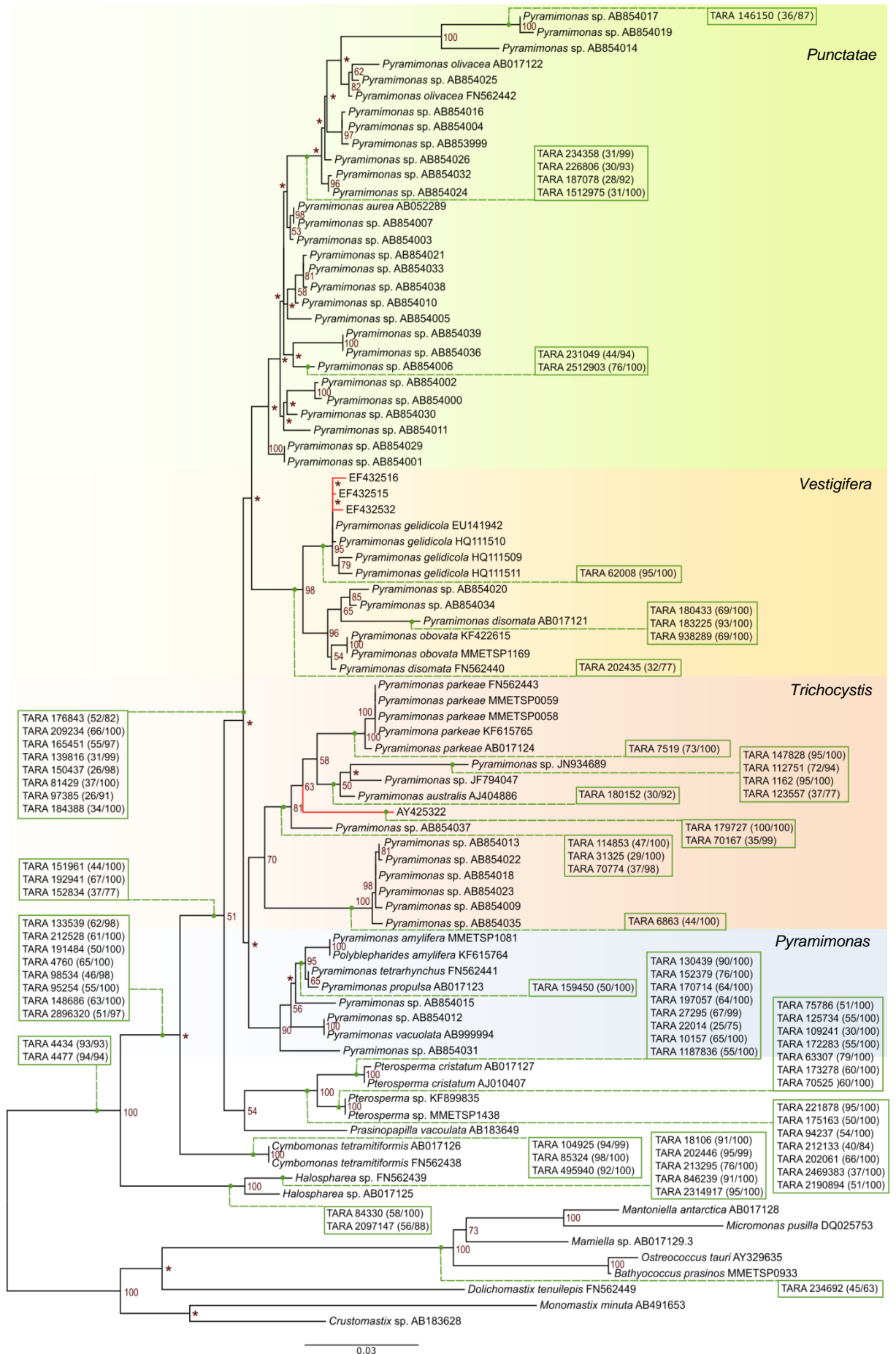
**Obr. 3. Fylogenetický strom genového klastru trnC(gca)-rps2.** Topologie byla vytvořena metodou maximum likelihood (ML) v RAxML. Environmentální sekvence jsou zvýrazněny červeně. Číselné hodnoty na uzlech vyjadřují bootstrapovou podporu větvení získanou metodou ML v RAxML s 500 opakováními. \* označuje uzle s podporou méně než 50 %.

## 5.2. Fylogeneze na základě 18S rRNA

Na základě sekvencí 18S rRNA stažených z databází GenBank, iMicrobe a expedice TARA (de Vargas et al., 2015) jsem zhotovila dva fylogenetické stromy metodou maximum likelihood (ML). První strom (Obr. 4) byl zaměřený na fylogenezi v rámci Euglenophyceae. Sladkovodní Euglenales vytváří monofyletický klad s maximální bootstrapovou podporou. Na bázi stromu se parafyleticky větví zástupci Eutreptiales. Jednotlivé rody nicméně nejsou monofyletické, neboť 18S rRNA sekvence pocházející z různých izolátů organismů označených jako *Eutreptiella gymnastica* tvoří dva oddělené klady. Druhý ze stromů (Obr. 5) ukazuje diverzitu v rámci skupiny Pyramimonadales. Pyramimonadales představují monofyletický klad sesterský linii Mamiellophyceae zvolené jako outgroup. Rod *Pyramimonas* vytváří korunní diverzifikovanou skupinu, na jejíž bázi se parafyleticky odvětvují ostatní klady *Pterosperma-Prasinopapilla*, *Cymbomonas* a *Halosphaera*. Do tohoto stromu byly navíc naznačeny pozice TARA sekvencí získané dílčími analýzami (viz kapitoly 4.5.2. a 4.5.3.).



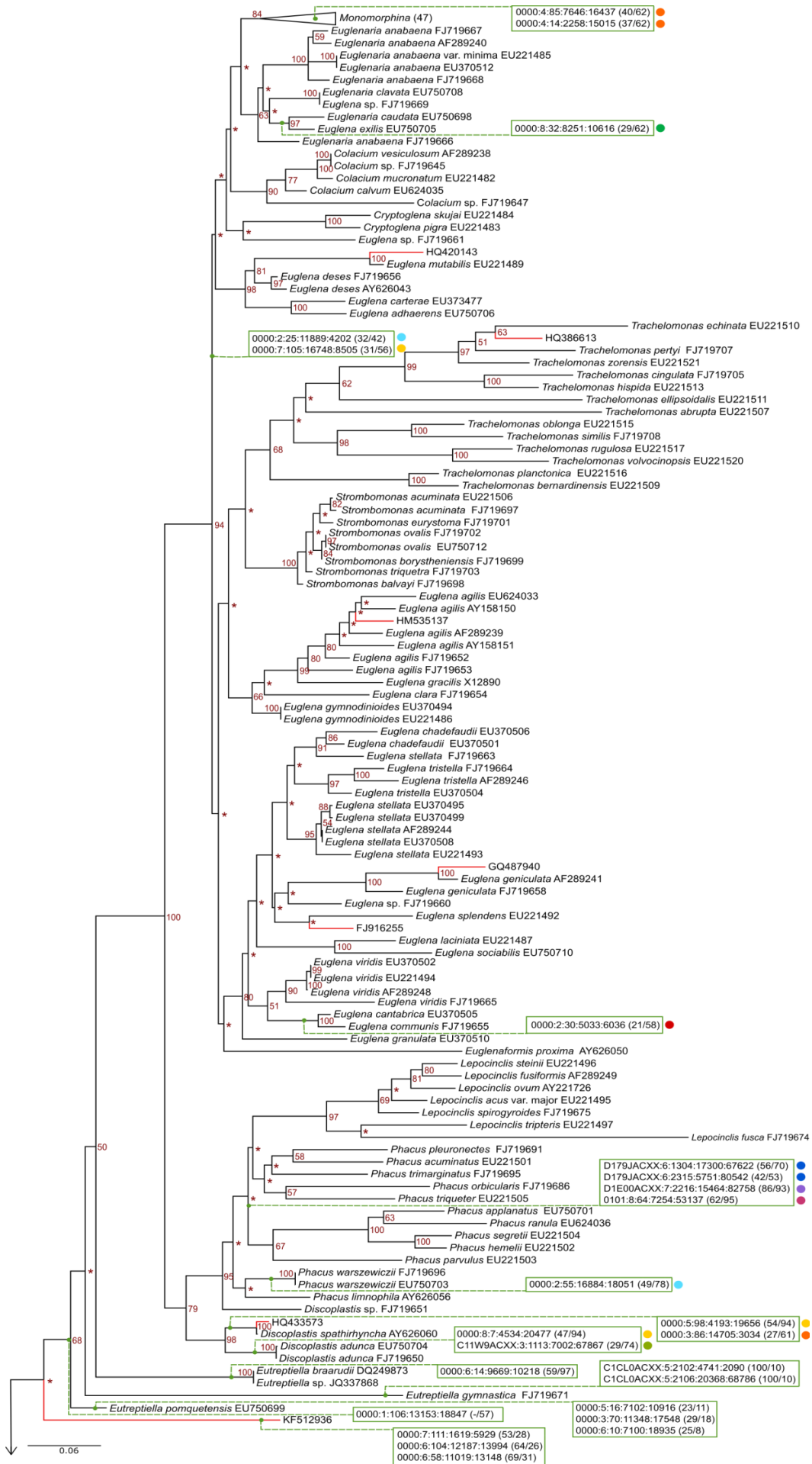
**Obr. 4. Fylogenetický strom na základě 18S rRNA sekvencí Euglenophyceae.** Topologie byla vytvořena metodou maximum likelihood v programu RAxML. Kódy jednotlivých sekvencí odkazují na záznam v příslušných databázích GenBank a iMicrobe (MMETSP..). Hodnota v závorce u skupiny Euglenales udává počet taxonů, které byly zkolabovány. Podpora uzlů je vyjádřena bootstrapem zjištěným metodou maximum likelihood s 500 opakováními.

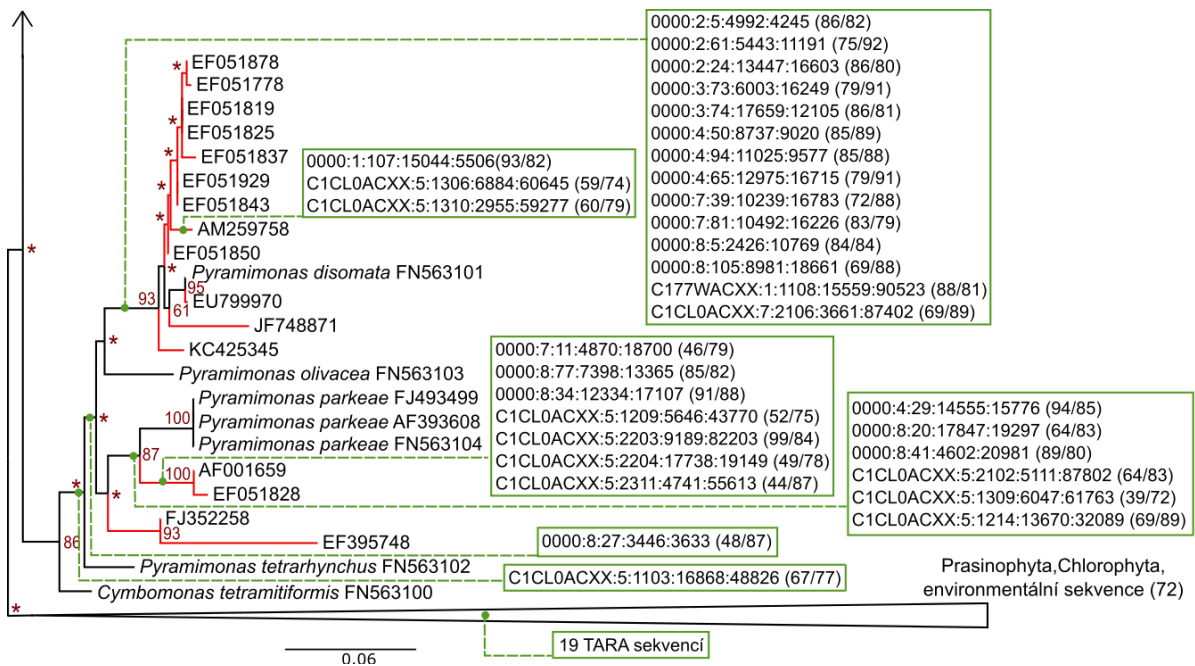


**Obr. 5. Fylogenetický strom na základě 18S rRNA sekvencí skupiny Pyramimonadales s vyznačenými pozicemi TARA sekvencí (předešlá strana).** Topologie byla vytvořena metodou maximum likelihood (ML) v programu RAxML. Značení jednotlivých sekvencí je převzato z příslušných databází GenBank a iMicrobe (MMETSP.). Environmentální sekvence z NCBI jsou zvýrazněny červeně. Podskupiny rodu *Pyramimonas* jsou barevně odděleny. Podporu uzlů vyjadřuje bootstrap zjištěným metodou ML s 500 opakováními. \* označuje uzly s podporou nižší než 50 %. Do fylogenetického stromu byly vyneseny pozice TARA sekvencí (zelené rámečky) určené na základě dílčích analýz. Číselné kódy TARA sekvencí udávají identifikační čísla OTU dostupné v databázi W5 (de Vargas et al., 2015). Bootstrap zjištěný ML metodou je vyjádřen v závorce: podpora daného uzlu, kde se větví TARA sekvence/bootstrapová podpora vyšší taxonomické skupiny, do které příslušná TARA sekvence spadá (Pyramimonadales či Mamiellophyceae).

### 5.3. Fylogeneze na základě 16S rRNA

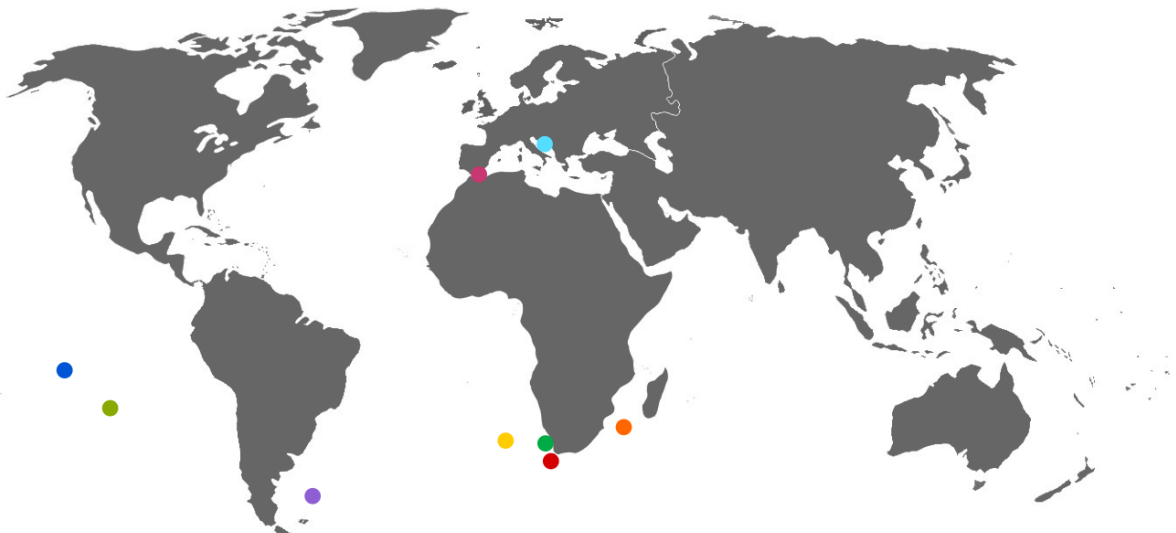
Metodou maximum likelihood jsem provedla fylogenetickou analýzu genového markeru 16S rRNA, která zahrnovala dostupné sekvence environmentálních vzorků, zástupců zelené linie řas a Euglenophyceae stažené z databází GenBank a expedice TARA. Ve výsledném stromu (Obr. 6) vytváří Euglenophyceae monofyletický klad sesterský skupině Pyramimonadales. Pyramimonadales nicméně nepředstavuje monofyletickou linii, protože se *Pterosperma* větví uvnitř zkolabovaného klanu zahrnujícího ostatní Chlorophyta, Prasinophyta a environmentální sekvence. Mořské environmentální TARA sekvence se umístily především uvnitř Pyramimonadales, bazálních Eutreptiales a zkolabované linie (Obr. 6). Značnou environmentální diverzitu můžeme pozorovat zejména v kladu zahrnující organismus *Pyramimonas disomata*, do kterého spadá celkem 29 environmentálních sekvencí s dobrou bootstrapovou podporou. Zajímavé jsou také TARA sekvence větvcí se spolu se sladkovodními taxony Euglenales, jež pocházejí z různých světových moří a oceánů (Mapa E).





**Obr. 6. Fylogenetický strom zkonstruovaný na základě 16S rRNA sekvencí Euglenophyceae a Chlorophyta s vyznačenými pozicemi TARA sekvencí (pokračování na předešlé straně).** Topologie byla vytvořena metodou maximum likelihood (ML) v programu RAxML. Celkem 72 sekvencí ostatních zástupců Prasinophyta, Chlorophyta a environmentálních vzorků bylo zkolabováno do jednoho celku pro lepší přehlednost. Sekvence pocházející z databáze GenBank jsou označeny příslušnými kódy. Environmentální sekvence z NCBI jsou zvýrazněny červeně. Hodnoty jednotlivých uzlů vyjadřují bootstrapovou podporu větvení zjištěnou metodou ML s 500 opakováními. \* označuje hodnotu nižší než 50 %. Do fylogenetického stromu byly vyneseny pozice TARA sekvencí (zelené rámečky) určené na základě dílčích analýz. TARA sekvence jsou označeny kódy, dle kterých je lze vyhledat v příslušné 16S rRNA databázi (Sunagawa et al., 2015). Hodnoty v závorkách určují bootstrapovou podporu větvení TARA sekvencí: hodnota u daného uzle/hodnota pro vyšší taxonomickou skupinu (*Pyramimonadales*, *Euglenales*, *Euglenophyceae* či *KF512936* + *Euglenophyceae*). Barevná kolečka u vybraných sekvencí odkazují na odběrová místa vzorků (viz mapa E).

**Mapa E. Odběrová místa vzorků vybraných 16S rRNA TARA sekvencí.** Barevná kolečka odpovídají TARA sekvencím vyznačeným v obrázku 6.



#### 5.4. Plastidový genom *Pterosperma cristatum*

Součástí mé diplomové práce byla rovněž sekvenace a anotace plastidového genomu řasy *Pterosperma cristatum*. Osekvenovaný genom (Illumina MiSeq) se podařilo složit do dvou scaffoldů o délkách 83 149 bp (párů bází) a 46 429 bp. První scaffold byl poskládán ze 454 950 čtení, druhý scaffold vznikl složením 142 307 čtení. Celková délka cpDNA *P. cristatum* činí 129 578 bp a GC obsah je 38,81 %. Ačkoliv se mi podařilo amplifikovat chybějící část mezi scaffoldy, po vyříznutí z gelu a přečištění kitem QIAquick® Gel Extraction Kit (Qiagen) byla koncentrace DNA velmi nízká dosahující maximálně 6,1 ng/μl. Veškerá provedená sekvenace z tohoto důvodu selhala a plastidový genom *Pterosperma cristatum* se nepodařilo spojit do jednoho scaffoldu. Některé oblasti genomu mohou obsahovat chyby a pokrytí čteními je v těchto oblastech slabé (viz níže).

V plastidovém genomu *Pterosperma cristatum* jsem anotovala celkem 92 genů. Většina těchto genů souvisí s fotosyntetickou funkcí. Dále se zde nachází velké množství genů kódujících ribozomální proteiny, genů spojených s transkripcí a translací, rRNA a tRNA. Všechny nalezené plastidové geny a jejich funkce jsou shrnuty v tabulce 12. Tabulka 13 shrnuje frekvenci používání jednotlivých kodonů. Rozložení jednotlivých genů, jejich délku a směr přepisu ukazují obrázky 7 a 8. Některé geny nicméně zůstaly ne zcela dořešené. Uvnitř genu chlB se nachází oblast s nízkým pokrytím čteními, která zřejmě způsobuje posunutí čtecího rámce na začátku sekvence genu. Gen chlN jsem označila za částečný (Obr. 8), neboť chybí velká část genu, která je u ostatních prasinofyt přítomna. Dalšími problémovými geny označenými v obrázcích 7 a 8 jako hypotetické jsou chlI, ftsH, rpoA a rpoC2. Potíž spočívá v jejich příliš dlouhém 3' konci, který v případě rpoC2 genu zapříčiňuje celkovou délku sekvence 9930 bp (Obr. 8). V genomu jsem objevila dva introny uvnitř genů psbD a ycf3 (Obr. 7 a 8). Zatímco oblast intronu genu ycf3 se zdá být nepřekládaná, v intronu genu psbD leží otevřený čtecí rámec (ORF) pro intronovou reverzní transkriptázu/maturázu. Veškeré geny vyjma jednoho začínají klasickým start kodonem ATG kódujícím methionin. Gen rpl5 má zřejmě jako alternativní start kodon valin (GTG).

Ve srovnání s blízkou příbuznou řasou *Pyramimonas parkeae* postrádá *Pterosperma cristatum* v plastidovém genomu 19 genů. Protože genomová sekvence není kompletní, je možné, že jsou některé tyto geny v genomu přítomny. Tabulka 14 ukazuje rozdíly v genomovém obsahu mezi prasinofytními a euglenofytními řasami.

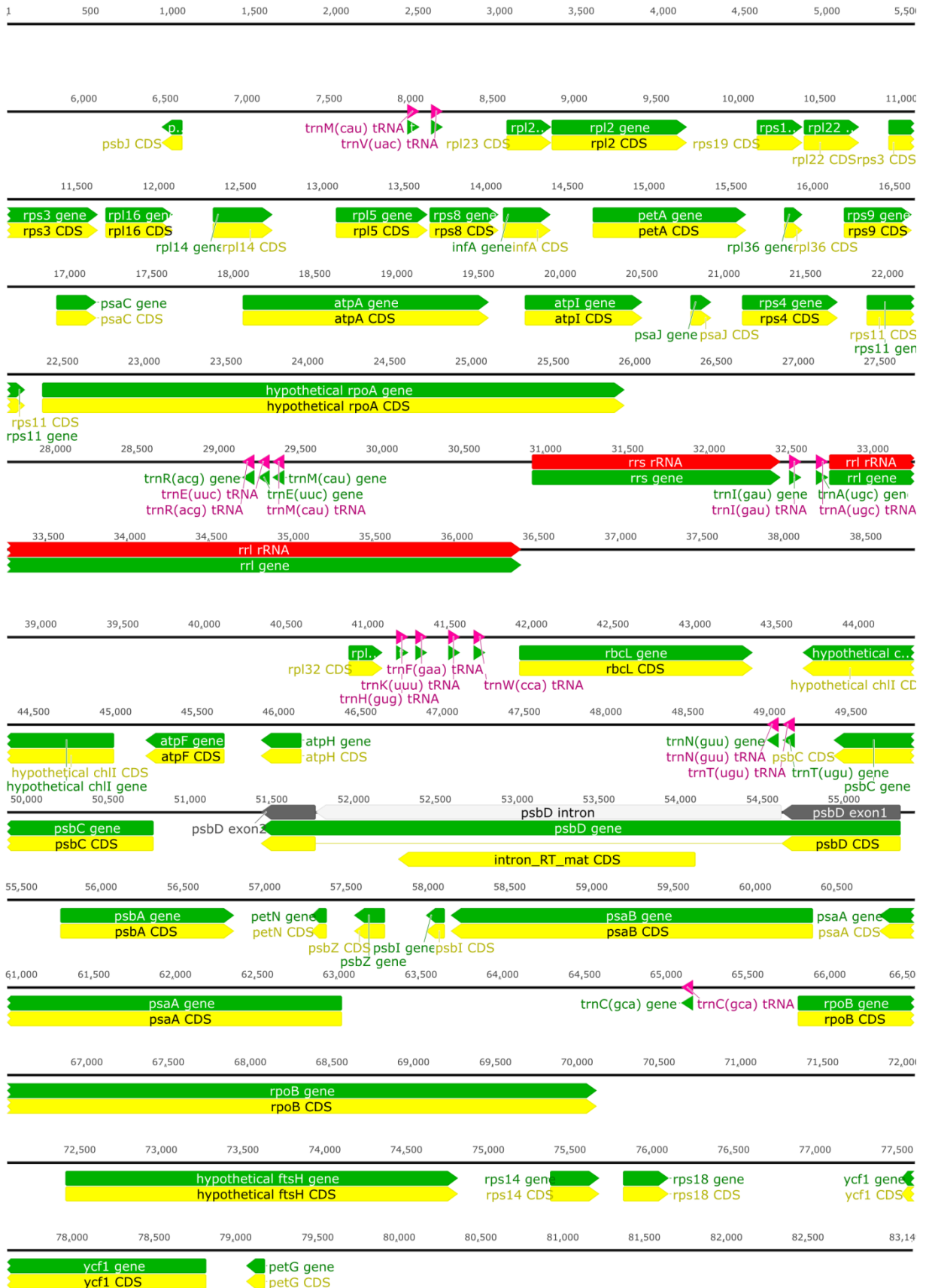


**Tab 12. Anotované geny *Pterosperma cristatum* a jejich funkce.** Legenda: ? - funkce zatím ne zcela objasněna, \* - gen pravděpodobně není správně předpovězen, PS - fotosystém.

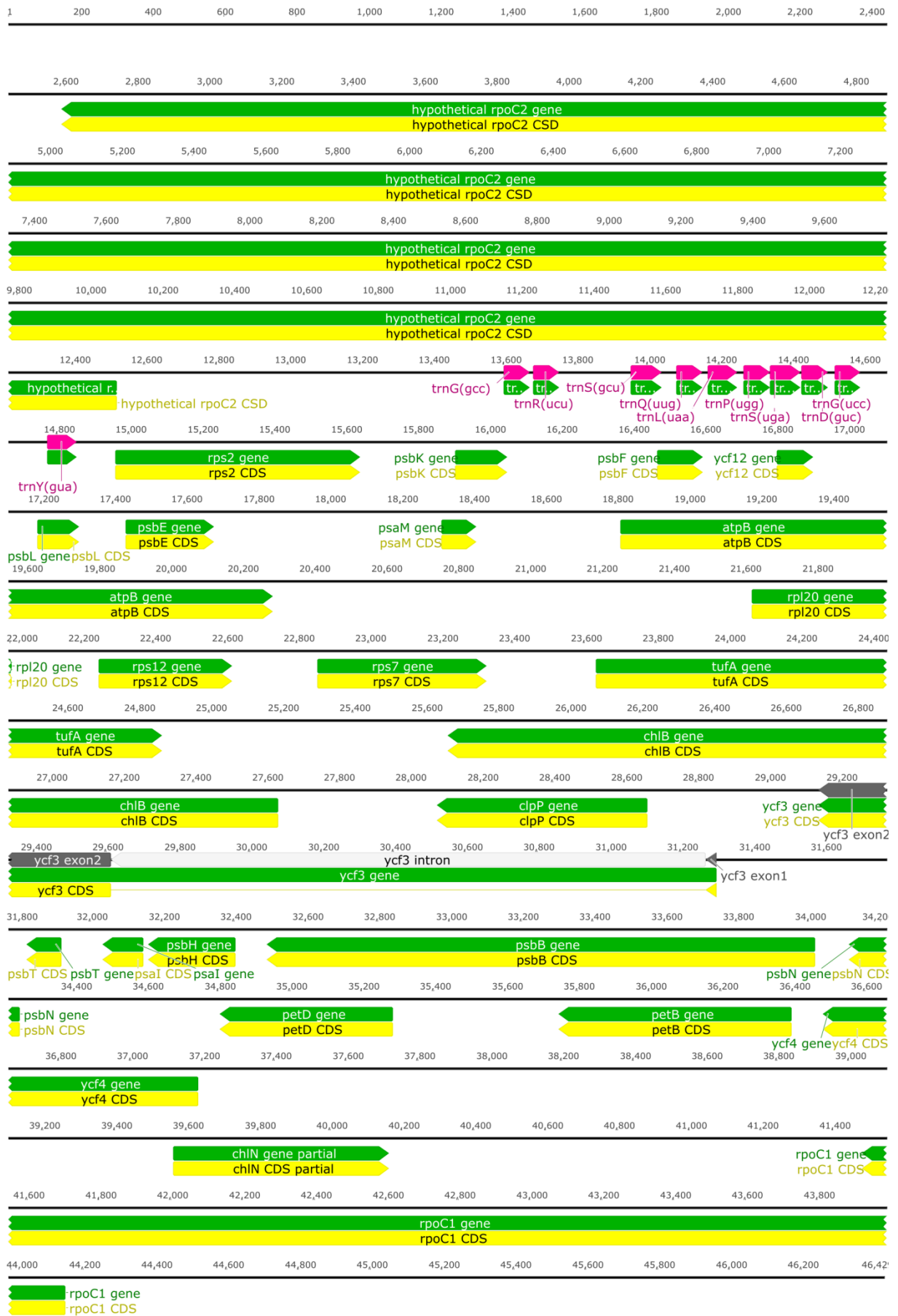
Označení genu	Genový produkt	Funkce
atpA, B, F, H, I	podjednotky ATP syntázy	zajištění fotosyntézy
chlB*, I*, N*	podjednotky protochlorofylid reduktázy	
petA, B, D, G, N	podjednotky cytochrom b6/f komplexu	
psaA, B, C, I, J, M	podjednotky PSI	
psbA, B, C, D, E, F, H, I, J, K, L, N, T, Z	podjednotky PSII	
rbcL	velká podjednotka RuBisCO	
ycf3, 4	proteiny zodpovídající za složení PSI	
ycf12	součást reakčního centra PSII	
clpP	podjednotka ATP-dependentní proteázy	proteolýza
ftsH*	ATP-dependentní metaloproteáza	
infA	translační iniciační faktor	translace
tufA	translační elongační faktor	
rpoA*, B, C1, C2*	podjednotky RNA polymerázy	transkripce
rpl2, 5, 14, 16, 20, 22, 23, 32, 36	ribosomální proteiny velké jednotky	ribozóm
rps2, 3, 4, 7, 8, 9, 11, 12, 14, 18, 19	ribosomální proteiny malé jednotky	
rrl, rrs	rRNA	
trn A(ugc), C(gca), D(guc), E(uuc), F(gaa), G(gcc), G(uch), H(gug), I(gau), K(uuu), L(uaa), Me(cau), Mf(cau), N(guu), P(ugg), Q(uug), R(acg), R(ucu), S(gcu), S(uga), T(ugu), V(uac), W(cca), Y(gua)	tRNA	proteosyntéza
ycf1	součást membránového komplexu	transport?

**Tab. 13. Frekvence používání kodonů v cpDNA *P. cristatum*.** N/1000 = počet vztažený na 1000 AK.

Kodon	N/1000	Kodon	N/1000	Kodon	N/1000	Kodon	N/1000
Ala (GCG)	3,51	His (CAT)	15,47	Pro (CCG)	4,18	Ser (TCT)	13,90
Ala (GCA)	17,93	His (CAC)	12,85	Pro (CCA)	13,60	Ser (TCC)	4,63
Ala (GCT)	20,85	Ile (ATA)	17,86	Pro (CCT)	12,03	Thr (ACG)	5,16
Ala (GCC)	4,18	Ile (ATT)	38,63	Pro (CCC)	3,21	Thr (ACA)	15,69
Cys (TGT)	6,80	Ile (ATC)	17,34	Gln (CAG)	15,54	Thr (ACT)	15,09
Cys (TGC)	3,51	Lys (AAG)	27,57	Gln (CAA)	34,08	Thr (ACC)	4,78
Asp (GAT)	19,20	Lys (AAA)	43,27	Arg (AGG)	4,41	Val (GTG)	15,99
Asp (GAC)	10,01	Leu (TTG)	23,76	Arg (AGA)	8,59	Val (GTA)	21,45
Glu (GAG)	11,66	Leu (TTA)	46,93	Arg (CGG)	2,84	Val (GTT)	28,84
Glu (GAA)	28,84	Leu (CTG)	12,18	Arg (CGA)	7,40	Val (GTC)	9,49
Phe (TTT)	37,36	Leu (CTA)	16,81	Arg (CGT)	12,48	Trp (TGG)	16,29
Phe (TTC)	30,04	Leu (CTT)	20,62	Arg (CGC)	2,32	Tyr (TAT)	16,81
Gly (GGG)	5,68	Leu (CTC)	8,67	Ser (AGT)	10,39	Tyr (TAC)	13,60
Gly (GGA)	16,07	Met (ATG)	26,53	Ser (AGC)	5,98	Stop (TGA)	6,80
Gly (GGT)	27,65	Asn (AAT)	27,80	Ser (TCG)	4,18	Stop (TAG)	12,18
Gly (GGC)	4,71	Asn (AAC)	17,49	Ser (TCA)	14,42	Stop (TAA)	21,82



**Obr. 7. Scaffold 1 cpDNA *Pterosperma cristatum*.** Číselné hodnoty ukazují délku scaffoldu. Značení: zelená – anotované geny, žlutá – kódující oblasti genů (CDS), červená – rRNA, růžová – tRNA, tmavě šedá – exony, světle šedá – introny.



Obr. 8. Scaffold 2 cpDNA *Pterosperma cristatum*. Číselné hodnoty ukazují délku scaffoldu. Značení: zelená – anotované geny, žlutá – kódující oblasti genů (CDS), růžová – tRNA, tmavě šedá – exony, světle šedá – introny.

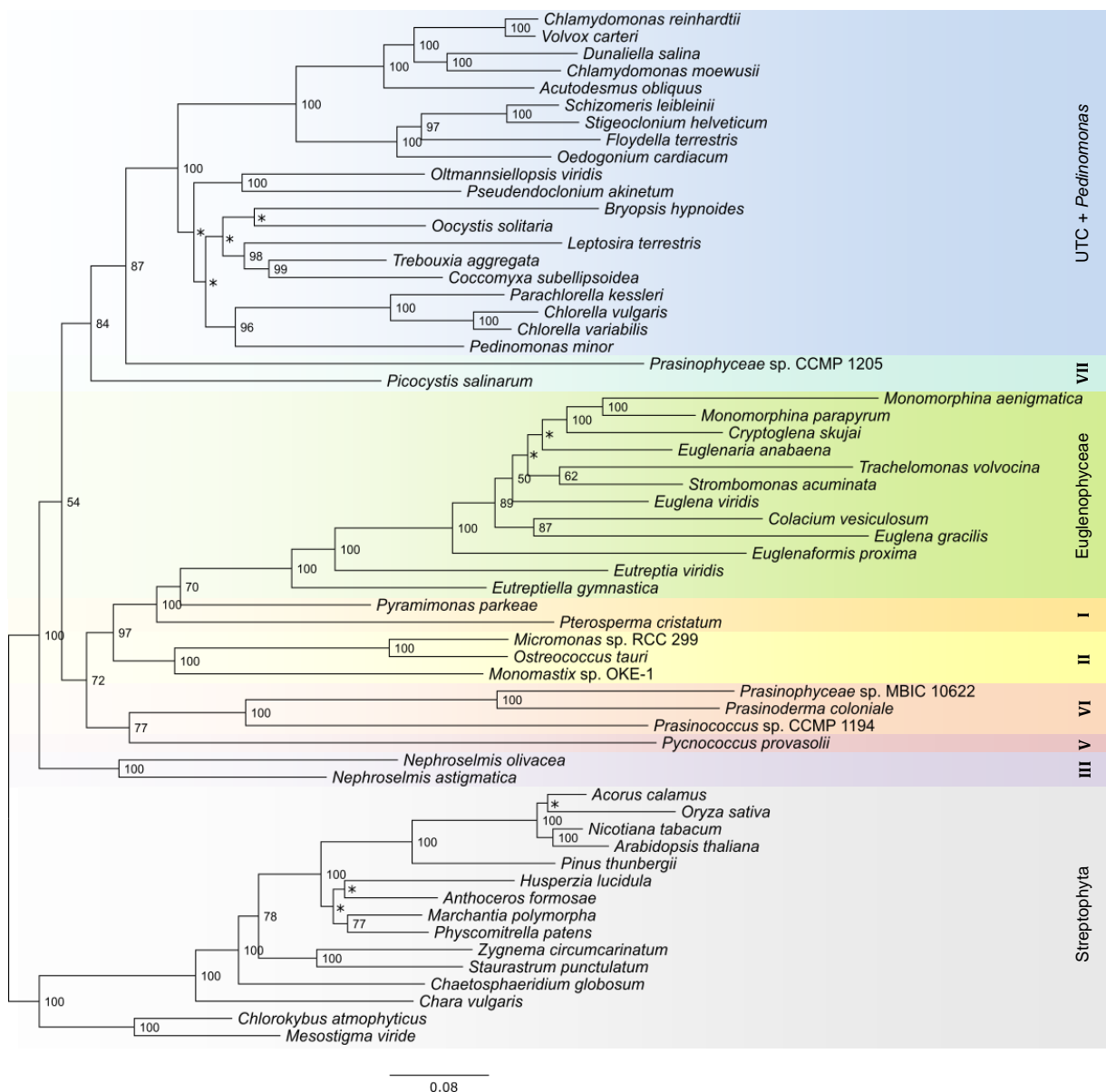
**Tab. 14. Rozdíly v genovém obsahu mezi prasinofytními a euglenofytními řasami.** Tabulka shrnuje pouze plastidové geny, v jejichž obsahu se jednotlivé skupiny liší. Legenda: × nepřítomnost genu, ● přítomnost genu, ● přítomnost částečného nebo hypotetického genu, P pseudogen, prázdné políčko značí možnou přítomnost genu při neúplné cpDNA.

Gen	Klad I		Klad II			Klad III		Klad V	Klad VI		Klad VII	Euglenophyceae		
	<i>Pterosperma cristatum</i>	<i>Pyramimonas parkeae</i> (NC_012099)	<i>Micromonas</i> sp. (FJ858267)	<i>Monomastix</i> sp. (FJ493497)	<i>Ostreococcus tauri</i> (NC_008289)	<i>Nephroselmis astigmatica</i> (NC_024829)	<i>Nephroselmis olivacea</i> (AF137379)	<i>Pycnococcus provasolii</i> (FJ493498)	<i>Prasinococcus</i> sp. (KJ746597)	<i>Prasinoderma coloniale</i> (NC_024817)	<i>Picocystis salinarum</i> (NC_024828)	<i>Eutreptia viridis</i> (JN643723)	<i>Eutreptiella gymnastica</i> (NC_017754)	<i>Euglena gracilis</i> (NC_001603)
accD		×	×	×	×	×	●	×	×	●	×	×	×	×
atpE		●	●	●	●	●	●	●	●	●	●	●	●	●
ccsA		●	×	●	×	●	●	×	×	×	×	×	×	×
cemA		×	×	×	×	×	●	●	×	×	●	×	×	×
chlB	●	●	×	×	×	●	●	×	●	●	×	×	×	×
chlI	●	●	×	●	×	●	●	●	●	●	●	●	●	●
chlL		●	×	×	×	●	●	●	●	●	×	×	×	×
chlN	●	●	×	×	×	●	●	●	●	●	×	×	×	×
clpP	●	●	●	●	●	●	●	●	●	●	●	×	×	×
cysA		×	×	×	×	●	●	×	×	×	×	×	×	×
cysT		×	×	×	×	●	●	●	×	×	×	×	×	×
ftsH	●	●	×	●	×	●	×	●	●	●	●	×	×	×
ftsI		×	×	×	×	×	●	×	×	●	●	×	×	×
ftsW		×	×	×	×	×	●	×	×	●	●	×	×	×
infA	●	●	●	●	●	●	●	●	×	×	●	×	×	×
minD		×	×	×	×	●	●	×	●	●	●	×	×	×
ndhA		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhB		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhC		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhD		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhE		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhF		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhG		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhH		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhI		●	×	×	×	●	●	×	●	×	●	×	×	×
ndhJ		×	×	×	×	×	×	×	●	×	×	×	×	×
ndhK		●	×	×	×	●	●	×	●	×	●	×	×	×
petA	●	●	●	●	●	●	●	●	●	●	●	×	×	×
petD	●	×	×	●	×	●	●	●	●	●	●	×	×	×
petL		×	×	×	×	●	●	●	×	×	●	×	×	×
petN	●	●	×	●	×	●	●	●	●	●	×	×	×	×
psaC	●	●	●	×	●	●	●	●	●	●	●	●	●	●
psaI	●	●	●	●	●	●	●	●	●	●	●	●	×	×
psaJ	●	●	●	●	●	●	●	×	●	●	×	●	●	●
psaM	●	●	●	●	●	×	×	×	●	●	●		●	●

Gen	Klad I		Klad II			Klad III		Klad V	Klad VI		Klad VII	Euglenophyceae		
	<i>Pterosperma cristatum</i>	<i>Pyramimonas parkeae</i> (NC_012099)	<i>Micromonas</i> sp. (FJ858267)	<i>Monomastix</i> sp. (FJ493497)	<i>Ostreococcus tauri</i> (NC_008289)	<i>Nephrolelmis astigmatica</i> (NC_024829)	<i>Nephrolelmis olivacea</i> (AF137379)	<i>Pycnococcus provasolii</i> (FJ493498)	<i>Prasinococcus</i> sp. (KJ746597)	<i>Prasinoderma coloniale</i> (NC_024817)	<i>Picocystis salinarum</i> (NC_024828)	<i>Eutreptia viridis</i> (JN643723)	<i>Eutreptiella gymnastica</i> (NC_017754)	<i>Euglena gracilis</i> (NC_001603)
psbM		×	×	●	×	●	●	×	×	×	●	×	×	×
rbcR		×	×	×	×	×	×	×	●	●	×	×	×	×
rne		×	×	×	×	×	●	×	×	×	×	×	×	×
rnpB		×	×	●	×	×	●	●	×	×	×	×	×	×
rpl12		●	×	×	×	●	●	×	●	×	×	●	●	●
rpl19		×	×	×	×	●	●	●	●	●	×	×	×	×
rpl21		×	×	×	×	×	×	×	●	●	×	×	×	×
rpl22	●	●	×	×	×	×	×	×	●	●	×	●	●	●
rpl32	●	●	●	●	●	●	●	×	●	×	●	●	●	●
rpl36	●	●		●	●	●	●	●	●	●	●	●	●	●
rpoB	●	●	●	●	●	●	●	×	●	●	●	●	●	●
rps9	●	●	●	×	●	●	●	×	●	●	●	●	●	●
rps15		×	×	×	×	×	×	×	●	●	×	×	×	×
rps16		×	×	×	×	×	×	×	●	●	×	×	×	×
rrf		×	●	●	●	●	●	×	●	●	●	×	×	●
tilS		×	×	×	×	×	●	×	×	×	●	×	×	×
ycf1	●	●	×	●	×	●	●	●	●	●	●	×	×	×
ycf3	●	●	●	●	●	●	●	●	●	●	●	×	×	×
ycf4	●	●	×	×	×	●	●	●	●	●	●	●	●	●
ycf20		●	×	●	×	●	P	×	×	×	×	×	×	×
ycf47		×	×	×	×	●	●	×	×	×	×	×	×	×
ycf62		×	×	×	×	×	●	×	×	×	●	×	×	×
ycf65		●	×	×	×	×	×	×	×	×	×	×	●	×
ycf66		×	×	×	×	×	×	×	●	●	×	×	×	×
ycf81		×	×	×	×	●	●	×	×	×	×	×	×	×
trnG(gcc)	●	●	×	●	×	●	●	●	×	●	●		●	●
trnH(gug)	●	●		●	●	●	●	●	●	●	●	●	●	●
trnI(cau)		●	×	●	×	●	●	×	●	●	●		●	●
trnL(caa)		×	×	×	×	●	●	×	●	●	●	●	×	●
trnL(gag)		×	●	×	●	●	●	●	×	×	●	×	×	×
trnL(uag)		●	●	●	●	●	●	●	●	●	●	●	●	●
trnP(ggg)		×	×	×	×	×	×	●	×	×	×	×	×	×
trnR(ccg)		●	×	×	×	●	●	●	×	×	●	×	×	×
trnS(cga)		×	×	×	×	●	●	●	×	×	×	×	×	×
trnS(gga)		×	×	×	×	●	●	●	×	×	●	×	×	×
trnT(ggu)		×	×	×	×	●	●	×	×	×	×	×	×	×
trnV(gac)		×	×	×	×	●	×	×	×	×	●	×	×	×
trnV(uac)	●	●		●	●	●	●	●	●	●	●	●	●	●

## 5.5. Fylogenomická analýza

Dataset 71 proteinů plastidových genů z celkem 60 zástupců skupin Chlorophyta, Streptophyta a Euglenophyceae byl použit k rekonstrukci fylogenetického stromu metodou maximum likelihood v RAxML (Stamatakis, 2014). Ve výsledném stromu (Obr. 9) vytváří Chlorophyta a Streptophyta dva monofyletické klany. Na bázi chlorofytního klanu se parafyleticky odvětvují linie skupiny Prasinophyta. Euglenophyceae se odděluje uvnitř Pyramimonadales. Vrchol chlorofytní linie tvoří UTC klad (Ulvophyceae, Trebouxiophyceae, Chlorophyceae) a *Pedinomonas*.



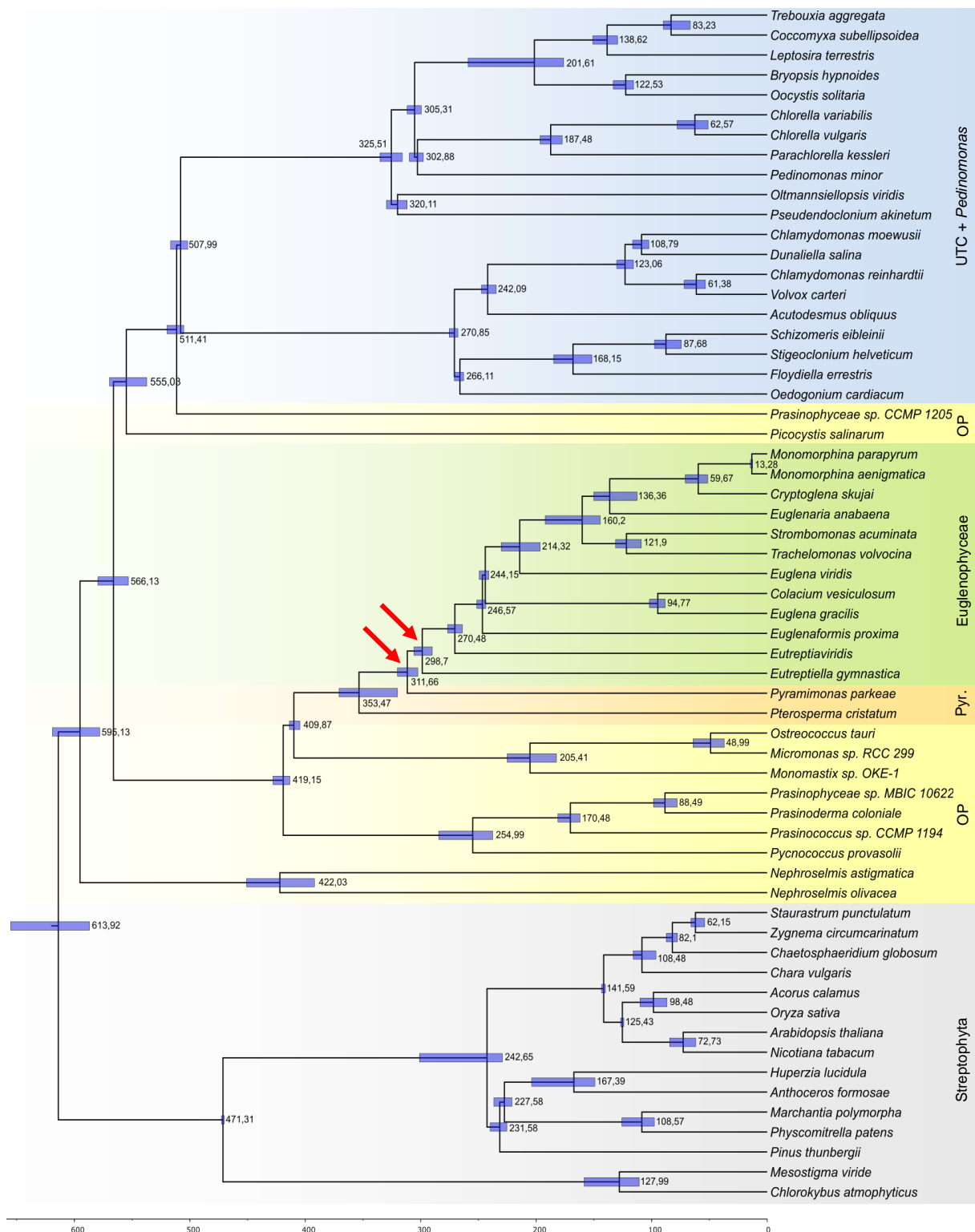
**Obr. 9. Fylogenetický strom zkonstruovaný na základě 71 plastidových genů.** Topologie byla vytvořena metodou maximum likelihood (ML) v programu RAxML. Hodnoty jednotlivých uzlů vyjadřují bootstrapovou podporu větvení zjištěnou metodou ML. \* označuje hodnotu nižší než 50 %. Římské číslice odpovídají prasinofytním kladům: Pyramimonadales (I), Mamiellophyceae (II), Nephroselmidophyceae (III), Pycnococcaeae (V), Prasinococcales (VI) a *Picocystis* klad (VII).

## 5.6. Molekulární datování

Pro analýzu molekulárních hodin byl rovněž použit dataset 71 proteinů plastidových genů. Strom byl kalibrován pomocí několika fosilních záznamů z publikovaných studií (Tab. 10). Byly provedeny dvě různé analýzy s odlišnými kalibračními body a burn-in (viz kapitola 4.5.4.). Všechny provedené analýzy vedly k posunutí stáří uzlů (především UTC kladu) blíže do současnosti, než bylo původně použito pro kalibraci (Obr. 10). Průměrná hodnota stáří uzlu Euglenophyceae vyšla v analýze se všemi kalibračními body 298,7 milionů let a uzlu Euglenophyceae + *Pyramimonas* 311,66 milionů let (červené šipky v Obr. 10). Odhady stáří obou uzlů se v jednotlivých analýzách nepatrně liší (Tab. 15).

**Tab. 15.** Stáří uzlů Euglenophyceae a Euglenophyceae + *Pyramimonas* podle analýz s odlišnými kalibračními body a burn-in. 95% kon. int. je zkratka pro 95% konfidenční interval.

analýza		uzel fylogenetického stromu (červené šipky v Obr. 10)			
kalibrace	burn-in (počet stromů)	Euglenophyceae		Euglenophyceae + <i>Pyramimonas</i>	
		průměr (milionů let)	95% kon. int. (milionů let)	průměr (milionů let)	95% kon. int. (milionů let)
s UTC	3600	298,7	305,9 – 290,0	311,7	320,2 – 302,6
bez UTC	4000	305,3	322,0 – 280,6	351,3	372,8 – 330,5
	9000	303,9	311,3 – 297,1	339,3	353,0 – 331,8



**Obr. 10. Časově kalibrovaný fylogenetický strom založený na analýze 71 plastidových proteinů (kalibrace s UTC uzlem).** Topologie a kalibrace byly vytvořeny Bayéskou metodou v programu BEAST. Strom byl kalibrován s použitím kalibračních bodů včetně uzlu UTC kladu shrnutých v tabulce 10. Číselné hodnoty na uzlech vyjadřují průměrnou hodnotu stáří uzlů v milionech. Modré čtverce znázorňují 95% konfidenční interval. Červené šipky ukazují na oddělení plastidů euglen od Pyramimonadales a na společného předka Euglenophyceae. Pyr. je zkratka pro Pyramimonadales, OP pro ostatní Prasinophyta.



## 6. Diskuse

### 6.1. Fylogeneze Pyramimonadales a Euglenophyceae na základě 16S a 18 rRNA

Evoluce taxonomických skupin organismů může být hodnocena studováním vývoje morfologických nebo molekulárních znaků. Rychlý rozvoj molekulárních metod nám v současnosti umožňuje získání velkého objemu molekulárních dat v relativně krátkém čase. Hlavními výhodami užití molekulárních znaků místo morfologických je především jejich uniformnost, známý způsob dědičnosti a použitelnost na různých taxonomických úrovních. Ve své diplomové práci jsem posuzovala fylogenezi Euglenophyceae a Pyramimonadales především na základě široce využívaných molekulárních markerů 16S rRNA (Obr. 6) a 18S rRNA (Obr. 4 a 5).

Skupina Euglenophyceae vykazuje ve všech fylogenetických stromech monofyletické uspořádání. Také Euglenales vytváří monofyletickou skupinu, od jejíž stonku se parafyleticky odvětvují mořští zástupci Eutreptiales (Obr. 4 a 6). To souhlasí s publikovanými výsledky fylogenetických studií (Kim et al., 2010; Linton et al., 2010; Karnkowska et al., 2015). Zajímavé jsou především pozice jednotlivých sekvencí označených jako *Eutreptiella gymnastica* na 18S rRNA stromu (Obr. 4). Zatímco sekvence stažené z databáze GenBank se větví spolu s ostatními druhy rodu *Eutreptiella*, sekvence pocházející z databáze iMicrobe vytváří oddělený klad sesterský linii *Eutreptia* s vysokou bootstrapovou podporou 98 % (Obr. 4). To ukazuje na větší genetickou diverzitu v rámci skupiny a možná existenci nového rodu. Analýza 16S rRNA (Obr. 6) i 18S rRNA ukázala vnitřní členění Euglenales na třídy Euglenaceae a Phacaceae (Kim et al., 2010; Karnkowska et al., 2015). Vnitřní vztahy mezi jednotlivými rody nebyly nicméně dobře statisticky podpořené kvůli nemonofyletickému uspořádání některých rodů (např. rody *Discoplastis*, *Phacus* a *Euglena*) (Obr. 6). Výsledky fylogenetických analýz 16S rRNA (Obr. 6) a zejména plastidových genů (Obr. 9) ukázaly na příbuznost plastidů euglen ke skupině Pyramimonadales.

Pyramimonadales vykazují na základě 18S rRNA monofyletické uspořádání (Obr. 5). Ve studii využívající 16S rRNA sekvence se nicméně *Pteroperma cristatum* větví uvnitř klanu zahrnujícího ostatní Prasinophyta, Chlorophyta a environmentální sekvence (Obr. 6), čímž dělá skupinu Pyramimonadales nemonofyletickou. Tato její pozice nebyla nicméně dostatečně podpořena, a pokládám jí proto za chybnou. Na rozdíl od jiných studií zabývajících se fylogenezí v rámci Pyramimonadales (Viprey et al., 2008) vytváří *Pterosperma* a *Pyramimonas* v mé 18S rRNA analýze oddělené klady (Obr. 5). Podpora větvení při uskupení do oddělených kladů i do komplexu *Pterosperma-Pyramimonas* (Viprey et al., 2008) je však poměrně nízká nepřesahující 60 %. Pro vyjasnění této situace bude proto zapotřebí dalšího zkoumání. Také rozdělení kladu *Pyramimonas* na podskupiny (Obr. 5) souhlasí s dřívějšími publikacemi (Daugbjerg et al., 1994; Suda, 2004; Viprey et al., 2008; Harðardóttir et al., 2014), avšak příbuzenské vztahy mezi jednotlivými podrody se liší.

## 6.2. Analýzy environmentálních sekvencí

Jedním z hlavních cílů této práce bylo hledání sekvencí blízké příbuzných chloroplastu euglenidů v environmentálních vzorcích a to jak vlastních, tak sekvencí dostupných v databázích (NCBI, TARA). Snažila jsem se tak důkladněji prozkoumat diverzitu v této části stromu, protože dosavadní studie zabývající se vznikem chloroplastů euglen používaly v analýzách jen jediného zástupce Pyramimonadales, a to *Pyramimonas parkeae*.

Environmentální studie zabývající se zkoumáním moří a oceánů v minulosti odhalily dříve nepozorovanou diverzitu uvnitř prasinofytních linií (Díez et al., 2001; Guillou et al., 2004; Shi et al., 2009). Naprostá většina získaných sekvencí byla nicméně příbuzná široce rozšířeným pikoplanktonním příslušníkům rodů *Ostreococcus*, *Bathycoccus*, *Micromonas*, *Mantoniella* a prasinofytním kladům VII a IX (Díez et al., 2001; Guillou et al., 2004; Viprey et al., 2008, Shi et al., 2009). Daleko menší část osekvenovaných vzorků byla přiřazena ke skupině Pyramimonadales (Guillou et al., 2004; Viprey et al., 2008; de Vargas et al., 2015) poukazujíc na nižší abundanci v prostředí. Ve své práci jsem se však cíleně zaměřila na environmentální sekvence příbuzné linii Pyramimonadales a ukázala jsem, že i tato skupina je v prostředí hojně zastoupená. Značnou genetickou diverzitu jsem detekovala především v druhovém komplexu *Pterosperma* a *Halosphaera* (Obr. 5), ale také v příbuzenstvu *Pyramimonas parkeae* a *Pyramimonas disomata* (Obr. 5 a 6). Žádná ze zkoumaných environmentálních sekvencí nicméně nevykazuje bližší příbuznost k plastidům euglen než sekvence popsaných organismů.

Značnou environmentální diverzitu jsem zjistila také uvnitř Euglenophyceae. Analýza 16S rRNA odhalila devět TARA a jednu NCBI sekvenci (KF512936), které se, mnohdy s malou podporou, větví na bázi Euglenophyceae (Obr. 6). Některé pravděpodobně pochází z *Rapaza viridis*, mixotrofního zástupce euglen (Yamaguchi et al., 2012), jejíž plastidovou 16S rRNA zatím neznáme. Jiné by mohly reprezentovat neznámé mořské linie bazálních Euglenophyceae. Zřejmě tedy existuje více mořských linií Eutreptiales, než se původně předpokládalo. Dalších 15 environmentálních 16S rRNA sekvencí z expedice TARA spadá ke sladkovodním taxonům *Discoplastis*, *Euglena*, *Monomorphina* a *Phacus* (Obr. 6). TARA sekvence větvící se uvnitř Euglenales jsou dokonce četnější než uvnitř Eutreptiales. Vzhledem k popsanému výskytu daných skupin, bych očekávala spíše opačný trend. V mořském prostředí by se tudíž mohla nacházet dříve nepozorovaná diverzita Euglenales (diskutováno níže).

### 6.2.1. Amplifikace částí plastidového genomu

Důvodem, proč je zelená řasa *Pyramimonas parkeae* považována za nejbližšího známého příbuzného chloroplastu euglenidů, je mimo jiné také konzervované pořadí některých genů v plastidových genomech organismů rodu *Pyramimonas* a *Euglena* (Obr. 1) (Turmel et al., 2009). Jim příbuzné organismy by měly ve svém genomu rovněž tyto genové oblasti obsahovat. Toho jsem využila pro amplifikaci sekvencí v environmentálních vzorcích. Analýza genových klastrů je oproti 16S a 18S

rRNA zaměřená jen na konkrétní organismy, jenž tyto genové klastry sdílí. Otevírá tak možnost, jak zacílit naše sekvenování pouze na skupinu příbuzných organismů.

K amplifikaci sekvencí v environmentálních vzorcích (Tab. 1) jsem ve své práci použila zejména dva genové klastry: rps4-rps11 a trnC(gca)-rps2 (Obr. 1). Důvodem byla především jejich dostatečná délka, která je nezbytná pro správnou interpretaci příbuzenských vztahů mezi organismy. U řady environmentálních vzorků se mi bohužel nepodařilo amplifikovat žádný produkt nebo byla koncentrace produktu po přečištění natolik nízká, že její sekvenace neprodukovala čitelný výsledek. Problém pravděpodobně spočíval v nedostatečném množství DNA cílových organismů ve vzorku. Jiné environmentální vzorky (Dánsko23, 24) naopak poskytly několik různých sekvencí (Obr. 3) poukazující na přítomnost více než jednoho příbuzného organismu. V ojedinělých případech se mi podařilo získat sekvence ze stejného vzorku u obou klastrů (Lod'1, Německo14) (Tab. 11). Zatímco sekvence vzorku Lod'1 se v obou fylogenetických stromech větví uvnitř Pyramimonadales, sekvence pocházející ze vzorku Německo14 náleží k odlišným skupinám. V analýze klastru rps4-rps11 (Obr. 2) spadá mezi Pyramimonadales a u klastru trnC(gca)-rps2 (Obr. 3) vytváří vnitřní euglenofytní linii. To opět ukazuje na přítomnost více než jednoho příbuzného organismu ve vzorku Německo14.

Hlavním záměrem analýz environmentálních vzorků bylo nalezení linie větvcí se na bázi Euglenophyceae, která by mohla představovat bližšího předka plastidů euglen, než jsou dosud známí Pyramimonadales. To se však nepodařilo, protože všechny sekvence se větví buď uvnitř Pyramimonadales, nebo Euglenophyceae a žádná nevychází z vnitřní větve oddělující oba taxony (Obr. 2 a 3). Sekvence environmentálních vzorků větvcí se uvnitř Pyramimonadales pravděpodobně pochází z organismů rodu *Pyramimonas* či jim příbuzným. Zajímavý vzorek by mohl představovat Norsko19 (amplifikovaný klastr rps4-rps11), který leží na bázi všech ostatních členů kladu *Pyramimonas* (Obr. 2). Kromě organismu *Pyramimonas parkeae* nebyl dosud přečten plastidový genom z žádného jiného zástupce skupiny Pyramimonadales. Vzorek Norsko19 reprezentuje organismus, jehož plastidový genom by spolu s cpDNA (chloroplastová DNA) řasy *Pterosperma cristatum* a *Rapaza viridis* mohl pomoci v rekonstrukci podoby chloroplastu získaného sekundární endosymbiotickou událostí u předka všech fototrofních euglen. Z fylogenetického stromu 16S rRNA (Obr. 6) a plastidových proteinů (Obr. 9) totiž vyplývá, že k pohlcení plastidového donora předkem Euglenophyceae došlo až po oddělení *Pterosperma* a *Pyramimonas*, ale ještě před diverzifikací uvnitř rodu *Pyramimonas*.

Fylogenetický strom rekonstruovaný ze sekvencí genového klastru trnC(gca)-rps2 navíc odkryl mořskou environmentální linii zahrnující sekvence vzorků Německo14, Kanada28 a 30, která se větví uvnitř sladkovodních Euglenales (Obr. 3). Kontaminace vzorků během přípravy a detekce je velmi nepravděpodobná, neboť jsem v této studii nepracovala s DNA pocházející z euglenofytních organismů (diskutováno níže).

### 6.2.2. Sladkovodní skupiny Pyramimonadales

Vzhledem k mořským bazálním skupinám Eutreptiales (Marin et al., 2003; Kim et al., 2010; Linton et al., 2010; Karnkowska et al., 2015) a skutečnosti, že většina zástupců Prasinophyceae obývá rovněž mořské prostředí, se předpokládá, že k pohlcení zelené řasy předkem fototrofních euglen došlo právě v moři (Marin, 2004). Mezi Pyramimonadales jsou však známé některé sladkovodní druhy (Swale a Belcher, 1968). To může vyvolat otázky, jestli se endosymbióza odehrála opravdu v moři. Typický sladkovodní druh *Pyramimonas tetraarhynchus* byl do analýz rovněž zahrnut. Zatímco v analýze genového klastru trnC(gca)-rps2 (Obr. 3) vytváří sesterskou větev k linii zahrnující *P. parkeae* a environmentální sekvence, v 16S rRNA stromu (Obr. 6) se větví na bázi kladu *Pyramimonas*. Ani v jednom případě není však jeho postavení dobře statisticky podpořeno s bootstrapem nepřesahujícím hodnotu 50 %. Mnohem více podpořenou pozici zaujímá *P. tetraarhynchus* ve fylogenetickém stromu 18S rRNA, kde se větví uvnitř podskupiny *Pyramimonas* (Obr. 5). Nepodařilo se mi amplifikovat žádný genový klastř ze vzorků Německo1, 2 a 8, které byly odebrány ze sladkého bahna či potoka (Tab. 1). Podobná studie zaměřená na sladkovodní plochy by nicméně mohla přinést zajímavé poznatky porovnatelné s výsledky této práce, doplnit poznatky o diverzitě sladkovodních Pyramimonadales a případně ukázat, zdali k endosymbiotické události mezi předkem Euglenophyceae a zelenou řasou došlo skutečně v moři.

### 6.2.3. Mořské linie Euglenales

Přechod mezi mořským a sladkovodním prostředím není v říši jednobuněčných organismů zcela neobvyklý. Mezi skupiny se známým výskytem v obou typech prostředí patří například Cryptophyta (Shalchian-Tabrizi et al., 2008), Haptophyta (Shalchian-Tabrizi et al., 2011) a spousta dalších. Kolonizace může probíhat v obou směrech – z moře do sladkovodních ploch i naopak (Shalchian-Tabrizi et al., 2008). Již dříve bylo jasné, že v evoluci Euglenophyceae došlo minimálně jednou k takovému přechodu, a to zřejmě z moře do sladkovodního prostředí. Bazální linie Euglenophyceae totiž tvoří mořské skupiny *Eutreptia*, *Eutreptiella* a *Rapaza*, zatímco ostatní zástupci (Euglenales) obývají sladkovodní areály. Tuto topologii lze nejnázat vysvětlit tím, že původní mořské organismy jednou přešly do sladkovodního prostředí, kde záhy diverzifikovaly.

Výsledky mé práce ovšem odhalily skrytou diverzitu jinak sladkovodních Euglenales v mořských vzorcích. Objevila jsem totiž několik 16S rRNA sekvencí a amplifikovala jsem mořskou linii v environmentálních vzorcích Německo14, Kanada28 a 30 (genový klastř trnC(gca)-rps2), které se větví spolu se sladkovodními taxony Euglenophyceae (Obr. 3 a 6). Tyto sekvence přitom pocházejí z moří a oceánů různých částí světa (Mapa E, Tab. 1). Provedené analýzy tak potvrdily existenci dříve nepozorované diverzity Euglenales v mořském prostředí, která je patrně široce rozšířená. Organismy kolonizující určitý typ prostředí většinou vytváří ve fylogenetických analýzách nezávislé linie vmezežené mezi linie zástupců původního areálu (Shalchian-Tabrizi et al., 2011; Yamaguchi et al.,

2011). Sladkovodní Euglenales navíc vykazují ve fylogenetických stromech monofyletické uspořádání (Kim et al., 2010; Linton et al., 2010; Karnkowska et al., 2015). Domnívám se proto, že nové euglenofytní environmentální linie spadající mezi Euglenales vznikly spíše rekolonizací moře než opakovaným osídlením sladkovodního prostředí.

#### 6.2.4. Úskalí environmentálního sekvenování

Analýzy environmentálních sekvencí sice odhalily možnost linií větvcích se uvnitř skupin Pyramimonadales či Euglenophyceae. Jediné informace, které o těchto environmentálních liniích máme, jsou ale pouze nukleotidové sekvence a místa odběru vzorků. Použití environmentálního přístupu přináší samo o sobě jistá omezení. Data získaná touto metodou nám sice umožňují studovat diverzitu (Guillou et al., 2004) či distribuci (Viprey et al., 2008) konkrétní taxonomické skupiny, o chování, vzhledu či stylu života nám však neřeknou téměř nic. Mnohé studie se proto uchylují k zařazení i dalších rozšiřujících metod. Kolodziej a Stoeck (2007) například použili specifické oligonukleotidové sondy ve spojení s fluorescenční *in situ* hybridizací (FISH) a skenovací elektronovou mikroskopií (SEM) k zacílení a prozkoumání struktury konkrétní buňky environmentální linie stramenopil. Seenivasan et al. (2013) popsali dokonce úplně nový druh za pomoci izolace konkrétní buňky z environmentálního vzorku použitím fluorescenčně-aktivovaného třídění buněk (FACS).

Známe tedy metody, které umožňují blíže studovat environmentální vzorky. Cílem mé práce však bylo především získat nukleotidové sekvence a podívat se, jestli se v prostředí nachází zajímavé linie, které by mohly lépe vysvětlit endosymbiotickou událost mezi zelenou řasou a předkem dnešních fototrofních Euglenophyceae. Fylogenetické studie provedené na genových klastrech a molekulárních markerech 16S a 18S rRNA sice neprokázaly existenci environmentální sekvence, která by se větvila mezi Euglenophyceae a Pyramimonadales, zajímavá by však mohla být sekvence pocházející ze vzorku Norsko19, větvcí se na bázi kladu *Pyramimonas* (Obr. 2). Získání více informací o daných skupinách například za pomoci kultivace by každopádně bylo velmi přínosné. Studie by se proto mohla v budoucnu ubírat právě tímto směrem a vzorek Norsko19 by mohl být jedním z potenciálních adeptů na získání kultury a následnou sekvenaci plastidového genomu. Znalost uspořádání plastidového genomu organismu, z kterého sekvence pochází, by mohla totiž pomoci při rekonstrukci podoby chloroplastu u předka všech Euglenophyceae. Poměrně překvapivé bylo nalezení environmentálních linií Euglenales. Jejich existenci bychom rádi potvrdili kultivací. Náš výzkumný tým proto plánuje další výpravu do těchto končin s hlavním cílem kultivovat mořské Euglenales.

#### 6.3. Plastidový genom *Pterosperma cristatum*

Prasinofytní řada *Pterosperma cristatum* představuje po *Pyramimonas parkeae* teprve druhého zástupce skupiny Pyramimonadales, jejíž plastidový genom byl alespoň částečně osekvenován. Získali jsme tak další bod pro rekonstrukci fylogeneze, který pomůže lépe vymezit klad Pyramimonadales

a postavení plastidů euglen vůči tomuto kladu. Plastidový genom se mi nepodařilo zkompletovat, takže je tvořen dvěma scaffoldy (Obr. 7 a 8).

Částečná cpDNA *Pterosperma cristatum* dosahuje celkové délky 129 578 bp (párů bází). Velikosti plastidových genomů dosud osekvenovaných prasinofytních řas se pohybují od cca 70 000 bp u drobného organismu *Ostreococcus tauri* (Robbens et al., 2007; Blanc-Mathieu et al., 2013) po více než 200 000 bp řasy *Nephroselmis olivacea* (Turmel et al., 1999). V porovnání s blízkým příbuzným *Pyramimonas parkeae*, jehož plastidový genom obsahuje celkem 101 605 bp (Turmel et al., 2009), se délka cpDNA *Pterosperma cristatum* poměrně liší. Rozdíly ve velikosti plastidových genomů jednotlivých zástupců jsou však v rámci prasinofytních kladů poměrně časté. *Nephroselmis olivacea* má kupříkladu cpDNA o zhruba 75 000 bp delší než příslušník stejného rodu *N. astigmatica*. Zelená řasa z třídy Mamiellophyceae *Monomastix* sp. svou velikostí cpDNA dosahující 114 528 bp (Turmel et al., 2009) také výrazně převyšuje všechny ostatní zástupce tohoto kladu. Není tedy příliš zvláštní, že by se podobné odlišnosti vyskytovaly i v rámci skupiny Pyramimonadales.

Velikost cpDNA *P. cristatum* může být nicméně také způsobena špatným složením některých úseků genomu. Některé oblasti genomu totiž nezahrnují žádné anotované geny. Jedná se především o začátky a konce obou scaffoldů (Obr. 7 a 8). Není běžné, že by plastidové genomy obsahovaly dlouhé nekódující oblasti. Naopak geny leží typicky naskládány těsně vedle sebe. Domnívám se proto, že v případě cpDNA *P. cristatum* se v podobných místech genomu vyskytují nějaké nesrovnalosti, jež mohly vzniknout během skládání genomu. Příslušné části genomu by měly být v budoucnu ještě jednou překontrolovány sekvenací. Tím by se také mohly vyřešit nesrovnalosti uvnitř genu chlB (viz kapitola 5.4.).

Částečný plastidový genom řasy *Pterosperma cristatum* obsahuje celkem 92 anotovaných genů. Převážná část genů zajišťuje fungování samotného chloroplastu a fotosyntézy, nachází se zde ale také geny důležité pro stavbu ribozómů, proteosyntézu, transkripci a translaci (Tab. 12). Některé geny byly označeny jako částečné nebo hypotetické (Obr. 7 a 8), protože se domnívám, že jejich sekvence nebyly předpovězeny správně. Soudím tak ze srovnání jejich sekvencí a sekvencí jim příbuzných homologů z jiných organismů. Domnívám se, že problémy pramení právě ve špatném složení některých míst genomu, následkem čehož mohlo dojít k posunutí čtecího rámce, a tím k pozměnění proteinových sekvencí některých genů. Ani v případě necelého genu chlN, kterému schází velká část 5' konce sekvence, ani v případě příliš dlouhých genů chlI, ftsH, rpoA a rpoC2 se mi nepodařilo zjistit příčinu nesrovnalostí. Zdá se, jako by otevřené čtecí rámce (ORFs) genů chlI, ftsH, rpoA a rpoC2 nebyly včas ukončeny stop kodonem. To by mohlo ukazovat na přítomnost alternativního stop kodonu. Frekvence výskytu jednotlivých kodonů však tuto skutečnost nepotvrdila (Tab. 13). Přítomnost alternativního stop kodonu by se prokázala nulovou frekvencí jeho použití v ostatních proteinech. *P. cristatum* však používá všechny kodony vyjma klasických stop kodonů pro kódování aminokyselin (Tab. 13), takže žádný z použitých kodonů nemůže představovat alternativní stop kodon.

Mezi anotovanými geny jsem zjistila dva geny *psbD* a *ycf3* obsahující každý jeden intron (Obr. 7 a 8). Plastidové genomy prasinofytních řas nejsou obecně moc bohaté na introny, několik jich však bylo v publikovaných cpDNA již detekováno. V naprosté většině případů se v genomech prasinofytních řas vyskytuje jeden či dva introny. Nejčastějším nalezeným genem s intronem je *atpB* popsáný například v genomu *Pycnococcus provasolii*, *Pyramimonas parkeae* (Turmel et al., 2009) či *Ostreococcus tauri* (Robbens et al., 2007). Častá je také přítomnost intronů v genech *psbA* či *rrl* (Lemieux et al., 2014). *Picocystis salinarum* má stejně jako *P. cristatum* intron ležící v genu *ycf3*. Gen *ycf3* *Picocystis salinarum* je v intronu rozdělený na dvě části, které vypadají jako samostatné transkripční jednotky (Lemieux et al., 2014). Na rozdíl od toho gen *ycf3* *P. cristatum* vytváří souvislý úsek cpDNA. V plastidových genomech příslušníků Euglenophyceae se oproti skupině Prasinophyta vyskytuje velké množství genů obsahující introny. Počet intronů zástupců bazálních linií Eutreptiales se pohybuje od dvou intronů u *Eutreptiella gymnastica* (Hrdá et al., 2012) až po 23 u *Eutreptia viridis* (Wiegert et al., 2012). Rapidní nárůst intronů můžeme zaznamenat u sladkovodních Euglenales, jejichž počet se může blížit až 150 (Hallick et al., 1993; Wiegert et al., 2013). K navýšení počtu intronů v plastidových genomech Euglenophyceae však zřejmě došlo teprve až během evoluce jednotlivých taxonů, dlouho po polapení a ustálení sekundárního endosymbionta v podobě chloroplastu (Thompson et al., 1995).

### 6.3.1. Rozdíly v genovém obsahu prasinofytních řas a Euglenophyceae

Genový obsah cpDNA se v rámci jednotlivých taxonů prasinofyt poměrně liší. U drobných pikoplanktonních organismů, jako je *Ostreococcus tauri*, došlo zřejmě v důsledku zmenšování k redukci mimo jiné také organelové DNA. Jeho cpDNA kóduje celkem 86 genů (Robbens et al., 2007). Oproti tomu řasy s poměrně velkými buňkami mají rozsáhlejší plastidové genomy. Mezi zatím největší známé, co se obsahu týče, patří genomy zástupců rodu *Nephroselmis* čítající 123 genů v případě *N. astigmatica* (Lemieux et al., 2014) a 127 genů u *N. olivacea* (Turmel et al., 1999). V plastidovém genomu řasy *Pterosperma cristatum* jsem anotovala celkem 92 genů (Tab. 12), což se blíží 110 genům kódovaných cpDNA *Pyramimonas parkeae*. Tento počet nicméně nemusí být konečný, neboť chybí část cpDNA mezi dvěma scaffoldy.

Ve srovnání s ostatními prasinofyty schází v plastidovém genomu *P. cristatum* dva geny *atpE* a *trnL(uag)* (Tab. 14), které jsou konzervované napříč celou skupinou. Tyto geny budou s největší pravděpodobností přítomny v chybějící části genomu. Genom *P. cristatum* pravděpodobně neobsahuje geny kódující podjednotky komplexu I elektronového transportního řetězce (*ndhA-K*), neboť jsem žádný z těchto genů v její cpDNA nenašla. To je srovnatelné například s plastidovými genomy *Pycnococcus provasolii*, *Prasinoderma coloniale* a zástupci skupiny Mamiellophyceae (Tab. 14). Oproti blízké příbuzné řase *Pyramimonas parkeae* postrádá *P. cristatum* dalších sedm plastidových genů (Tab. 14). Konkrétně se jedná o geny: *ccsA*, *chlL*, *rpl12*, *ycf20*, *65*, *trnI(cau)*, *trnR(ccg)*. V rámci

jednoho kladu se nicméně přítomnost určitých genů často do jisté míry liší (Tab. 14). Není proto tak překvapivé, že i obsah genomů rodů *Pyramimonas* a *Pterosperma* se může různit a některé geny opravdu nejsou v genomu přítomny. Příkladem takové situace je přítomnost genu *petD* u *P. cristatum*, který naopak chybí v cpDNA *Pyramimonas parkeae* (Tab. 14). To, jaké ze zbývajících genů cpDNA *P. cristatum* obsahuje či nikoliv, odhalí nicméně pouze dosekvenování chybějící cpDNA mezi oběma scaffoldy.

Zástupci Euglenophyceae se v genovém obsahu ve srovnání s prasinofytními skupinami významně liší (Tab. 14). V průběhu evoluce vedoucí od předka po dnešní zástupce Euglenophyceae došlo k úpravám plastidového genomu původního pohlceného endosymbionta. Během toho byla celá řada genů pravděpodobně ztracena (Hrdá et al., 2012). Následovala velká expanze intronů, až cpDNA nabyla podoby, jakou můžeme pozorovat dnes. V plastidovém genomu *Pterosperma cristatum* se nachází některé konzervované genové oblasti nebo jejich části, které sdílí ve své cpDNA *Pyramimonas parkeae* a *Euglena gracilis* (Obr. 1). Pořadí genů *rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rpl5-rps8-infA* a *rrs-trnI(gau)-trnA(ugc)-rrl* je odvozeno z ancestrálního uspořádání cpDNA vyskytujícího se také u ostatních chlorofyt a streptofyt (Turmel et al., 2009). Genové klastry *rpl20-rps12-rps7-tufA* a *rps4-rps11* představují ovšem odvozené uspořádání plastidových genů. Pořadí ostatních unikátních genových oblastí je v plastidovém genomu *P. cristatum* odlišné oproti *Pyramimonas* a *Euglena*. Tato skutečnost spolu s výsledky fylogenomické analýzy (Obr. 9) potvrzuje bližší příbuznost chloroplastů euglen k *Pyramimonas parkeae* než k *Pterosperma cristatum*.

#### 6.4. Fylogenomická analýza

Osekvenování a anotace *Pterosperma cristatum* poskytla další data pro rozsáhlou fylogenomickou analýzu zelené linie řas a rostlin. Do původního datasetu použitého v publikaci Lemieux et al. (2014) jsem kromě *P. cristatum* přidala také proteiny plastidových genů Euglenophyceae. Výsledný strom (Obr. 9) se oproti jiným studiím zaměřených na Prasinophyta (Guillou et al., 2004; Nakayama et al., 2007; Turmel et al., 2009, Lemieux et al., 2014) liší hlavně v topologii jednotlivých prasinofytních kladů. První linii, která se se oddělila od společného předka skupiny Chlorophyta, představují Nephroselmidophyceae s maximální bootstrapovou podporou (Obr. 9). Dříve nepozorovaná příbuznost je také mezi klady Pycnococceae a Prasinococcales, které navíc vytváří sesterskou větev komplexu Pyramimonadales-Mamiellophyceae (Obr. 9). Podpora větvení mezi touto částí stromu a zbytkem stromu je nicméně poměrně slabá dosahující hodnoty 54 %. Domnívám se proto, že je pozice některých linií nesprávná. Euglenophyceae se odvětvují uvnitř Pyramimonadales jako sesterské *Pyramimonas*, čímž opět potvrzují původ plastidů euglen v této skupině organismů.



## 6.5. Datování endosymbiotické události

Metoda molekulárních hodin umožňuje datovat přibližné stáří nodů fylogenetického stromu, a tím odhalit předpokládanou dobu vzniku jednotlivých taxonomických skupin. Podobným způsobem je možné datovat také endosymbiotickou událost, která se odehrála mezi posledním společným předkem všech Euglenophyceae a odvětvením Euglenophyceae od Pyramimonadales (červené šipky v Obr. 10). V této práci jsem provedla dvě analýzy molekulárního datování s odlišnými kalibračními body a burn-in (viz kapitola 4.5.4. a Tab. 10). Výsledky těchto analýz se v odhadu stáří sekundární endosymbiózy, která dala vzniku plastidům euglen, nepatrně liší (Tab. 15). Staří uzlu Euglenophyceae + *Pyramimonas* dosahuje v konfidenčním intervalu nejvyšší hodnoty 372,8 milionů let a uzel Euglenophyceae má v konfidenčním intervalu nejnižší hodnotu 280,6 milionů let. Pokud předpokládáme, že k sekundární endosymbióze došlo někde mezi těmito dvěma uzly, potom můžeme vznik plastidů euglen datovat do období před 372,8–280,6 miliony lety. Diverzifikace skupiny Euglenophyceae se zdá být o něco starší, než předpověděli v dřívějších studiích Parfreyová et al. (2011) a Stiller (2014). Do analýz nicméně nebyla zahrnuta *Rapaza viridis*, která se odvětjuje od předka Euglenophyceae ještě před skupinou Eutreptiales (Yamaguchi et al., 2011). Její pozice na stromu posune uzel Euglenophyceae ještě o něco dále do minulosti.

Provedené analýzy molekulárního datování se nicméně potýkaly s několika problémy. Malý skok pozorovaný v hodnotách likelihoodu těsně před koncem řetězce MCMC (Markov Chain Monte Carlo) v analýze bez UTC kalibračního bodu poukazuje na to, že řetězec nestihl konvergovat do oblasti stromů s nejlepší topologií. Analýza se všemi kalibračními body se sice dostala do delší fáze plató, která běžně poukazuje na konvergenci řetězce. Jestli ale MCMC skutečně konvergoval, však nevíme. Špatné uspořádání topologií některých větví uvnitř linií UTC + *Pedinomonas* a Streptophyta (Obr. 10), například oddělení nahosemenných (*Pinus thunbergii*) ještě před mechorosty (*Anthoceros formosae*, *Physcomitrella patens*, *Marchantia polymorpha*) a plavuňovitými (*Huperzia lucidula*), ovšem ukazuje na to, že ke konvergenci nejspíš nedošlo. Z nám neznámého důvodu došlo rovněž k velkému posunutí některých nakalibrovaných bodů blíže do současnosti (Obr. 10, Tab. 10). Uvědomujeme si proto, že zde předkládané analýzy nejsou zcela bezchybné a nepovažujeme odhad stáří sekundární endosymbiotické události za velmi věrohodný. V budoucnu by měla být analýza znovu provedena s přihlédnutím k předchozím nedostatkům, především času potřebnému ke konvergenci řetězce a parametrům kalibrace stromu.

## 7. Závěr

Historické pátrání po donoru chloroplastu euglen dovedlo naše poznání až ke skupině Pyramimonadales (Turmel et al., 2009). Samotná endosymbióza se zřejmě uskutečnila v moři (Marin, 2004) mezi předkem všech Euglenophyceae a řasou, která se svým uspořádáním plastidového genomu hodně podobala *Pyramimonas parkeae*. Výsledky mé diplomové práce navíc ukázaly následující:

(1) Metoda amplifikace unikátních genových klastrů sdílených organismy *Euglena gracilis* a *Pyramimonas parkeae* se ukázala jako přínosná pro zacílení sekvenování na konkrétní skupinu příbuzných organismů. Odhalila několik environmentálních linií, ale nepotvrdila existenci organismu bližšího plastidům euglen, než je *P. parkeae*. Zajímavý environmentální vzorek nicméně představuje Norsko19, jenž by mohl vzhledem k bazální pozici v linii *Pyramimonas* pomoci při rekonstrukci podoby chloroplastu u předka Euglenophyceae. Překvapivé je také nalezení environmentální mořské linie větvcí se uvnitř sladkovodních Euglenales.

(2) Důležitým poznatkem je zde publikovaná anotace plastidového genomu *Pterosperma cristatum*, jež představuje teprve druhý organismus ze skupiny Pyramimonadales, u kterého byl osekvenovaný plastidový genom. Znalost obsahu a uspořádání plastidových genomů prasinofytních a euglenofytních řas nám přitom může pomoci při rekonstrukci evoluce endosymbiotické události.

(3) Rozsáhlé fylogenetické studie 16S rRNA, 18S rRNA a plastidových genů potvrdily příbuzenské vztahy mezi Pyramimonadales a Euglenophyceae. Odkryly rovněž značnou diverzitu Pyramimonadales. Analýzy 16S rRNA sekvencí z expedice TARA navíc odhalily dříve nepozorované environmentální mořské linie příbuzné sladkovodním taxonům Euglenales, které spolu s výsledky amplifikace genových klastrů (bod 1) potvrzují existenci mořských skupin Euglenales.

(4) Pomocí metody molekulárních hodin byla sekundární endosymbióza mezi předkem Euglenophyceae a řasou ze skupiny Pyramimonadales odhadnuta na období před 372,8–280,6 miliony lety. Vzhledem k některým problémům během molekulárního datování musíme však brát tento výsledek s rezervou.

Tato práce přinesla zpřesňující, ale i nové poznatky o environmentální diverzitě skupin Pyramimonadales a Euglenophyceae. Jaké zelené řasy se staly donorem plastidů euglen, a daly tak vzniknout unikátním fotosyntetickým organismům, však zatím stále přesně nevíme. Práce podobného typu by však mohly v budoucnu vést k odhalení záhad spojených s endosymbiotickými událostmi, jež tak zásadně ovlivnily život na naší planetě.

## 8. Seznam použitých zkratek

16S rRNA	ribosomální RNA prokaryotní velikosti, součást malé podjednotky ribozomu
18S rRNA	ribosomální RNA eukaryotní velikosti, součást malé podjednotky ribozomu
AK	aminokyselina
ATP	adenosintrifosfát
CDS	kódující oblast genu
cpDNA	chloroplastová DNA
DNA	deoxyribonukleová kyselina
FACS	fluorescenčně-aktivovaného třídění buněk
FISH	fluorescenční <i>in situ</i> hybridizace
IPTG	isopropyl $\beta$ -D-1-thiogalaktopyranosid
ML	maximum likelihood
ORF	otevřený čtecí rámeček
PCR	polymerázová řetězová reakce
RNA	ribonukleová kyselina
rRNA	ribosomální RNA
RuBisCO	ribulóza-1,5-bisfosfát karboxyláza/oxygenáza
SEM	skenovací elektronová mikroskopie
SSU rDNA	gen pro SSU rRNA
SSU rRNA	RNA malé podjednotky ribosomu
TAE	tris-acetát-EDTA
tRNA	transferová RNA
UTC	klad sdružující skupiny Ulvophyceae, Trebouxiophyceae a Chlorophyceae
UV	ultrafialové záření

## 9. Seznam použité literatury

Převzatá citace je označena \*

- Al-Dhaheeri R.S., Willey R.L., 1996. Colonization and reproduction of the epibiotic flagellate *Colacium vesiculosum* (Euglenophyceae) on *Daphnia pulex*. *Journal of Phycology* 32, 770–774.
- Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A. V., Sirotkin A. V, Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology: a journal of computational molecular cell biology* 19, 455–477.
- Becker B., Marin B., Melkonian M., 1994. Structure, composition, and biogenesis of prasinophyte cell coverings. *Protoplasma* 181, 233–244.
- Bennett M.S., Triemer R.E., 2015. Chloroplast genome evolution in the Euglenaceae. *The Journal of eukaryotic microbiology* 62, 773–785.
- 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, 66–73.
- Blanc-Mathieu R., Sanchez-Ferandin S., Eyre-Walker A., Piganeau G., 2013. Organellar inheritance in the green lineage: insights from *Ostreococcus tauri*. *Genome biology and evolution* 5, 1503–1511.
- Cavalier-Smith T., 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *The Journal of eukaryotic microbiology* 46, 347–366.
- Crisuolo 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 evolutionary biology* 10, 210.
- Daugbjerg N., Moestrup O., Arctander P., 1994. Phylogeny of the genus *Pyramimonas* (Prasinophyceae, chlorophyta) inferred from the rbcL gene. *Journal of Phycology* 30, 991–999.
- de Vargas C., Audic S., Henry N., Decelle J., Mahé F., Logares R., Lara E., Berney C., Le Bescot N., Probert I., Carmichael M., Poulain J., Romac S., Colin S., Aury J.-M., Bittner L., Chaffron S., Dunthorn M., Engelen S., Flegontova O., Guidi L., Horák A., Jaillon O., Lima-Mendez G., Lukeš J., Malviya S., Morard R., Mulot M., Scalco E., Siano R., Vincent F., Zingone A., Dimier C., Picheral M., Searson S., Kandels-Lewis S., Acinas S.G., Bork P., Bowler C., Gorsky G., Grimsley N., Hingamp P., Iudicone D., Not F., Ogata H., Pesant S., Raes J., Sieracki M.E., Speich S., Stemann L., Sunagawa S., Weissenbach J., Wincker P., Karsenti E., 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, 1261605.
- Díez B., Pedrós-Alió C., Massana R., 2001. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Applied and environmental microbiology* 67, 2932–2941.
- Dotzler N., Taylor T.N., Krings M., 2007. A prasinophycean alga of the genus *Cymatiosphaera* in the Early Devonian Rhynie chert. *Review of Palaeobotany and Palynology* 147, 106–111.
- Dragos N., Peterfi L., Popescu C., 1997. Comparative fine structure of pellicular cytoskeleton in *Euglena* EHRENBERG. *Archiv für Protistenkunde* 148, 277–285.

- Drummond A.J., Rambaut A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* 7, 214.
- Fawley M.W., Qin M., Yun Y., 1999. The relationship between *Pseudosourfieldia marina* and *Pycnococcus provasolii* (Prasinophyceae, Chlorophyta): evidence from 18S rDNA sequence data. *Journal of Phycology* 35, 838–843.
- Fawley M.W., Yun Y., Qin M., 2000. Phylogenetic analyses of 18S rDNA sequences reveal a new coccoid lineage of the prasinophyceae (Chlorophyta). *Journal of Phycology* 36, 387–393.
- 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.
- 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, 193–214.
- Hall T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 95–98.
- 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, 3537–3544.
- Harðardóttir S., Lundholm N., Moestrup Ø., Nielsen T.G., 2014. Description of *Pyramimonas diskoicola* sp. nov. and the importance of the flagellate *Pyramimonas* (Prasinophyceae) in Greenland sea ice during the winter–spring transition. *Polar Biology* 37, 1479–1494.
- Hori T., Moestrup Ø., Hoffman L.R., 1995. Fine structural studies on an ultraplanktonic species of *Pyramimonas*, *P. virginica* (Prasinophyceae), with a discussion of subgenera within the genus *Pyramimonas*. *European Journal of Phycology* 30, 219–234.
- 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, e33746.
- Inouye I., Hori T., Chihara M., 1990. Absolute configuration analysis of the flagellar apparatus of *Pterosperma cristatum* (Prasinophyceae) and consideration of its phylogenetic position. *Journal of Phycology* 26, 329–344.
- Janouškovec J., Horák A., Barott K.L., Rohwer F.L., Keeling P.J., 2012. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Current biology : CB* 22, 518–519.
- Jouenne F., Eikrem W., Le Gall F., Marie D., Johnsen G., Vaultot D., 2011. *Prasinoderma singularis* sp. nov. (Prasinophyceae, Chlorophyta), a solitary coccoid Prasinophyte from the South-East Pacific Ocean. *Protist* 162, 70–84.
- 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, 1133–1140.
- 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, 362–373.

- Katoh K., Misawa K., Kuma K., Miyata T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30, 3059–3066.
- 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 (Oxford, England)* 28, 1647–1649.
- \* Keeling P.J., 2010. The endosymbiotic origin, diversification and fate of plastids. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365, 729–748.
- Kim J.I., Shin W., Triemer R.E., 2010. Multigene analyses of photosynthetic euglenoids and new family, Phacaceae (Euglenales)1. *Journal of Phycology* 46, 1278–1287.
- Kolodziej K., Stoeck T., 2007. Cellular identification of a novel uncultured marine stramenopile (MAST-12 Clade) small-subunit rRNA gene sequence from a norwegian estuary by use of fluorescence in situ hybridization-scanning electron microscopy. *Applied and environmental microbiology* 73, 2718–2726.
- Kustatscher E., Dotzler N., Taylor T.N., Krings M., 2014. Microfossils with suggested affinities to the Pyramimonadales (Pyramimonadophyceae, Chlorophyta) from the Lower Devonian Rhynie Chert. *Acta Palaeobotanica* 54, 163–171.
- Latasa M., Scharek R., Gall F. Le, Guillou L., 2004. Pigment suites and taxonomic groups in Prasinophyceae. *Journal of Phycology* 40, 1149–1155.
- \* Leander B.S., 2004. Did trypanosomatid parasites have photosynthetic ancestors? *Trends in microbiology* 12, 251–258.
- Leander B.S., Triemer R.E., Farmer M.A., 2001. Character evolution in heterotrophic euglenids. *European Journal of Protistology* 37, 337–356.
- \* Leliaert F., Verbruggen H., Zechman F.W., 2011. Into the deep: new discoveries at the base of the green plant phylogeny. *BioEssays : news and reviews in molecular, cellular and developmental biology* 33, 683–692.
- 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 genomics* 15, 857.
- Lewin R.A., Krienitz L., Goericke R., Takeda H., Hepperle D., 2000. *Picocystis salinarum* gen. et sp. nov. (Chlorophyta) – a new picoplanktonic green alga. *Phycologia* 39, 560–565.
- Li R., Yu C., Li Y., Lam T.-W., Yiu S.-M., Kristiansen K., Wang J., 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics (Oxford, England)* 25, 1966–1967.
- 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, 603–619.
- Marin B., 2004. Origin and fate of chloroplasts in the Euglenoida. *Protist* 155, 13–14.
- Marin B., Melkonian M., 2010. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* 161, 304–336.

- Marin B., Nowack E.C.M., Melkonian M., 2005. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* 156, 425–432.
- Marin B., Palm A., Klingberg M.A.X., 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, 99–145.
- Maruyama S., Kim E., 2013. A modern descendant of early green algal phagotrophs. *Current biology : CB* 23, 1081–1084.
- McFadden G.I., Hill D.R.A., Wetherbee R., 1986. A study of the genus *Pyramimonas* (Prasinophyceae) from south-eastern Australia. *Nordic journal of botany* 209–234.
- Miyashita H., Ikemoto H., Kurano N., Miyachi S., Chihara M., 1993. *Prasinococcus capsulatus* gen. et sp. nov., a new marine coccoid prasinophyte. *The Journal of General and Applied Microbiology* 39, 571–582.
- Moczyłowska M., Landing E., Zang W., Palacios T., 2011. Proterozoic phytoplankton and timing of Chlorophyte algae origins. *Palaeontology* 54, 721–733.
- Monfils A.K., Triemer R.E., Bellairs E.F., 2011. Characterization of paramylon morphological diversity in photosynthetic euglenoids (Euglenales, Euglenophyta). *Phycologia* 50, 156–169.
- Nakayama T., Marin B., Kranz H.D., Surek B., Huss V.A.R., Inouye I., Melkonian M., 1998. The basal position of scaly green flagellates among the green algae (Chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences. *Protist* 149, 367–380.
- Nakayama T., Suda S., Kawachi M., Inouye I., 2007. Phylogeny and ultrastructure of *Nephroselmis* and *Pseudoscourfieldia* (Chlorophyta), including the description of *Nephroselmis anterostigmatica* sp. nov. and a proposal for the Nephroselmidales ord. nov. *Phycologia* 46, 680–697.
- Parfrey L.W., Lahr D.J.G., Knoll A.H., Katz L.A., 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* 108, 13624–13629.
- Pereira M.J., Azeiteiro U.M.M., Gonçalves F., Soares A.M.V.M., 2003. Inorganic composition of the envelopes of *Trachelomonas* Ehr. (Euglenophyta). *Acta Oecologica* 24, 317–324.
- Rambaut A., 2014. FigTree v1.4.2. Dostupné z: <http://tree.bio.ed.ac.uk/software/figt>.
- Rambaut A., Suchard M.A., Xie D., Drummond A.J., 2014. Tracer v1.6. Dostupné z: <http://beast.bio.ed.ac.uk/tracer>.
- 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. *Molecular biology and evolution* 24, 956–968.
- 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, 54–62.
- Sagan L., 1967. On the origin of mitosing cells. *Journal of theoretical biology* 14, 255–274.

- Seenivasan R., Sausen N., Medlin L.K., Melkonian M., 2013. *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as “picobiliphytes”. *PLoS one* 8, e59565.
- Shalchian-Tabrizi K., Bråte J., Logares R., Klaveness D., Berney C., Jakobsen K.S., 2008. Diversification of unicellular eukaryotes: cryptomonad colonizations of marine and fresh waters inferred from revised 18S rRNA phylogeny. *Environmental microbiology* 10, 2635–2644.
- Shalchian-Tabrizi K., Reier-Røberg K., Ree D.K., Klaveness D., Bråte J., 2011. Marine-freshwater colonizations of haptophytes inferred from phylogeny of environmental 18S rDNA sequences. *The Journal of eukaryotic microbiology* 58, 315–318.
- Shi X.L., Marie D., Jardillier L., Scanlan D.J., Vaultot D., 2009. Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS one* 4, e7657.
- Schattner P., Brooks A.N., Lowe T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research* 33, 686–689.
- Stamatakis A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 30, 1312–1313.
- \* Stiller J.W., 2014. Toward an empirical framework for interpreting plastid evolution. *Journal of phycology* 50, 462–471.
- Suda S., 2004. Taxonomic characterization of *Pyramimonas aurea* sp. nov. (Prasinophyceae, Chlorophyta). *Phycologia* 43, 682–692.
- Sunagawa S., Coelho L.P., Chaffron S., Kultima J.R., Labadie K., Salazar G., Djahanschiri B., Zeller G., Mende D.R., Alberti A., Cornejo-Castillo F.M., Costea P.I., Cruaud C., D’Ovidio F., Engelen S., Ferrera I., Gasol J.M., Guidi L., Hildebrand F., Kokoszka F., Lepoivre C., Lima-Mendez G., Poulain J., Poulos B.T., Royo-Llonch M., Sarmiento H., Vieira-Silva S., Dimier C., Picheral M., Searson S., Kandels-Lewis S., Bowler C., de Vargas C., Gorsky G., Grimsley N., Hingamp P., Iudicone D., Jaillon O., Not F., Ogata H., Pesant S., Speich S., Stemmann L., Sullivan M.B., Weissenbach J., Wincker P., Karsenti E., Raes J., Acinas S.G., Bork P., 2015. Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348, 1261359.
- Swale E.M.F., Belcher J.H., 1968. The external morphology of the type species of *Pyramimonas* (*P. tetra-rhynchus* Schmarida) by electron microscopy. *Proceedings of the Linnean Society of London* 179, 77–81.
- Takahashi F., Okabe Y., Nakada T., Sekimoto H., Ito M., Kataoka H., Nozaki H., 2007. Origins of the secondary plastids of Euglenophyta and Chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene psbO 1. *Journal of Phycology* 43, 1302–1309.
- Tao T., 2014. Standalone BLAST Setup for Windows PC. Dostupné z: <http://www.ncbi.nlm.nih.gov/books/NBK5>.
- 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, 4745–4752.
- Triemer R.E., 1980. Role of Golgi apparatus in mucilage production and cyst formation in *Euglena gracilis* (Euglenophyceae). *Journal of Phycology* 16, 46–52.



- Turmel M., de Cambiaire J.-C., Otis C., Lemieux C., 2016. Distinctive architecture of the chloroplast genome in the chlorodendrophycean green algae *Scherffelia dubia* and *Tetraselmis* sp. CCMP 881. *PLoS one* 11, e0148934.
- 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, 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. *Proceedings of the National Academy of Sciences of the United States of America* 96, 10248–10253.
- 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, 171–180.
- \* Vaultot D., Eikrem W., Viprey M., Moreau H., 2008. The diversity of small eukaryotic phytoplankton (< or =3 microm) in marine ecosystems. *FEMS microbiology reviews* 32, 795–820.
- 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, 1804–1822.
- 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). *The Journal of eukaryotic microbiology* 60, 214–221.
- 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, 832–843.
- Wyman S.K., Jansen R.K., Boore J.L., 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics (Oxford, England)* 20, 3252–3255.
- 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.
- Yamaguchi H., Suda S., Nakayama T., Pienaar R.N., Chihara M., Inouye I., 2011. Taxonomy of *Nephroselmis viridis* sp. nov. (Nephroselmidophyceae, Chlorophyta), a sister marine species to freshwater *N. olivacea*. *Journal of plant research* 124, 49–62.