

UNIVERZITA PALACKÉHO V OLMOUCI

DIPLOMOVÁ PRÁCE

Olomouc 2017

Bc. Marek Glombik

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra buněčné biologie a genetiky



**Celogenomové sekvenování vybraných druhů
sinic**

Diplomová práce

Marek Glombik

Studijní program: Biologie
Studijní obor: Molekulární a buněčná biologie
Forma studia: Prezenční

Olomouc 2017

Vedoucí práce: Mgr. Petr Dvořák, Ph.D.

Prohlášení

Prohlašuji, že jsem předloženou diplomovou práci vypracoval zcela samostatně s použitím uvedených literárních zdrojů.

V Olomouci dne

.....

Marek Glombik

Souhrn

Diplomová práce se zabývá celogenomovým sekvenováním a anotací 2 vybraných druhů sinic, sladkovodní sinice *Jacksonvillea apiculata* 2014/77b a epizoické nebo endozoické sinice *Synechococcus* sp. S4.

Teoretická část obsahuje poznatky o fylogenomice, genomice sinic a genome streamliningu, využití celogenomového sekvenování ve výzkumu evoluce sinic a sekundárním metabolismu.

Praktická část zahrnuje zpracování sekvenčních dat a kompletní anotaci kódujících sekvencí, repetitivních sekvencí, tRNA, analýzu ontologie genů, přítomnosti sekundárního metabolismu a zařazení do systému sinic na základě fylogenomického přístupu u obou vybraných druhů sinic. U *Synechococca* sp. S4 byla ještě navíc provedena analýza horizontálního přenosu genů, analýza přítomnosti delecí a inverzí v porovnání s nejbližším příbuzným, sinicí *Cyanobium gracile* PCC 6307 a datovací analýza. Velikost genomu *J. apiculata* 2014/77b byla stanovena na 10,7 Mb s 44% obsahem G:C bází a 10 096 protein kódujícími sekvencemi. Fylogenomickou analýzou byla *J. apiculata* zařazena do řádu Oscillatoriales, ale byla odhalena možná kontaminace kmenu jinou sinicí. Velikost genomu *Synechococca* sp. S4 byla stanovena na 2,7 Mb s obsahem G:C bází 70 % a 2 672 protein kódujícími sekvencemi. Fylogenomickou analýzou byla potvrzena předpokládaná příbuznost se sinicí *C. gracile* PCC 6307. Nebylo potvrzeno spojení redukce genomu *Synechococca* sp. S4 se symbiózou. Datovací analýzou bylo určeno stáří tohoto druhu na přibližně 90 milionů let.

Summary

This diploma thesis is focused on a whole-genome sequencing and annotation of two selected cyanobacterial species, freshwater cyanobacterium *Jacksonvillea apiculata* 2014/77b and epizoic or endozoic cyanobacteria *Synechococcus* sp. S4.

Theoretical part of this thesis contains principal knowledge of phylogenomics, cyanobacterial genomics and genome streamlining, using whole-genome sequencing to study the evolution of cyanobacteria and secondary metabolism.

Practical part of this thesis involves processing primary sequence data, complete annotation of coding regions, repetitive sequences, tRNA, gene ontology analysis, presence of secondary metabolism and an inference of cyanobacterial phylogeny of the two selected cyanobacterial species by the phylogenomic approach. This part also contains the analysis of horizontal gene transfer, presence of deletions and inversions of *Synechococcus* sp. S4 by comparison with its closest neighbour *Cyanobium gracile* PCC 6307 and the divergence time analysis. The size of the genome of *J. apiculata* 2014/77b is 10,7 Mb with the G:C content of 44 % and 10 096 coding sequences. Phylogenomic analysis pinpointed the location of *J. apiculata* to an Oscillatoriales order but a possible contamination by a different cyanobacteria in the culture has been found. The size of the genome of *Synechococcus* sp. S4 is 2,7 Mb with the G:C content of 70 % and 2 672 coding sequences. Phylogenomic analysis has confirmed its close relationship with *C. gracile* PCC 6307. A connection between the genome reduction of *Synechococcus* sp. S4 and symbiosis hasn't been found. Divergence time analysis dated the origin of this species to approximately 90 million years ago.

Poděkování

Rád bych touto cestou poděkoval panu Mgr. Petru Dvořákovi, Ph.D. za odborné vedení diplomové práce, věnovaný čas, ochotu a velmi cenné rady při vyhodnocování výsledků a lekce bioinformatiky. Dále bych rád poděkoval vedení algologické laboratoře Univerzity Palackého v Olomouci za možnost konání své diplomové práce na jejich pracovišti a pracovníkům Ústavu experimentální botaniky AV ČR za sekvenování vybraných genomů sinic pro tuto práci, na jejich odborném pracovišti.

Obsah

1	Úvod.....	9
2	Současný stav řešené problematiky.....	10
2.1	Fylogenomika	10
2.2	Genomika sinic a genome streamlining.....	16
2.3	Využití celogenomového sekvenování v evoluci sinic.....	19
2.4	Sekundární metabolismus sinic	29
3	Cíl práce	30
4	Materiál a metody.....	31
4.1	Biologický materiál.....	31
4.2	Sekvenování DNA	31
4.3	Bioinformatická analýza	31
4.3.1	Zpracování primárních sekvenčních dat	31
4.3.2	Anotace kódujících oblastí.....	32
4.3.3	Anotace repetitivních sekvencí	33
4.3.4	Anotace tRNA.....	33
4.3.5	Analýza přítomnosti genů sekundárního metabolismu.....	33
4.3.6	Fylogenomická analýza.....	33
4.3.7	Analýza horizontálního přenosu genů.....	35
4.3.8	Analýza přítomnosti delecí, inverzí	35
4.3.9	Ontologie genů (GO)	35
4.3.10	Datování fylogeneze <i>Synechococcus</i> sp. S4	36
5	Výsledky.....	38
5.1	<i>Jacksonvillea apiculata</i>	38
5.1.1	Obecné informace o genomu	38
5.1.2	Anotace repetitivních sekvencí	39
5.1.3	Anotace tRNA.....	41

5.1.4	Ontologie genů	42
5.1.5	Analýza přítomnosti genů sekundárního metabolismu	44
5.1.6	Fylogenomická analýza.....	47
5.2	Synechococcus sp. S4	49
5.2.1	Obecné informace o genomu	49
5.2.2	Anotace repetitivních sekvencí	50
5.2.3	Anotace tRNA.....	52
5.2.4	Ontologie genů	53
5.2.5	Analýza přítomnosti genů sekundárního metabolismu	54
5.2.6	Analýza horizontálního přenosu genů.....	55
5.2.7	Analýza přítomnosti delecí, inverzí	57
5.2.8	Fylogenomická analýza.....	60
5.2.9	Datování fylogeneze Synechococcus sp. S4	61
6	Diskuze.....	63
6.1	Jacksonvillea apiculata	63
6.2	Synechococcus sp. S4	66
7	Závěr.....	71
8	Literatura	73
9	Seznam použitých zkratk a symbolů	85
10	Přílohy	89

1 Úvod

Fylogenomika je nástrojem sloužícím k určení evolučních vztahů mezi organismy. Vychází z fylogenetiky a oproti ní pracuje s velkým množstvím genů, čímž podává věrohodnější výsledky (Philippe *et al.*, 2005; Philippe *et al.*, 2011). I fylogenomika má ovšem svá omezení. Při analýzách neodpadají systematické chyby, které mohou vést ke vzniku silně podpořených, ale nesprávně zkonstruovaných fylogenomických stromů (Jeffroy *et al.*, 2006; Wang *et al.*, 2015). V dnešní době lze tyto chyby odhalit a dosáhnout tak kvalitních výsledků analýzy (Huson *et al.*, 2006; Wägele *et al.*, 2007).

Sinice jsou organismy oplývající velmi širokou ekologickou a morfologickou diverzitou. Dříve byly sinice rozdělovány do 5 skupin právě na základě morfologických znaků (Rippka *et al.*, 1979), toto rozdělení ovšem s přibývajícím počtem nově nalezených a osekvenovaných druhů nepodává přesné informace o jejich fylogenezi. Každý nově představený genom může vést ke změně fylogenetického stromu, a proto je důležité k odvození fylogeneze používání spolehlivých molekulárních markerů (Gupta *et al.*, 2010), jako jsou třeba proteiny typické pro určité druhy (Mulikidjanian *et al.*, 2006). Momentálně se předpokládá nejstarší výskyt sinic před 3,5 miliardami let (Schopf, 2000). Jsou to organismy, které výrazně přispěly k nárůstu hladiny atmosférického kyslíku na Zemi v období, které se označuje jako „Great Oxidation Event“ (velké okysličení, GOE), a ke kterému došlo asi před 2,4 miliardami let (Bekker *et al.*, 2004; Anbar *et al.*, 2007). Jedním z evolučně nejúspěšnějších rodů sinic je rod *Synechococcus* (Dvořák *et al.*, 2014). Ten patří k nejrozšířenějším a nejvýznamnějším primárním producentům (Flombaum *et al.*, 2013) a jeho původ se datuje do doby před 3 miliardami let (Dvořák *et al.*, 2014).

2 Současný stav řešené problematiky

2.1 Fylogenomika

Nejběžnějším nástrojem, který slouží k určení evolučních vztahů mezi všemi organismy je fylogenetika. Tento nástroj je klíčový pro pochopení evoluce života na Zemi (Philippe *et al.*, 2011). Tradičně se ke studiu fylogeneze používaly fenotypové znaky organismů a s rozvojem sekvenačních technologií se začaly používat nukleotidové a aminokyselinové sekvence (Snel *et al.*, 2005). Nejčastěji používané geny pro studium fylogeneze jsou mitochondriální geny, a to díky jejich univerzálnímu výskytu u eukaryot a prokaryot a evoluční konzervativnosti (Woese *et al.*, 1990), protože mitochondrie vznikla endosymbiózou prokaryotní a eukaryotní buňky (Andersson *et al.*, 2003). Dříve byly tyto geny vysoce využívány ke klasifikaci mikroorganismů a pomohly rozlišit třetí říši života, Archaea (Woese *et al.*, 1977). Nicméně fylogeneze založená na jednotlivých genech vedla často ke konfliktním výsledkům (Jeffroy *et al.*, 2006) a určité evoluční analýzy je možno provádět pouze analýzou celého genomu, například chceme-li porovnat ztrátu genů mezi určitými druhy organismů (Eisen *et al.*, 2003).

Novější oblastí vycházející z fylogenetiky se stala fylogenomika, která k porovnávání evolučních vztahů používá velká množství genů (Philippe *et al.*, 2011). Její vznik a vývoj byl umožněn rychlým rozvojem sekvenačních technologií, díky nimž máme možnost zjistit kompletní sekvence DNA organismů. Hlavními zájmy fylogenomiky jsou studia vztahů mezi organismy na základě molekulárních dat a studia evoluce na molekulární úrovni podle známých informací o evoluční historii organismů. Díky tomu, že fylogenomika používá k práci velké množství genů, odpadá stochastická chyba a celá analýza podává statisticky věrohodnější výsledky (Philippe *et al.*, 2005; Jeffroy *et al.*, 2006). I to vedlo k názoru, že fylogenomické analýzy odstraní nesrovnalosti ve fylogenezi založené na jednotlivých genech (Gee, 2003; viz níže).

Mezi metody konstrukce fylogenetických stromů, které byly převzaty dále do fylogenomiky patří:

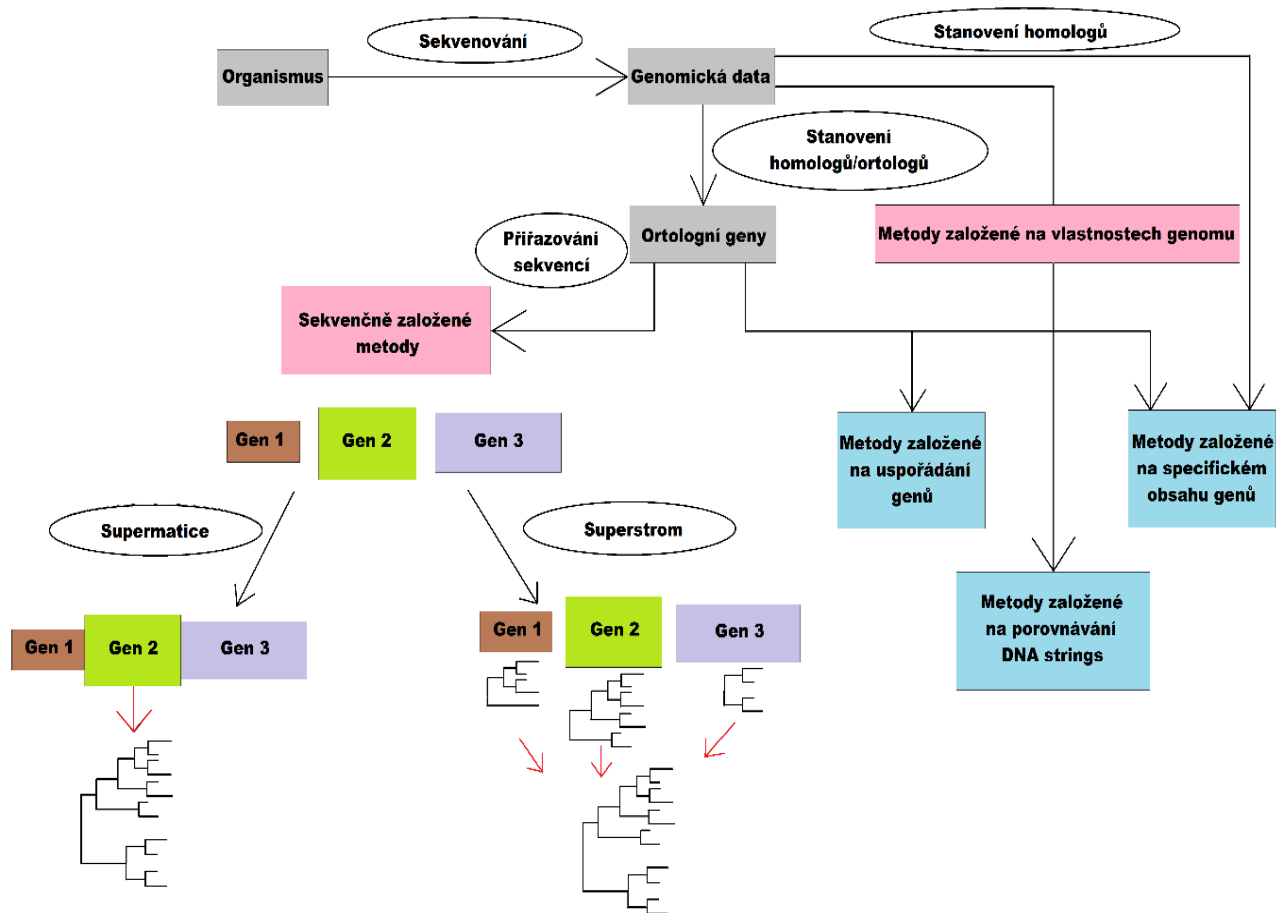
- 1) Distanční metoda – fylogenetický strom je zde odvozen z distanční matice, která reprezentuje evoluční vzdálenost mezi dvojicemi taxonů.

- 2) Metoda maximální parsimonie (maximum parsimony) – touto metodou je konstruováno mnoho fylogenetických stromů a vybrán je ten, který odpovídá nejmenšímu množství evolučních událostí, nutných k jeho konstrukci (Delsuc *et al.*, 2005).
- 3) Metoda maximální věrohodnosti (maximum likelihood) – základem je funkce, kterou je vypočítána pravděpodobnost, s jakou dochází ke konstrukci daných fylogenetických stromů. Následně je vybrán ten, jehož pravděpodobnost vzniku je nejvyšší (Felsenstein, 1981).

Zmíněné metody konstrukce fylogenetického stromu se volí podle přístupu k samotné analýze. Podle metod, které používáme k analýze určitých charakteristik genomu je dělíme na sekvenčně založené metody a metody založené na vlastnostech genomu. Postup u sekvenčně založených metod začíná vzájemným porovnáním sekvencí, odvozením ortologních genů a následnou konstrukcí fylogenetického stromu na základě výsledku mnohočetného přiřazování sekvencí (Delsuc *et al.*, 2005). Mnohočetné přiřazování má však své limity. Občas může docházet k nejasným či nedostatečným výsledkům při výskytu strukturních variací v genech, u kterých čekáme, že si zachovávají relativní jednotnost v průběhu evoluce (Chan *et al.*, 2013). K determinaci a následnému odstranění nevěrohodně přiřazených sekvencí se používá například program Gblocks (Castresana, 1999). V dalším kroku lze použít 2 různé postupy k rekonstrukci fylogeneze. Prvním je tvorba a analýza „supermatice“ vzniklé spojením jednotlivých genů, pro kterou je vhodné využití metod konstrukce fylogenetického stromu založené na věrohodnosti. Druhým postupem je tvorba „superstromu“ individuální analýzou jednotlivých genů a následnou kombinací výsledných fylogenetických stromů do jednoho (Delsuc *et al.*, 2005). Pro tento postup se využívá metoda reprezentace matice parsimonií (MRP – Matrix Representation using Parsimony) (Bininda-Emonds, 2004), která ale může mít problém se zařazením rychle se vyvíjejících linií (Philippe *et al.*, 2005). Nověji se ke konstrukci superstromu používá metoda maximální věrohodnosti (maximum likelihood) (Steel *et Rodrigo*, 2008). Ze statistického hlediska ovšem není ani jeden ze zmíněných postupů zcela ideální. Metoda s využitím supermatice ignoruje rozdíly mezi geny co do frekvence substitučních změn, složení bází, či dalších aspektů evoluce. Metoda superstromu zase může různě nadhodnocovat data, vlivem nezávislého nastavení parametrů pro každý gen (Ren *et al.*, 2008). Metodami založenými na vlastnostech genomu se zkoumá uspořádání genů, specifický obsah genů v genomu či distribuce oligonukleotidů v genomu,

tzv. DNA strings. První dvě zkoumané vlastnosti jsou nezávislé na mnohočetném přiřazování sekvencí, ale stále vyžadují určení homologů či ortologů.

Metody používané ke konstrukci fylogenetických stromů na základě uspořádání genů mohou být distančně (Luo *et al.*, 2009) či parsimonně založené (Wolf *et al.*, 2001). Metody u tohoto typu analýzy jsou výpočetně náročnější než např. sekvenčně založené metody. Je nutno vypočítat minimální počet událostí nutných ke změně jednoho uspořádání genů v druhé (Wang *et al.*, 2006). Obecně se za tyto události počítají inverze či transpozice v genomu (Moret *et al.*, 2001). Ke konstrukci fylogenetických stromů na základě specifického obsahu genů se používají distanční metody, kde je vzdálenost reprezentována proporcí sdílených ortologních genů mezi genomy a metody parsimonie, které hodnotí přítomnost či absenci homologů nebo ortologů. Metoda DNA strings je založena na frekvenci výskytu krátkých oligonukleotidů, jejichž kombinace je pro genomy charakteristická (Delsuc *et al.*, 2005). Ke konstrukci fylogenetických stromů se používá distanční metoda (Edwards *et al.*, 2002). K podpoře určitých uzlů se při fylogenetické analýze dá také využít dalších charakteristik genomů jako jsou pozice intronů, integrace retrovirozů či fúze a štěpení genů (Delsuc *et al.*, 2005). Shrnutí výše zmíněných metod odvození fylogeneze je vyobrazeno na Obr. 1.

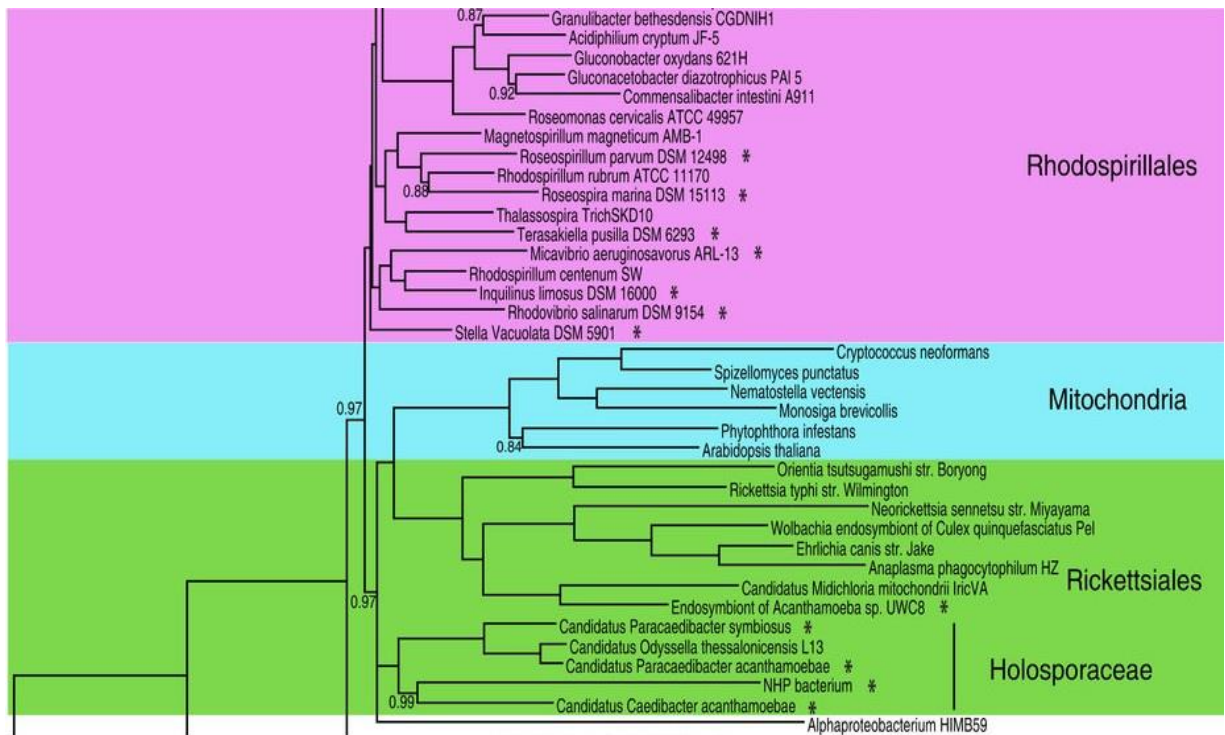


Obr. 1: Metody k odvození fylogeneze při fylogenomické analýze. Předloha podle (Delsuc *et al.*, 2005)

Jak již bylo zmíněno výše, s rozvojem fylogenomiky se očekávalo, že se vyřeší nejasnosti ve fylogenetických stromech založených na analýze jednoho genu. Nicméně bylo zjištěno, že i když vymizí stochastická chyba tak zůstávají systematické chyby jako sekvenční tahy a přitažlivost dlouhých větví (Long Branch Attraction, dále jen LBA), což znamená, že jsou chybně přiřazeny jinak vzdálené linie jako příbuzné. Z důvodu takovýchto chyb pak vznikají nesprávné fylogenetické stromy se silnou podporou (Jeffroy *et al.*, 2006; Wang *et Wu*, 2015).

Řešením takového problému se zabývali autoři Wang *et Wu* (2015) ve své práci o využití fylogenomického přístupu k určení přesné pozice mitochondrií na stromu života. Již starší fylogenetické studie naznačovaly, že původce mitochondrií pochází ze skupiny alfa proteobakterií (Lang *et al.*, 1999). Jejich přesné určení nebylo možné právě z faktorů, jako jsou LBA, sekvenční tahy a slabý fylogenetický signál z důvodu saturace mnoha informativních míst sekvencí DNA mutacemi. V dřívějších studiích byly mitochondrie zařazeny jako nejbližší příbuzné k rodu *Rickettsia* (Emelyanov, 2003) či k bakterii

Rhodospirillum rubrum (Esser *et al.*, 2004). Z důvodu takovýchto nejasných výsledků autoři provedli analýzy „NeighborNet“ v programu SplitsTree (Huson *et Bryant*, 2006) a spektrální analýzu metodami „Split Analyses“ (Wagele *et Mayer*, 2007), které jim pomohly odhalit přítomnost systematických chyb v tehdejší současném datasetu potřebném k fylogenomické analýze. Protože většina genomů, které byly v datasetu, byla osekvenována z ekonomických či lékařských důvodů, chyběla zde dostatečná fylogenetická diverzita, což také přispívá k LBA. Autoři proto osekvenovali celkem 18 nových genomů z 6 bakteriálních druhů, přičemž celkem 9 z těchto osekvenovaných druhů patřilo do řádů Rickettsiales a Rhodospirillales z důvodu jejich dřívějšího přidružení k mitochondriím v jiných pracích. Po přidání těchto nových genomů do datasetu se systematická chyba prokazatelně výrazně snížila. Pro finální fylogenomickou analýzu byly z mitochondriálních genů různých organismů využity z nich odvozené jaderné geny, které vznikly laterálním přenosem z mitochondrií. Bylo prokázáno, že oproti genům kódovaným mitochondriemi mají ty jaderné nižší frekvenci substitučních mutací v sekvenci, podléhají méně sekvenčním tahům a produkují fylogenetický strom v souladu se vzorovým uspořádáním genů, a tudíž jsou vhodnými markery k odvození původu mitochondrií na stromu života. Na Obr. 2 lze vidět finální Bayesiánský konsenzuální fylogenomický strom, k jehož konstrukci byl použit soubor přiřazených proteinových sekvencí složený z 6 201 aminokyselin. Jako vnější skupina byl použit soubor Beta a Gammaproteobakterií. Vnitřní uzly, u kterých není jinak specifikovaná hodnota podpory, měly hodnotu podpory 1.0 (tj. nejvyšší možná podpora).



Obr. 2: Bayesiánský konsenzuální fylogenomický strom sestřený z datasetu 72 Alfaproteobakterií a 6 eukaryotických organismů, zacílený na mitochondrie a nejbližší vývojové větve. Převzato z Wang *et* Wu (2015).

Výsledný fylogenomický strom ukázal, že nejbližšími příbuznými mitochondrií jsou čeledi bakterií Rickettsiaceae a Anaplasmataceae. To by mohlo znamenat, že původce mitochondrií byl s největší pravděpodobností endosymbiontem bakterií řádu Rickettsiales. Zřejmě ovšem nemusel být obligátním symbiontem a díky tomu mohl opustit hostitele a dát původ obligátním vnitrobuněčným liniím Rickettsiales, jak je známe dnes. Na rozdíl od předchozích studií byly zde mitochondrie umístěny pod řád Rickettsiales, a ne pouze jako jeho sesterská vývojová větev. Nejhluběji umístěnou větev v řádu Rickettsiales tvoří čeleď Holosporaceae a mitochondrie vznikly až po oddělení této čeledi. Nakonec na základě vysoké podpory (0,97) vývojové větve Rickettsiales/mitochondria autoři došli k závěru, že mitochondrie se vyvinuly jako odvozená linie z řádu Rickettsiales. Jak dále autoři uvádějí, velký význam v dalších fylogenetických analýzách o určení přesného původu mitochondrií bude mít čeleď Holosporaceae, jakožto nejbližší vnější skupina k vývojové větvi mitochondria/Rickettsiaceae/Anaplasmataceae, protože skrývá potenciál k dalšímu zvýšení přesnosti budoucích analýz.

2.2 Genomika sinic a genome streamlining

Sinice jsou velmi rozmanitou a jedinou skupinou prokaryotických organismů, která provádí oxygenní fotosyntézu (Gupta *et Mathews*, 2010). Díky autotrofnímu stylu života kolonizovaly velké množství prostředí od slaných či sladkých vod až po půdy, v rozsahu teplot polárních oblastí po extrémní teploty horských pramenů. K této jejich široké ekologické diverzitě jsou sinice také velmi rozdílné morfologicky (Schirrmeister *et al.*, 2015).

Většina sinic má cirkulární chromozóm a malý počet plazmidů, které mohou dosahovat délky až několik stovek kilobází. Vyskytují se také výjimky jako u mořských pikocyanobakterií rodu *Prochlorococcus* a *Synechococcus*, které nemají žádný plazmid. Rozdíly jsou samozřejmě také v počtech kopií chromozómů (Hess, 2011). Mořské pikocyanobakterie jsou hlavně mono či diploidní. Některé sinice mohou být také oligoploidní (3-10 kopií genomu v buňce) a v extrémních případech vysoce polyploidní jako tomu je u *Synechocystis* PCC 6803 s nejvyšším počtem až 218 kopií chromozómu v buňce (Griese *et al.*, 2011). Genom sinic je obecně jako u ostatních prokaryot tvořen z více jak 85 % kódujícími sekvencemi (Lynch, 2006). Z historického hlediska se vznik sinic předpokládá někdy v době před 3,5 mld. let (Schopf, 2000). Tak dlouhá doba evoluce mohla rozhodně přispět k široké genomové diverzitě mezi sinicemi. Velikost jejich genomů se pohybuje od 1,44 Mb u mořské sinice UCYN-A (*Candidatus Atelocyanobacterium thalassa*) (Tripp *et al.*, 2010) až po 11,6 Mb sinice *Calothrix* sp. PCC 7103 (Shih *et al.*, 2012).

Komparativní analýzou 15 kompletních genomových sekvencí různých druhů sinic bylo odhaleno celkem 3 188 proteinových rodin (klastrů ortologních skupin proteinů sinic - CyOGs) kódovaných alespoň u 3 z analyzovaných genomů sinic. Z toho celkem 892 CyOGs bylo přítomných u všech analyzovaných genomů a dalších 162 bylo přítomných ve 14 genomech. Celkem tedy bylo nalezeno 1 054 CyOGs přítomných alespoň ve 14 analyzovaných genomech, které autoři práce označili jako „core“ sadu u sinic. Zajímavé je, že pouze 50 z těchto genů bylo pro sinice unikátních a 84 genů sinice sdílely s fotosyntetickými eukaryoty (Mulikidjanian *et al.*, 2006).

Určité druhy sinic mají vysoce redukované a kompaktní genomy. Kupříkladu analýza již zmíněné sinice UCYN-A odhalila přítomnost pouhých 1 241 genů. Došlo u ní k redukci genomu, pravděpodobně za účelem vyšší efektivity metabolismu (Lynch, 2006). Tento jev se nazývá „genome streamlining“ (zjednodušení a zefektivnění genomu). Zjistilo se, že byly

eliminovány všechny geny zodpovědné za fotolýzu vody a geny centra vyvíjející kyslík, což jsou důležité součásti fotosyntetického aparátu. Dále chybí enzymy pro fixaci uhlíku, Krebsův cyklus a dráhy pro syntézu některých aminokyselin a purinu (Tripp *et al.*, 2010). V mnoha sinicích je Krebsův cyklus neúplný s chybějící α -ketoglutarát dehydrogenázou a NADH oxidázou (Smith *et al.*, 1967). Sinice UCYN-A má pouze jeden enzym Krebsova cyklu, a to malát dehydrogenázu. Genetický charakter, metabolické nedostatky této sinice, přítomnost na kyslík citlivé nitrogenázy, redukovaný genom, prozatímni neschopnost kultivace a další faktory poukazují na symbiotický způsob života této sinice. Předpokládá se, že UCYN-A je buď obligátním endosymbiontem kryptického symbiotického partnera nebo se adaptovala a získává organické živiny z organické hmoty v oceánu (Tripp *et al.*, 2010).

Typickým příkladem bakteriálního symbionta s redukovaným genomem, který by se dal porovnat s UCYN-A, je *Buchnera aphidicola*. U této bakterie také chybí geny pro enzymy Krebsova cyklu. Přítomny jsou pouze geny 2-oxoglutarátového komplexu. Dále chybí geny pro biosyntézu ubichinonů a geny zodpovědné za fermentaci a anareobní respiraci. Navíc je *B. aphidicola* omezena v opravě DNA a rekombinaci. Nemá gen *recA*, který je nejdůležitější komponentou pro homologní rekombinaci. Další geny důležité k SOS opravě, které chybí, jsou *lexA*, *umuCD*, *uvrABC*, což činí tuto bakterii velmi zranitelnou. Dále chybí geny potřebné pro metylaci a restrikci DNA. Chybí téměř všechny geny pro tvorbu fosfolipidů a lipopolysacharidů. V neposlední řadě nejsou v genomu *B. aphidicola* přítomny geny řídicí expresi při změně vnějších podmínek, chybí všechny regulátory transkripce kromě *dnaA*, a také chybí gen pro adenylát cyklázu *cyaA* a gen *crp* pro cAMP receptor (Shigenobu *et al.*, 2000). Zachovány zůstaly například geny pro biosyntézu aminokyselin (Baumann *et al.*, 1997). U symbionta *Wigglesworthia glossinidia* se gen *dnaA* naopak vůbec nenachází (Akman *et al.*, 2002) a u symbiotické sinice *Nostoc Azollae* došlo k pseudogenizaci *dnaA* genu. Dalšími dotčenými geny jsou replikační geny *dnaX*, *dnaN*, *holB*, *ssb*, geny pro opravu DNA *recD*, *alkA* a gen *recQ* kódující DNA helikázu (Ran *et al.*, 2010).

Známým příkladem sinice s redukovaným genomem je *Prochlorococcus marinus*. U kmenů MED4 a SS120 této sinice nebyly nalezeny geny *ada* kódující 6-o-metylguanin-DNA metyltransferázu, která slouží k opravě alkylovaného guaninu a tyminu, dále gen *recQ* kódující DNA helikázu, a nakonec gen kódující enzym z NUDIX rodiny hydroláz. Samotnému *P. marinus* MED4 navíc schází gen *mutY* pro A/G-specifickou DNA glykolázu, gen *recJ* pro exonukleázu specifickou pro jednovláknovou DNA a geny

exol/xseA a *xseB* kódující podjednotky Exonukleázy VII (Dufresne *et al.*, 2005). Co se týče genů zapojených do fotosyntézy, *P. marinus* SS120 schází gen *psbU* kódující podjednotku fotosystému II sloužící k ochraně kyslík vyvíjejícího komplexu před inaktivací teplem, a gen *psbV* kódující cytochrom *c* asociovaný s fotosystémem II. Na druhou stranu má SS120 8 kopií *pcb* genu kódující proteiny světloběrného komplexu oproti MED4, která má jen 1 kopii tohoto genu. *P. marinus* SS120 má také neúplný Krebsův cyklus. Chybí geny *sucA* a *sucB* kódující E1 (dehydrogenáza) a E2 (dihydrolipoamid sukcinyl transferáza) komponentu 2-oxoglutarát dehydrogenázového komplexu, dále schází geny pro obě podjednotky sukcinyl-CoA syntetázy a NAD-dependentní malát dehydrogenázy (Dufresne *et al.*, 2003).

Mnoho organismů s redukováným genomem má také velmi nízký poměr G:C k A:T bázím (Ussery *et Hallin*, 2004). To a další jevy, jako malý počet pseudogenů a nízké procento nekódujících sekvencí, se považují za typické pro genome streamlining (Giovannoni *et al.*, 2014). Jako příklad lze uvést střevního komenzála *Lactobacillus johnsonii* kmen NCC 533, který má procentuální zastoupení A:T bází 65 %, či parazit *Mycoplasma mycoides* subsp. *mycoides*, u kterého hladina A:T bází dosahuje až 76 % (Ussery *et Hallin*, 2004). Zástupce *Buchnera aphidicola* je na tom podobně, procentuální zastoupení A:T bází má 74 %. (Tamas *et al.*, 2002). Ze zástupců sinic lze uvést *Prochlorococcus marinus* SS120 u něhož je zastoupení A:T bází 64 %.

Je velice zajímavé, že genomy sinic rodu *Prochlorococcus* vykazují podobnosti s parazitickými či symbiotickými prokaryoty ve smyslu redukce genomu, zmenšování obsahu G:C bází a ztráty genů pro reparaci DNA, protože jde o volně žijící sinice (Dufresne *et al.*, 2003). Předpokládá se, že existence tohoto fenoménu není z důvodu genetického driftu či relaxované selekčního tlaku. Je možné, že pro život v oceánu, kde může být nedostatek živin poskytují redukované genomy jakousi selekční výhodu. Funkční analýza ztracených genů neukázala vztah mezi jejich ztrátou a specializací pro určitou ekologickou niku. Bylo ovšem zjištěno, že z genomu *Prochlorococca* se neztrácí jen neesenciální geny. Chybí totiž geny důležité pro ochranu před oxidativním stresem. Je možné, že zde došlo k uplatnění hypotézy Černé Královny (Morris *et al.*, 2012), kdy ztráta těchto genů nebyla pro *Prochlorococca* fatální, protože se ve stejném prostředí nachází organismus jako např. sinice rodu *Synechococcus*, které tyto geny mají a zvládají produkovat dostatečné množství enzymů snižující reaktivní oxidativní částice na úroveň bezpečnou pro okolní organismy (Sun *et Blanchard*, 2014). Je nutné ovšem dodat, že tento vztah mezi *Prochlorococem* a

Synechococem je mutuální a tyto 2 rody spolu spolupracují a sdílí produkty metabolismu ve svém životním prostředí (Moore *et al.*, 2002).

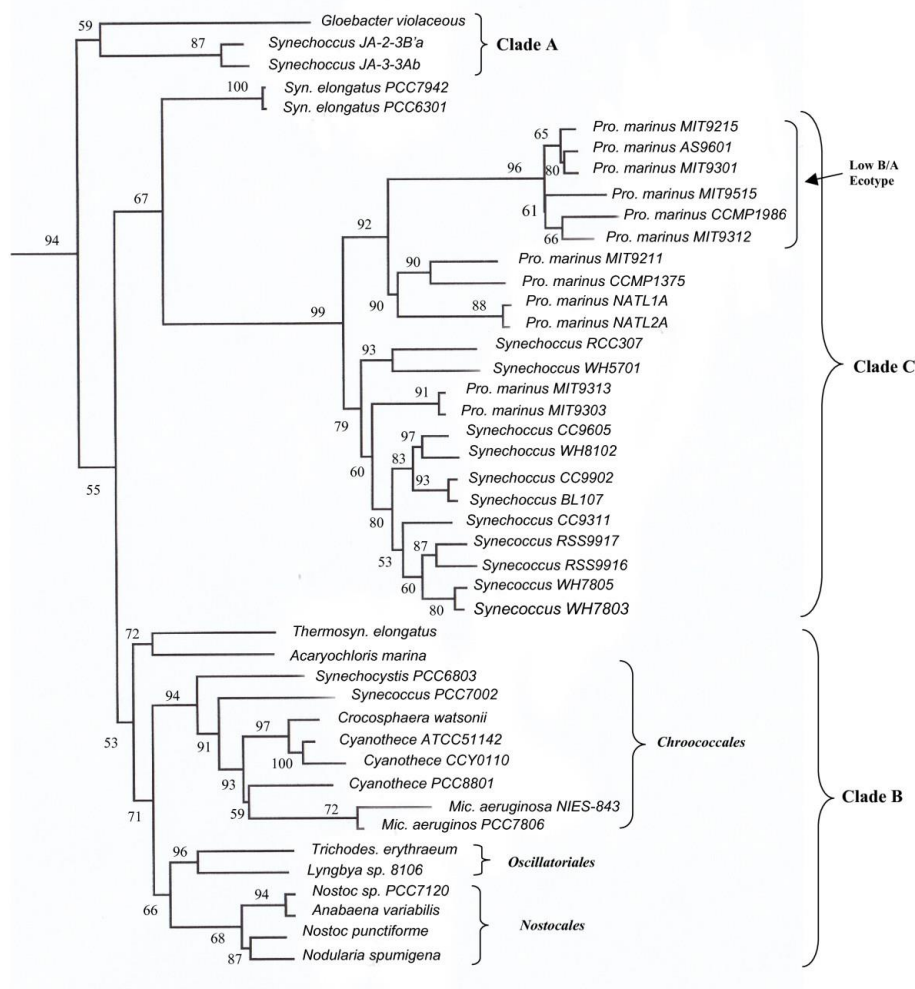
2.3 Využití celogenomového sekvenování v evoluci sinic

Jak již bylo zmíněno výše, sinice oplývají širokou ekologickou a morfologickou diverzitou. Na základě morfologie byly již dříve sinice taxonomicky rozděleny do 5 oddílů. Oddíly I a II zahrnují jednobuněčné druhy, oddíly III-V mnohobuněčné druhy. Sinice v oddílech IV a V jsou schopny tvorby diferenciovaných typů buněk jako jsou heterocysty pro fixaci dusíku, akinety pro přežití nepříznivých podmínek a další (Rippka *et al.*, 1979; Shih *et al.*, 2013).

Fylogeneze založená pouze na morfologii není samozřejmě přesná a až rozvoj sekvenačních metod umožnil jasnější porozumění fylogenezi sinic. Příkladem dřívější studie je např. ta od Sanchez-Barcaldo *et al.* (2005), ve které prováděli fylogenetickou analýzu 14 druhů sinic na základě 34 proteinů. S dalším pokrokem a rozvojem sekvenování se postupně prováděly analýzy se stále větším množstvím sekvencí a organismů. Swingley *et al.* (2008) již prováděl fylogenomickou analýzu 24 druhů sinic na základě 583 ortologních proteinů.

S každou introdukcí nově osekvenovaného genomu může docházet ke změnám v uspořádání fylogenomických stromů. Je tedy důležité identifikovat a používat spolehlivé markery, které odpovídají výsledkům současných fylogenomických analýz a zároveň umožní definitivní vymezení různých vývojových větví. Příkladem takovýchto užitečných markerů jsou indely, což jsou konzervované inserce a delece v široce rozmístěných proteinech. Ty mohou být typické pro určitý kmen či jeho hlavní podskupiny (Gupta *et al.*, 2003; Gupta *et Mathews*, 2010). Druhým typem vhodných markerů jsou proteiny unikátní pro určité druhy organismů, v tomto případě pro sinice (Mulkidjanian *et al.*, 2006).

Gupta *et Mathews* (2010) se na rozdíl od dřívějších studií soustředili nejen na hledání proteinů typických pro všechny či většinu sinic, ale také na hledání proteinů typických pro některé hlavní skupiny sinic. Na základě přiřazení 44 proteinových sekvencí ze 44 druhů sinic pak sestavili fylogenetický strom metodou maximum likelihood. Sinice v tomto stromu rozdělili do 3 vývojových větví (clade) A, B a C. (viz Obr. 3).



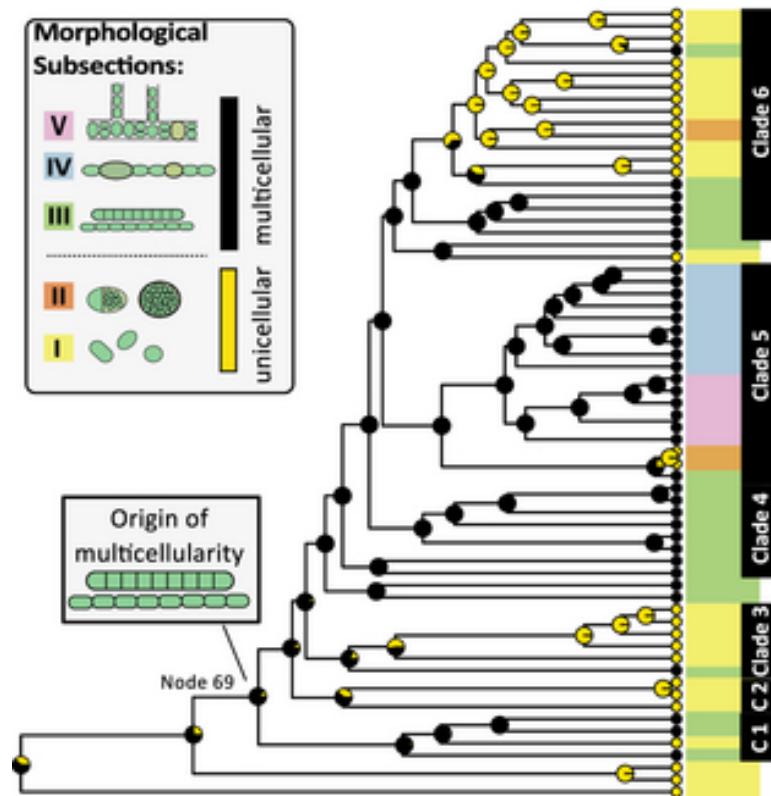
Obr. 3: Maximum-likelihood fylogenetický strom vybraných druhů sinic založený na přiřazení 44 proteinů. Vývojové větve jsou označeny A-C. Čísla u jednotlivých uzlů znamenají procentuální vyjádření podpory metodou quartet puzzling, měřítko vyjadřuje 0,1 substitucí na místo v sekvenci a low B/A ecotype odkazuje na *Prochlorococcus* spp. obsahující nižší poměr chlorofylu b/a2 jako adaptaci k růstu za jiných světelných podmínek. Převzato Gupta *et* Mathews (2010)

Za pomoci programu BLASTP byly poté identifikovány proteiny společné pro všechny sinice v tomto stromu a v jednotlivých vývojových větvích. Proteinů společných pro všechny sinice bylo nalezeno 39, s tím, že 33 z nich se nacházelo u všech sinic a zbývajících 6 chybělo u 1-2 druhů. Pro vývojovou větev A bylo nalezeno 14 specifických proteinů, pro vývojovou větev B 38 specifických proteinů a pro vývojovou větev C 60 specifických proteinů. I přesto, že se z velké části jednalo pouze o hypotetické proteiny s blíže neurčenou funkcí, jejich přítomnost v různých vývojových větvích indikuje, že se geny pro tyto proteiny vyvinuly u společného předka těchto vývojových větví a následně se udržely pouze u některých potomků. Díky tomu se mohou stát významnými markery pro tyto větve a

společně s indely mohou pomoci vyřešit přesnější vymezení druhů v rámci fylogenetických stromů. Dále se takovéto unikátní proteiny dají využít k určení kompozice v metagenomických analýzách, jako to provedli von Mering *et al.* (2007).

Jak bylo zmíněno výše, nejstarší výskyt sinic se podle předpokladů datuje do doby někdy před 3,5 miliardou let (Schopf, 2000). Jsou to jediné organismy, u kterých došlo k vývinu oxygenní fotosyntézy. Ostatní organismy jako řasy a rostliny tuto vlastnost získaly díky endosymbióze se sinicemi, která vyústila ve vznik chloroplastů (Sagan, 1967; Schirrmeyer *et al.*, 2015). Sinice výrazně přispěly k nárůstu hladiny atmosférického kyslíku na Zemi. Toto období se nazývá Great Oxidation Event a došlo k němu asi před 2,4 miliardou let (Bekker *et al.*, 2004; Anbar *et al.*, 2007). S GOE je také spojován vznik mnohobuněčnosti u sinic (Schirrmeyer *et al.*, 2013).

V práci Schirrmeyera *et al.* 2015 autoři studovali evoluci sinic kombinací fylogenomiky a fosilních záznamů. Metodou maximum likelihood zkonstruovali fylogenomický strom pro 65 taxonů založený na 756 identifikovaných proteinech. Tento strom porovnali s jiným maximum likelihood stromem, tentokrát založeným na sekvencích 16S a 23S rRNA pro stejný počet taxonů. Zatímco bootstrapová analýza u stromu založeného na ribozomální RNA poskytla 100% podporu pouze pro 24 ze 64 větví, u stromu založeného na 756 proteinech poskytla 100% podporu u 57 ze 64 větví. Na zkonstruovaném fylogenomickém stromu autoři identifikovali 6 hlavních vývojových větví. Následně provedli analýzu stavu znaku k odvození evoluce mnohobuněčnosti pomocí bayesiánských i maximum likelihood metod. Výsledek těchto analýz ukázal, že v průběhu evoluce byla mnohobuněčnost alespoň 5x ztracena a 1 nebo 2x znovu získána u některých linií. Dále se ukázalo, že většina mladších sinic jsou potomky staré mnohobuněčné linie. Analýzy naznačují, že ke vzniku mnohobuněčnosti došlo na uzlu 69 (viz Obr. 4).



Obr. 4: Rekonstrukce stavu znaku k odvození evoluce mnohobuněčnosti 64 taxonů sinic metodou maximum likelihood za předpokladu nesymetrických změn. Koláčové grafy na jednotlivých uzlech zobrazují zrekonstruované stavy znaku, kdy černá barva označuje mnohobuněčnost a žlutá jednobuněčnost. Taxony jsou představovány různými barvami na základě jejich zařazení do jednoho z 5 typů oddílů podle Rippka *et al.* (1979). Převzato z Schirmermeister *et al.* (2015).

Po určení místa vzniku mnohobuněčnosti bylo třeba určit pomocí datování, kdy k tomuto jevu došlo. Zde byly testovány 2 hypotézy, a to, zda došlo ke vzniku mnohobuněčnosti před nebo až po GOE. Ke kalibraci pro 4 analýzy (2 pro každou hypotézu) byly použity 3 body:

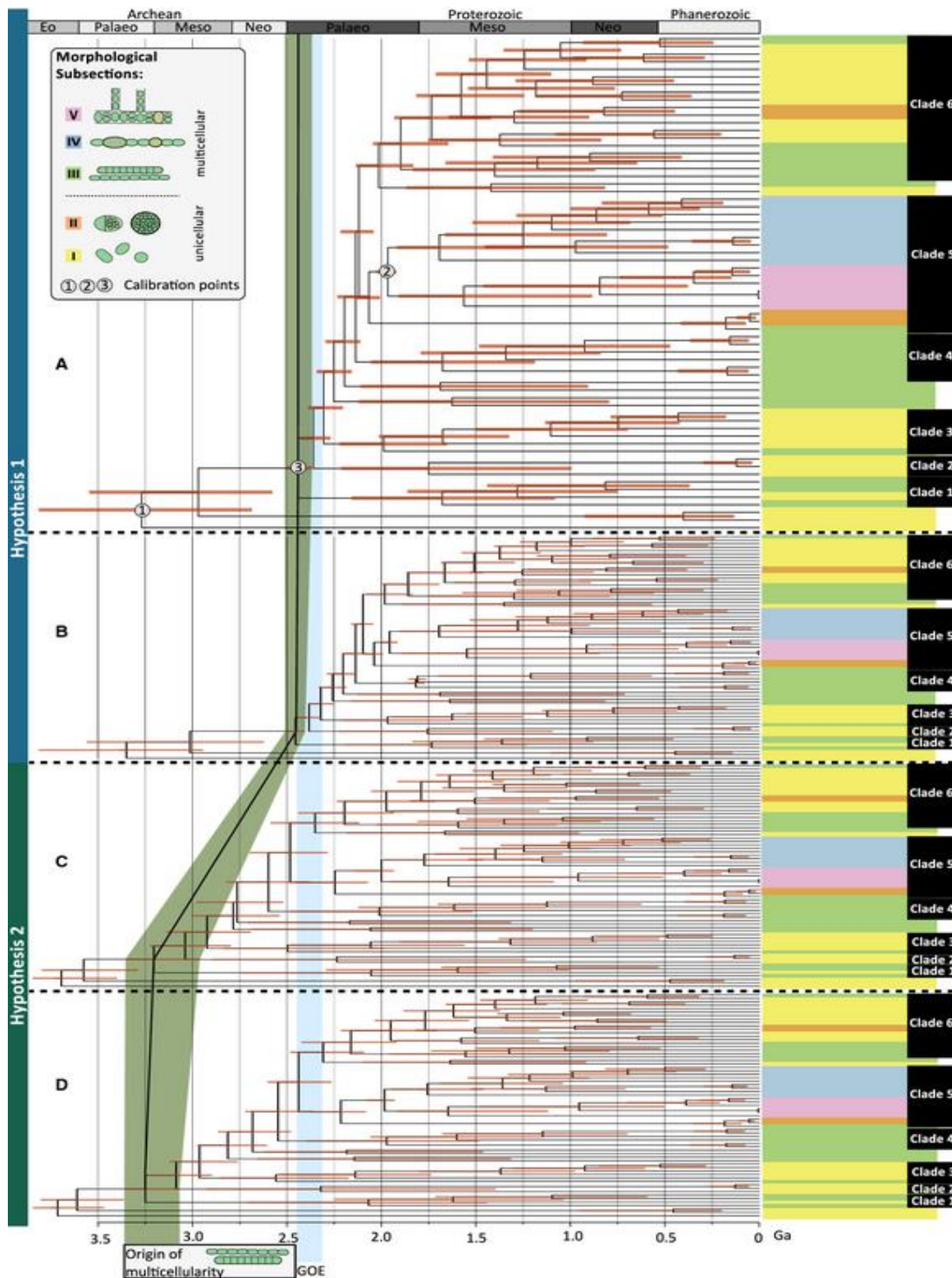
C1 – vznik sinic – Spodní hranice pro vznik byla stanovena na 3,85 mld. let pro všechny analýzy. Dále pro 2 analýzy byla stanovena vrchní hranice na 2,45 mld. let, což předpokládá vznik před GOE a pro další 2 analýzy byla stanovena vrchní hranice na 2,918 mld. let, což je doba před prvním nárůstem volného kyslíku v ovzduší.

C2 – vznik terminální diference buněk – Tato kalibrace byla založena na fosilních nálezech akinet, které jsou přítomny pouze u sinic schopné tvorby heterocystů, tudíž sinice oddílů IV a V podle Rippka *et al.* (1979). Spodní hranice byla stanovena na 2,32 mld. let za předpokladu, že existence heterocyst mohla poskytnout evoluční výhodu až po GOE,

vrchní hranice byla stanovena na 1,957 mld. let, což odpovídá nejstarším nálezům fosilních akinet.

C3 – evoluce mnohobuněčnosti – první hypotéza předpokládala vznik mnohobuněčnosti před GOE. Zde byla spodní hranice pro kalibraci stanovena na 3,33 mld. let představující nejstarší fosilní nálezy a vrchní hranice byla stanovena na 2,45 mld. let představující GOE. Druhá hypotéza naopak předpokládala vznik mnohobuněčnosti až po GOE. Zde byla tedy spodní hranice pro kalibraci stanovena na 2,45 mld. let a vrchní hranice na 1,957 mld. let.

Výsledek všech 4 analýz datoval vznik mnohobuněčnosti před GOE. Na Obr. 5 jsou 4 datované fylogenomické stromy z předchozí analýzy (viz Obr. 4), 2 pro každou hypotézu.



Obr. 5: Datování fylogeneze sinic pomocí molekulárních hodin. Jednotlivé body pro kalibraci jsou vyznačeny čísly ve vrchním fylogenomickém stromu. Barvy označují různé oddíly sinic podle rozdělení Rippka *et al.* (1979). Vývojové větve jsou vyznačeny v černém rámečku v pravé části obrázku. Místo vzniku mnohobuněčnosti je ve fylogenomických stromech vyznačeno zeleným pruhem, místo GOE je vyznačeno světle modrým pruhem. Geologická období jsou vyznačena ve vrchní části obrázku. Převzato z Schirrmester *et al.* (2015).

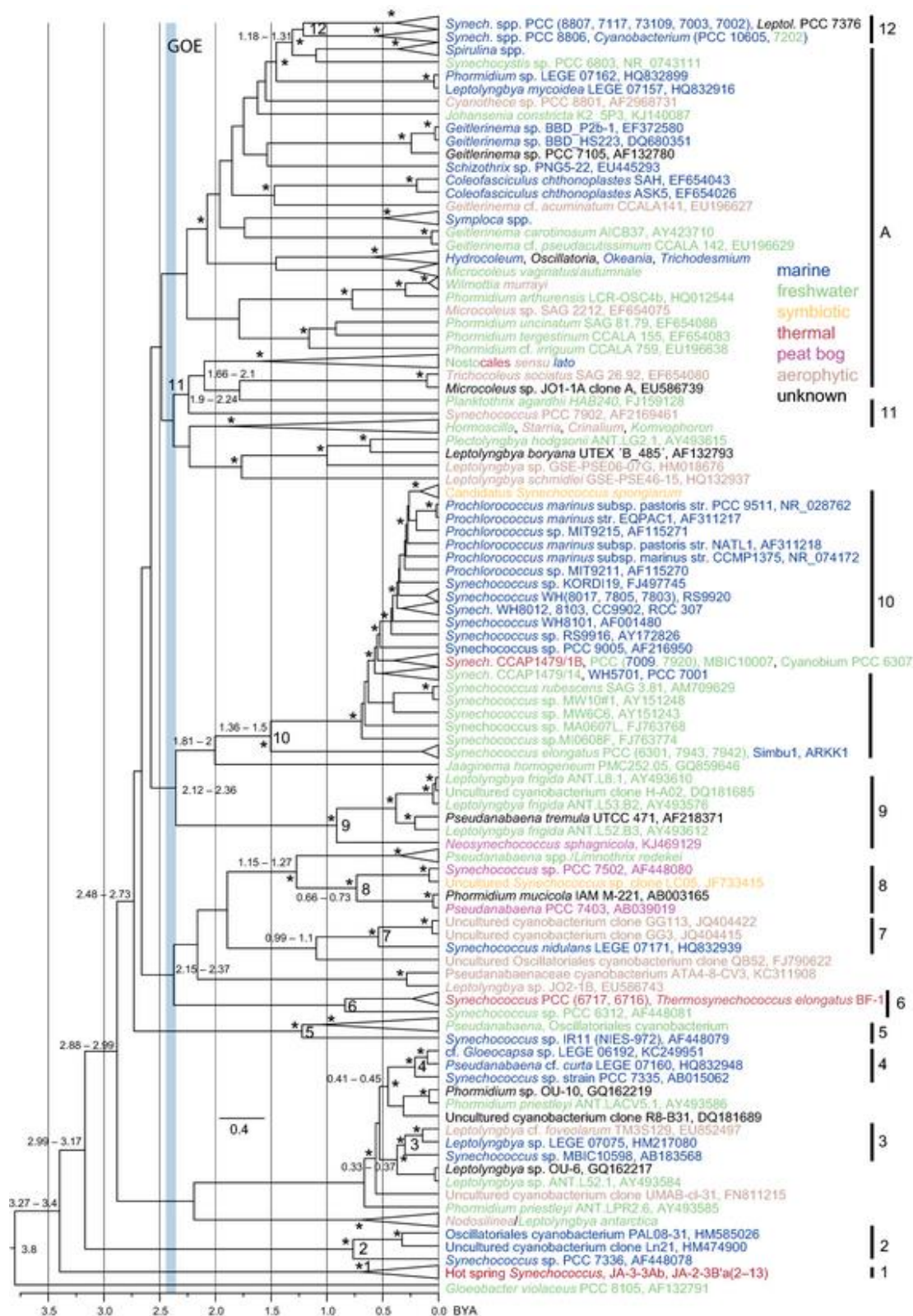
Použití značně velké sady dat v této studii pomohlo překonat limitace dřívějších analýz zlepšením podpory fylogeneze sinic. Samotný původ sinic byl již dříve diskutován a jeho řešení je stále v nedohlednu. Byl prokázán výskyt volného atmosférického kyslíku již v době před 3 miliardami let, což by mohlo naznačovat přítomnost sinic v této době (Lyons *et al.*, 2014). Bylo také vysvětleno, že původně produkovaný kyslík byl v tehdejší atmosféře vychytáván různými reduktanty, což by mohlo vysvětlit tak pozdní nástup GOE (Catling *et al.*, 2005). Výsledek z datovací analýzy naznačuje, že mnohobuněčnost mohla napomoci zvýšenému nárůstu sinic a tím ke zvýšenému metabolismu čili ke zvýšené fotosyntetické aktivitě a následnému rozšíření sítě producentů kyslíku. Tím mohlo následně dojít ke spuštění GOE. Mnohobuněčnost mohla dále poskytnout nové možnosti diversifikace sinic a adaptace na nové životní podmínky.

Jedním z nejzajímavějších rodů sinic z pohledu evoluce je určitě rod *Synechococcus*. Ten je obecně považován za jeden z nejrozšířenějších mikroorganismů na Zemi a s *Prochlorococem* patří mezi významné mořské primární producenty (Flombaum *et al.*, 2013). Evoluční vztahy u *Synechococca* jsou velmi spletité. Již dřívější studie ukázaly zařazení sinic tohoto rodu do několika vývojových větví bez odpovídajících ekologických či morfologických vlastností (Honda *et al.*, 1994; Robertson *et al.*, 2001). Ve své publikaci se těmito evolučními vztahy zabývali Dvořák *et al.* (2014). Jejich práce zahrnovala *de novo* sestavení a anotaci genomu nové *Synechococcu* podobné sinice *Neosynechococcus sphagnicola*, a dále porovnání fylogeneze založené na 16S rRNA sekvencích *Synechococců* s fylogenomickým stromem založeným na 192 ortologních genech ze současných dostupných genomů *Synechococců*. Topologie obou stromů byly velmi podobné, nicméně zrovna v případě zařazení *N. sphagnicola* se stromy od sebe lišily. Ten, který byl založený na 16S rRNA zařadil *N. sphagnicola* do vývojové linie k *Leptolyngbya frigida*, kdežto strom založený na 192 ortologních genech ji zařadil k *Leptolyngbya boryana*, která se nachází v 16S stromu v jiné vývojové větvi (viz Obr. 6, Obr. 7). Podle výsledků fylogenetické analýzy autoři došli k závěru, že *Synechococcus* je polyfyletickým rodem. Jednotlivé linie obsahovaly směs sladkovodních i mořských izolátů žijících jak v oblastech s mírnějšími, tak v oblastech s vysokými teplotami.

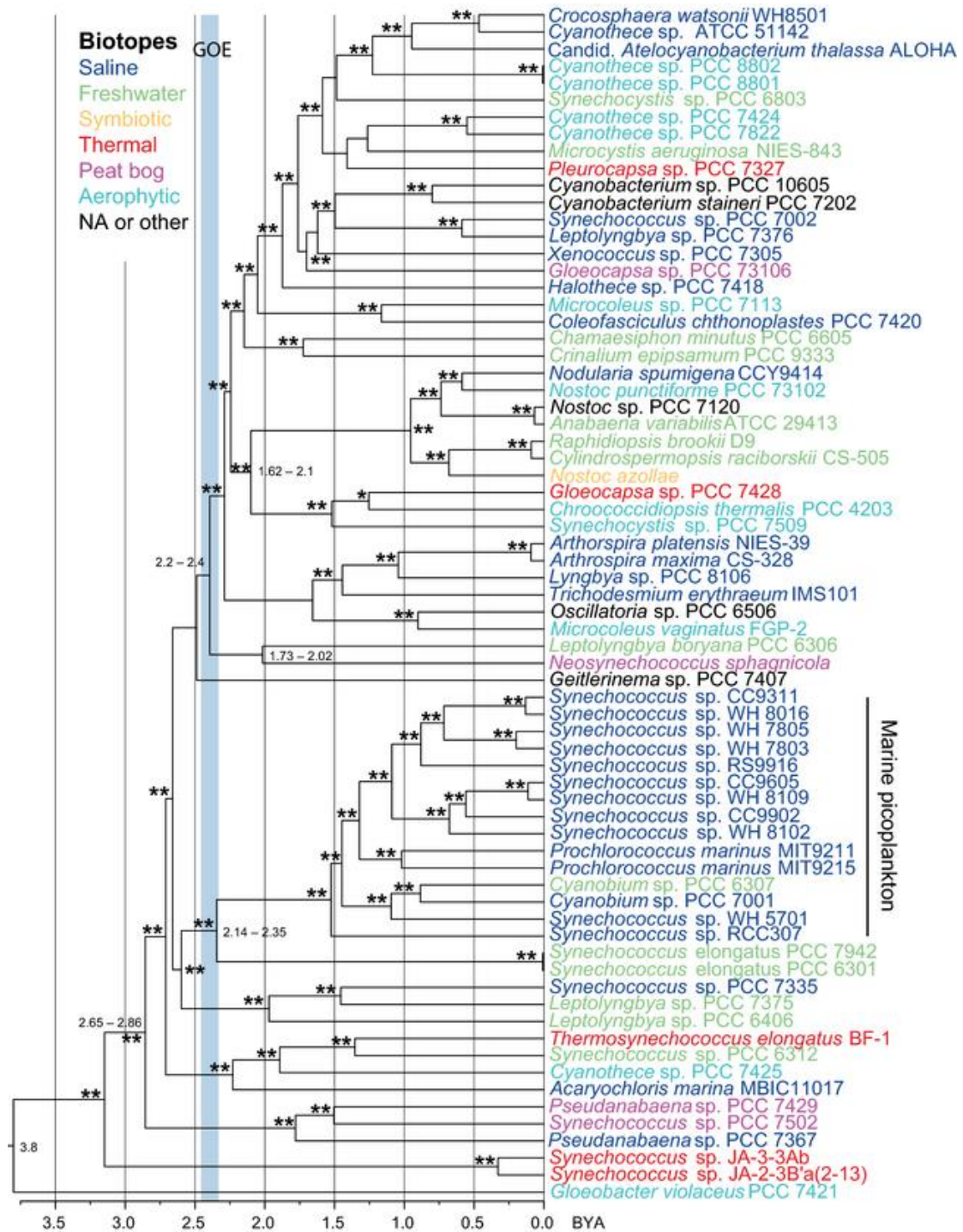
Dále autoři provedli datování obou zkonstruovaných stromů. Kalibrace datování byla totožná pro oba stromy. Kalibračním bodem č. 1 byl vznik akinet, který byl ustanoven na základě fosilních záznamů. Spodní hranice pro kalibraci byla nastavena na 2,1 mld. let

(Amard *et* Bertrand-Sarfati, 1997). Kalibračním bodem č. 2 byl kořen stromu. Spodní hranice pro kalibraci byla nastavena na 3,8 mld. let na základě fosilních nálezů fotosyntetizujících sinic (Nisbet *et* Sleep, 2001). Horní hranice byla nastavena na 2,7 mld. let představující pravděpodobný vznik sinic produkujících kyslík (Brocks *et al.*, 1999). Třetím kalibračním bodem byla doba vzniku sinic tvořící filamenta. Spodní hranice pro kalibraci byla nastavena na 3,08 mld. let a horní hranice na 2,38 mld. let (Shirrmeister *et al.*, 2013).

Chronogram založený na sekvencích 16S rRNA ukázal, že nejstaršími by mohly být kmeny obývající termální oblasti, ze kterých se pak dále šířily do moří a následně sladkovodního prostředí. Ne všechny vývojové linie byly podle analýzy evolučně stejně staré. Zatímco se zdá, že vývojová větev č. 10, která obsahovala směs sladkovodních a mořských kmenů, se vyvinula před 1,5-1,36 mld. let, vývojová větev 1 vznikla pravděpodobně před 3,4-3,27 mld. let. Rozpor mezi 16S rRNA chronogramem a chronogramem založeným na ortologních genech vznikl u nejvíce zastoupené linie mořských pikoplanktonních *Synechococců* ve vývojové větvi č. 10. Podle 16S rRNA se linie oddělila před 2-1,81 mld. let, zatímco podle ortologů před 2,35-2,14 mld. let. Pouze 3 z vývojových větví, 1, 2 a 5 vznikly před GOE, který byl v období před 2,32-2,45 mld. let (Bekker *et al.*, 2004) (viz Obr. 6, Obr. 7). Samotný původ celého rodu *Synechococcus* autoři podle výsledků datování předpokládají do doby před 3 mld. let. Z velké části se výsledky této analýzy shodovaly s dřívějšími publikacemi (Falcon *et al.*, 2010; Shirrmeister *et al.*, 2013). K těmto výsledkům bylo nutné ovšem dodat, že jsou do určité míry nejisté, a to hlavně z důvodu, že v porovnání s rostlinami a živočichy mají sinice mizivé množství fosilních záznamů. I přesto se dá *Synechococcus* označit za jeden z nejúspěšnějších a nejvlivnějších organismů na Zemi. Za více jak 3 mld. let došlo 12x k jeho divergenci. Jeho celosvětové dominanci připisují autoři několik faktorů. Zaprvé rychlou generační dobu (Moore *et al.*, 1995), dále schopnost kompetice o zisk živin (Young, 2007) a světla (Moore *et al.*, 1993) díky malé velikosti a tvaru buněk, a nakonec díky převládání horizontálního přenosu genů skvělá schopnost zisku a využívání cizí DNA, která může poskytnout selektivní výhodu.



Obr. 6: Chronogram všech linií *Synechococca* a dalších sinic založený na Bayesiánské fylogenezi sekvencí 16S rRNA. Identifikované vývojové větve jsou číselně označeny v pravé části obrázku. Hvězdičky znázorňují hodnoty posteriorních pravděpodobností nad 0,9. Informace o stáří v jednotkách miliard let jsou přiřazeny k jednotlivým uzlům. Organismy jsou barevně označeny podle svého prostředí výskytu, které je uvedeno na obrázku. GOE je znázorněno modrým pruhem. Převzato z Dvořák *et al.* (2014).



Obr. 7: Chronogram Maximum-likelihood fylogenomického stromu založeného na 192 ortologních genech. Hodnoty podpory uzlů bootstrapové analýzy nad 70 % jsou znázorněny jednou hvězdičkou, nad 90 % dvěma hvězdičkami. Informace o stáří v jednotkách miliard let jsou přiřazeny k jednotlivým uzlům. Organismy jsou barevně označeny podle svého prostředí výskytu, které je uvedeno v obrázku. GOE je znázorněno modrým pruhem. Převzato z Dvořák *et al.* (2014).

2.4 Sekundární metabolismus sinic

Sinice jsou známými producenty mnoha toxinů. Nejvýrazněji se objevují v letních měsících a jejich toxiny mají mnoho nežádoucích účinků na lidský organismus (Dittmann *et al.*, 2012). Kromě toxinů jsou ale sinice zdrojem mnoha jiných a užitečných látek (Burja *et al.*, 2001). Syntézu těchto látek mají na svědomí z velké části syntázy pro neribozomální peptidy (NRPS) a polyketid syntázy (PKS). Analýzou sekundárního metabolismu byly tyto syntázy nalezeny až u 70 % vybraných genomů sinic (Shih *et al.*, 2013).

Calteau *et al.* (2014) identifikovali u 89 ze 126 genomů sinic 190 klastrů genů pro NRPS, 162 pro PKS a 100 hybridních klastrů kódujících jak NRPS tak PKS. Proteinové sekvence všech z těchto 452 klastrů autoři porovnali za účelem seskupení podobných drah do společných rodin klastrů genů, které mohou kódovat megasyntetázy zapojené do biosyntézy blízké příbuzných metabolitů. Autoři byli schopni identifikovat pár zajímavých biosyntetických drah. První zahrnovala PUFA PKS (polyketid syntázy polynenasycených mastných kyselin), které jsou zapojeny do produkce glykolipidů heterocyst, což je poslední krok zrání heterocyst, který zajišťuje jejich nepropustnost kyslíku (Fan *et al.*, 2005; Kumar *et al.*, 2010). Dále byly nalezeny klastry obsahující geny spojené s transportními systémy sideroforů, což jsou látky jejichž náplní je získávání železa z prostředí (Neilands, 1995). Tento nález byl pro autory nečekaný, protože u sinic bylo dříve charakterizováno pouze malé množství sideroforů, příkladem může být anachelin 1 původně charakterizovaný v sinici *Anabaena cylindrica* CCAP 1403/2A (Beiderbeck *et al.*, 2014).

Jedny ze známých sekundárních metabolitů nalezených ve všech hlavních vývojových liniích bakterií, jsou bakteriociny (Riley *et al.*, 2002). Bakteriociny jsou antimikrobiální látky a používají se v potravinářském průmyslu jako konzervanty (Galvez *et al.*, 2008) či antibiotika (Piper *et al.*, 2009; Wieland Brown *et al.*, 2009). U sinic byly také nalezeny klastry genů zodpovídající za syntézu bakteriocinů. Tyto klastry lze klasifikovat do 7 skupin na základě odlišné organizace katalytických domén v klastrech. Mnoho konkrétních produktů těchto klastrů genů je zatím zcela neznámých (Wang *et al.*, 2011), ale u některých druhů sinic byl objeven protein sloužící k syntéze určitého typu bakteriocinů, lantibiotik (Begley *et al.*, 2009; Goto *et al.*, 2010). U sinice *Prochlorococcus marinus* MIT9313 byla navíc zjištěna přítomnost samotných lantipeptidů (Li *et al.*, 2010).

3 Cíl práce

1. Vypracování literární rešerše na téma fylogenomika a využití celogenomového sekvenování ve výzkumu evoluce sinic.
2. Bioinformatická analýza primárních sekvenačních dat: sestavení genomů, anotace kódujících regionů, repeticí, tRNA.
3. Rekonstrukce fylogenetických vztahů mezi osekvenovanými genomy s genomy v databázi.

4 Materiál a metody

4.1 Biologický materiál

Vzorky DNA izolované z 2 druhů sinic “*Synechococcus* sp. S4” a “*Jacksonvillea apiculata* 2014/77b”. Kmeny jsou uloženy ve sbírce kmenů sinic a řas Katedry botaniky PřF UP. Kultury byly pěstovány v 90mm Petriho miskách v tekutém Zehnderově médiu (Staub, 1961) za následujících podmínek: Teplota 26 ± 1 °C, osvětlení $20 \mu\text{mol m}^{-2} \text{s}^{-1}$, světelný režim 12 h světlo/12 h tma. Sinice *Jacksonvillea apiculata* byla získána z prostředí dna sladkovoního jezera, Sinice *Synechococcus* sp. S4 má pravděpodobně epizoický nebo endozoický způsob života, byla získána z prvoka *Paramecium* sp.

4.2 Sekvenování DNA

Sekvenování DNA bylo provedeno v analyzátoru Illumina MiSeq pracovníky Ústavu Experimentální Botaniky, AV ČR, Olomouc. Knihovna pro sekvenování byla připravena pomocí kitu TruSeq Nano DNA (Illumina Inc., USA) z 200 ng DNA podle instrukcí výrobce, ale digesce DNA byla provedena v ultrazvukovém sonikátoru Bioruptor Plus (Diagenode, Belgie) a velikostní selekce byla modifikována za účelem zisku inzertů o velikosti kolem 1 000 bp. Velikost inzertů byla určena kitem Agilent High Sensitivity (Agilent Technologies Inc., USA) a koncentrace knihoven byla stanovena pomocí kitu KAPA Library Quantification for Illumina (Kapa Biosystems, USA). Pro sekvenování párových konců a produkce čtení o délce 300 bp byl použit MiSeq Reagent Kit v3 (Illumina Inc., USA).

4.3 Bioinformatická analýza

Početně náročné analýzy byly zpracovány v prostředí Linux Ubuntu 14.04 LTS na výkonném HP Workstation Z820 (2x Intel Xeon 6ti jádrový procesor po 2 vláknech), 128 GB RAM, 2x 2 TB harddisk.

4.3.1 Zpracování primárních sekvenčních dat

Primární sekvenční data vzorku DNA druhu *Jacksonvillea apiculata* 2014/77b byla sestavena *de novo* do kontigů pomocí programu MIRA verze 4.0.2 (Chevreux, 1999). Primární sekvenční data vzorku DNA druhu *Synechococcus* sp. S4 byla sestavena do kontigů pomocí téhož programu podle referenčního genomu sinice *Cyanobium gracile* PCC 6307 na

základě předpokládané příbuznosti těchto dvou druhů. U *de novo* sestavení genomu *J. apiculata* bylo nastaveno přesné sestavování `job = genome,denovo,accurate`, dále způsob, jakým byl genom sekvenován, tedy sekvenování párových konců, `autopairing` a technologie použitá k sekvenování genomu, `solexa` od firmy Illumina, `technology = solexa`. Nakonec bylo v nastavení zakázáno přerušení analýzy z důvodu délky jednotlivých čtení >40 bází `parameters = -NW:cmrn1=no`. U sestavování genomu *Synechococcus* sp. S4 bylo nastaveno přesné sestavení podle reference `job = genome,mapping,accurate` a samotná referenční sekvence `is_reference`. Poté byl nastaven způsob sekvenování, tedy sekvenování párových konců, `autopairing`, dále byla nastavena technologie, která byla použita k sekvenování genomu, `solexa` od firmy Illumina, `technology = solexa` a zákaz přerušení analýzy z důvodu délky jednotlivých čtení >40 bází `parameters = -NW:cmrn1=no`.

4.3.2 Anotace kódujících oblastí

Po sestavení sekvence DNA *J. apiculata* do kontigů byla provedena jejich identifikace v programu BLASTN (Altschul *et al.*, 1990). Kontigy, jejichž sekvence DNA nebyla identifikována jako žádný známý druh sinice dostupný v databázi GenBank, ale odpovídala jiným bakteriálním druhům, byly ze sekvenčních dat manuálně odstraněny. Po odstranění kontaminantů zůstal celkový počet kontigů v sekvenčních datech 256. Genom *Synechococca* sp. S4 byl sestaven podle referenčního genomu sinice *Cyanobium gracile* PCC 6307. Byl tedy tvořen jen 1 kontigem. Poté byly anotovány kódující oblasti v sekvenčních datech obou sinic použitím nástroje RAST (Rapid Annotation using Subsystem Technology) v základním nastavení s povolením opravy posunu čtecích rámců. RAST nejdříve vyhledal tRNA sekvence v genomu pomocí programu tRNAscan-SE (Lowe *et Eddy*, 1997), a poté identifikoval rRNA kódující geny pomocí svého nástroje „search_for_rnas“. Následně provedl prvotní vyhledávání genů programem GLIMMER2 (Delcher *et al.*, 1999). Nalezené geny byly podrobeny porovnání s malou sadou proteinů univerzálních pro prokaryota. Na základě toho byly nalezeny nejbližší sousední genomy ke zkoumanému genomu. Poté byl nově přidáný genom porovnán se sadami proteinů těchto sousedů a byla určena funkce odpovídajících genů (Aziz *et al.*, 2008).

4.3.3 Anotace repetitivních sekvencí

Repetitivní sekvence v sekvenčních datech byly anotovány pomocí programu RepeatMasker verze 4.0.6, který slouží k vyhledávání rozptýlených repetit v genomech (Smit *et al.*, 1996). Konkrétním nástrojem pro vyhledávání repetit byl RMBlast, kompatibilní verze standardního NCBI BLASTu pro RepeatMasker. Za vyhledávací nástroj (-e) byl zvolen portál NCBI. Program byl spuštěn za paralelního použití 4 procesorů (-pa). Dále bylo nastaveno pomalejší a citlivější vyhledávání repetit v databázi (-s). Za příbuzný druh nacházející se v taxonomické databázi NCBI a zároveň v databázi repetitivních sekvencí RMBase (-species) byl zvolen organismus *Arabidopsis thaliana* (viz příkaz níže).

```
./RepeatMasker -e ncbi -pa 4 -s -species arabidopsis
```

Inzerční elementy byly v sekvenčních datech vyhledány programem BLAST v databázi ISfinder v základním nastavení (dostupné na <https://www-is.biotoul.fr/>) (Zhang *et al.*, 2000).

Segmenty nahromaděných pravidelně rozmístěných krátkých palindromických repetit (CRISPR) byly v sekvenčních datech vyhledány online programem CRISPRfinder v základním nastavení (dostupné na <http://crispr.i2bc.paris-saclay.fr/>) (Grissa *et al.*, 2007).

4.3.4 Anotace tRNA

Počet a typy tRNA sekvencí v sekvenčních datech byly zjištěny pomocí programu tRNAscan-SE v základním nastavení (dostupné na <http://lowelab.ucsc.edu/tRNAscan-SE/>) (Lowe *et Eddy*, 1997).

4.3.5 Analýza přítomnosti genů sekundárního metabolismu

Pomocí programu antiSMASH verze 4.0 v základním nastavení byly v genomech obou sinic vyhledány klastry genů pro biosyntézu sekundárních metabolitů (Tilman *et al.*, 2015) (dostupné na <http://antismash.secondarymetabolites.org>).

4.3.6 Fylogenomická analýza

Část této analýzy byla provedena v programu Hal pro automatickou fylogenetickou analýzu genomových dat (Robbertse *et al.*, 2011). Tento program využívá několik podprogramů spouštěných za sebou postupně v několika krocích. V prvním kroku byla

provedena identifikace ortologů. Každý protein z proteomů poskytnutých k fylogenomické analýze byl srovnáván se všemi ostatními proteiny z těchto proteomů v programu BLAST. Na základě výstupu z BLASTu byly poté pomocí programu MCL (Markov Cluster) všechny proteiny rozděleny do skupin podle příbuznosti. Následně byly vybrány skupiny proteinů obsahující vždy právě jeden protein z jednoho organismu, jehož sekvence proteinů byly použity k analýze. Uvnitř těchto skupin bylo prováděno mnohočetné přiřazování sekvencí proteinů v programu MUSCLE, a poté byly jednotlivé výstupy z přiřazování editovány programem Gblocks, který odstranil méně věrohodně přiřazené sekvence a ostatní rozdílné oblasti.

```
hal -a muscle -E -N 24 -o Se,Sg
```

a – program pro mnohočetné přiřazování sekvencí

E – vypsání zprávy o chybách

N – počet jader počítače použitých k analýze

o – vnější skupiny

Následně byly všechny ortology spojeny do jedné matice. Matice byla využita k sestavení fylogenomického stromu metodou Maximum likelihood v programu RAxMLHPC verze 8.0.0 (Stamatakis, 2014). Za substituční model byl zvolen CAT, který počítá s místně specifickými vlastnostmi proteinů v průběhu evoluce, kdy je každá pozice proteinu pod velmi specifickým selekčním tlakem, a na těchto pozicích se může vlivem tohoto tlaku nacházet jedna z 20 známých aminokyselin v průběhu evoluce (Lartillot, N. et Philippe, H., 2004) s proteinovou maticí LG (Le, S. Q. et Gascuel, O., 2008).

Pro fylogenomickou analýzu *Synechococca* sp. S4 byly jako vnější skupiny použity druhy *Synechococcus elongatus* PCC 6301 a *Synechococcus elongatus* PCC 7942.

```
raxmlHPC-PTHREADS-AVX -m PROTCATLG -p 12345 -T 4 -o Se,Sg
```

Všechny sinice použité pro fylogenomickou analýzu sinice *Synechococcus* sp. S4 jsou uvedeny v Tab. 1 v Příloze.

Pro fylogenomickou analýzu *Jacksonvillea apiculata* bylo nejdříve pomocí programu AMPHORA2-master vybráno 14 základních (core) proteinů kódovaných ve všech ostatních genomech sinic. Program byl spuštěn v základním nastavení. Následně bylo u těchto sad proteinů provedeno mnohočetné přiřazování v programu Hal pomocí programu MUSCLE.

Jako vnější skupiny byly vybrány sinice *Gloeobacter kilaueensis* JS1 a *Gloeobacter viloaceus* PCC 7421.

```
hal -a muscle -E -N 24 -o GJ,GV
```

Na základě mnohočetného přiřazení byl sestrojen fylogenomický strom metodou Maximum likelihood v programu RAxMLHPC verze 8.0.0 se substitučním modelem CAT a proteinovou maticí LG. Jako vnější skupiny byli zadáni výše zmínění zástupci rodu *Gloeobacter*.

```
raxmlHPC-PTHREADS-AVX -m PROTCATLG -p 12345 -T 4 -o GJ,GV
```

Všechny sinice použité pro fylogenomickou analýzu sinice *Jacksonvillea apiculata* 2014/77b jsou uvedeny v Tab. 1 v Příloze. Topologie obou fylogenomických stromů byla podpořena bootstrapovou analýzou s 1 000 opakováními přes portál CIPRES Science Gateway (Miller *et al.*, 2010). Výsledné stromy byly vizualizovány v programu FigTree verze 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

4.3.7 Analýza horizontálního přenosu genů

Za pomoci programu SigiHMM (Waack *et al.*, 2006) byla provedena analýza horizontálního přenosu genů v základním nastavení se zvýšenou citlivostí vyhledávání potenciálních přenesených genů u sinic *Synechococcus* sp. S4 a *Cyanobium gracile* PCC 6307.

```
./SigiHMM -sens 0.95
```

4.3.8 Analýza přítomnosti delecí, inverzí

DNA sekvence anotovaných genomů vzorku *Synechococcus* sp. S4 a *Cyanobium gracile* kmen PCC 6307 byly porovnány v programu BLASTN v základním nastavení. Výstupní soubor porovnání těchto genomů byl společně se soubory obou anotovaných genomů ve formátu GenBank nahrán do programu ACT (Artemis Comparison Tool) (Carver *et al.*, 2005) a manuálně prohledány na přítomnost delecí a inverzí.

4.3.9 Ontologie genů (GO)

Analýza funkce genů podle slovníku Gene Ontology, popisujícího funkce genů, byla provedena pro potvrzení jejich funkce určené RASTem, a u *Synechococcus* sp. S4 také pro

porovnání genů určité funkce s nejbližším příbuzným druhem, *C. gracile* PCC 6307. Analýza byla provedena za pomoci skriptů, které po převedení souborů z formátu GenBank do tabulkového formátu porovnávají soubory sekvencí s knihovnou všech genů anotovaných GO analýzou a vyjmou ze souboru geny, které GO analýzou nebyly anotovány.

4.3.10 Datování fylogeneze *Synechococcus* sp. S4

Sekvenční data proteomů 130 druhů sinic byla použita k vytvoření mnohočetného přiřazení v programu Hal s pomocí programu MUSCLE. Jako vnější skupiny byly vybrány sinice *Gloeobacter kilaueensis* JS1 a *Gloeobacter viloaceus* PCC 7421.

```
hal -a muscle -E -N 24 -o GJ,GV
```

Na základě mnohočetného přiřazení byl následně sestrojen fylogenomický strom metodou Maximum likelihood v programu RAXMLHPC verze 8.0.0 se substitučním modelem CAT a proteinovou maticí LG. Jako vnější skupiny byli opět zadáni oba zástupci rodu *Gloeobacter*.

```
raxmlHPC-PTHREADS-AVX -m PROTCATLG -p 12345 -T 4 -o GJ,GV
```

Pro tuto analýzu byly použity proteomy sinic uvedených v Tab. 1 v Příloze. Datování fylogeneze bylo následně provedeno na tomto sestrojeném fylogenomickém stromu v programu r8s verze 1.70 pro odvození tempa molekulární evoluce (Sanderson, 2003). Jednotky pro určení délek větví byly nastaveny na `lengths=persite`, což odpovídá očekávanému počtu substitucí na místo sekvence. Toto nastavení se typicky používá pro maximum likelihood analýzy. V nastavení, zda je strom ultrametrický bylo zvoleno `ultrametric=no`, protože koncové uzly představující druhy nejsou ve stejné rovině a nejsou stejně vzdálené od společného předka. Délka sekvence mnohočetného přiřazení byla `nsites=12052`. Větve s nulovou délkou byly převedeny do polytomických nastavením příkazu `collapse`. Dále byl nastaven odhad stáří všech uzlů `unfixage taxon=all`. Uzly určené ke kalibraci datování byly pojmenovány a označeny příkazem `mrca`. Zvolená místa byla 3, a poté pomocí příkazu `constrain taxon` byl u těchto míst nastaven odhad jejich stáří. Kalibrace byla z větší části provedena na základě kombinace fosilních a molekulárních kalibrací publikovaných v práci Shirmmeistera *et al.* (2013).

C1 – `mrca akinets Ft Rb` – Kalibračním bodem č. 1 byl vznik akinet, který byl ustanoven na základě fosilních záznamů. Spodní hranice pro kalibraci byla nastavena na 2,1 mld. let (Amard *et* Bertrand-Sarfati, 1997)

C2 – `mrca root Gv SU` – Kalibračním bodem č. 2 byl kořen stromu. Spodní hranice pro kalibraci byla nastavena na 3,2 mld. let na základě fosilních nálezů fotosyntetizujících sinic (Satkoski *et al.*, 2015). Horní hranice byla nastavena na 2,7 mld. let představující pravděpodobný vznik sinic produkujících kyslík (Brocks *et al.*, 1999).

C3 – `mrca filaments PQ AC` – Třetím kalibračním bodem byla doba vzniku sinic tvořících filamenta. Spodní hranice pro kalibraci byla nastavena na 3,08 mld. let a horní hranice na 2,38 mld. let (Shirrmeister *et al.*, 2013).

Příkazem `set checkGradient=yes` byla nastavena dodatečná kontrola výsledku analýzy. K odhadu časů oddělení větví a substitučního tempa byla použita metoda PL (penalized likelihood), která kombinuje parametrický model mající rozdílná substituční tempa na každé větvi s neparametrickou penalizací, která se zvyšuje, pokud se rychlost temp přílišně zvyšuje od větve k větvi (Sanderson, 2002). Smoothing parametr byl nastaven na hodnotu 1 000, vyhledávací algoritmus byl zvolen TN. Dále byl nastaven odhad intervalů spolehlivosti pro jednotlivé uzly `divtime method=pl smoothing=1000 algorithm=tn confidence=yes`. Příkazem `showage` byly odhady staří a temp substituce vypsány do tabulky. Výsledek analýzy byl vypsán jako chronogram `describe plot=chronogram plot=chrono_description`.

Výsledný chronogram byl vizualizován v programu FigTree verze 1.4.3.

5 Výsledky

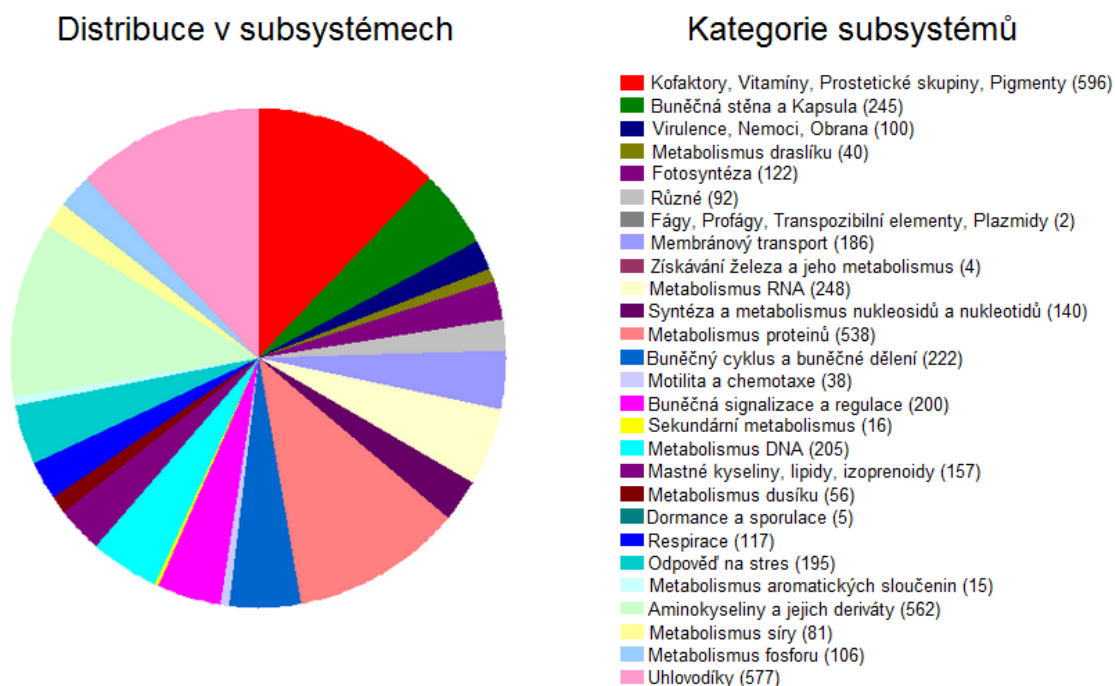
5.1 *Jacksonvillea apiculata*

5.1.1 Obecné informace o genomu

Jacksonvillea apiculata je sladkovodní sinicí obývající dna jezer (Hašler *et al.*, 2017). Genom této sinice byl sestaven *de novo*. Pomocí nástroje RAST byla provedena anotace kódujících sekvencí *J. apiculata*. Po manuálním odstranění kontaminací jinými bakteriálními sekvencemi tvořilo sekvenční data 255 kontigů. Velikost genomu *J. apiculata* byla stanovena na 10 689 215 Mb, s 44% obsahem GC bází. Průměrná délka sestavených kontigů (N50) byla 97 856 bází. V sekvenci bylo nalezeno 10 096 protein kódujících sekvencí a 136 různých RNA. Základem anotace pomocí RASTu je zařazení jednotlivých kódujících sekvencí do určitých subsystémů podle jejich funkce určené na základě identifikace programem BLAST. Do subsystémů bylo zařazeno celkem 34 % ze všech kódujících sekvencí. U 5 794 sekvencí byla identifikována funkce, 4 302 jich zůstalo pouze hypotetických. Přehled byl zapsán do Tab. 2. Zastoupení anotovaných kódujících sekvencí v subsystémech a kategorie subsystémů jsou zobrazeny na Obr. 8. Největší počet genů (596) byl zařazen do subsystému Kofaktory, Vitamíny, Prostetické skupiny, Pigmenty. Dalšími nejvíce zastoupenými subsystémy s počtem zařazených genů nad 500 byly Uhlovodíky (577), Aminokyseliny a jejich deriváty (562) a Metabolismus proteinů (538). Mezi nejméně zastoupené subsystémy, do kterých bylo zařazeno méně než 10 genů patřily Dormance a sporulace (5), Získávání železa a jeho metabolismus (4) a Fágy, Profágy, Transpozibilní elementy, Plazmidy (2).

Tab. 2: Přehled zařazení protein kódujících sekvencí sinice *J. apiculata* nástrojem RAST do subsystémů.

Celkový počet protein kódujících sekvencí	Zařazeno do subsystému		Nezařazeno do subsystému	
	Identifikovaných	Hypotetických	Identifikovaných	Hypotetických
10 096	3 434 (34 %)		6 662 (66 %)	
	3 319	115	2 475	4 187



Obr. 8: Zastoupení anotovaných genů *J. apiculata* v subsystémech a jejich kategorie podle RASTu.

5.1.2 Anotace repetitivních sekvencí

Program RepeatMasker verze 4.0.6. odhalil v sekvenčních datech sinice *Jacksonvillea apiculata* různé druhy repetitivních sekvencí na základě jejich porovnání s repeticemi nacházejícími se v databázi RMBase pro rostlinu Huseníček rolní (*Arabidopsis thaliana*). Tento organismus byl zvolen z důvodu jeho nejbližší příbuznosti k sinicím v rámci všech organismů nacházejících se v databázi RMBase. Z celkového počtu 60 nalezených repetitivních sekvencí 10 tvořilo retroelementy či jejich části, z toho 9 jich patřilo do nadrodiny Ty1/Copia a 1 do nadrodiny Gypsy/DIRS1. Všechny tyto nalezené retroelementy patřily do skupiny LTR, což znamená, že jejich sekvence je, nebo byla ohraničena dlouhými koncovými repeticemi. Dalších 14 nalezených retroelementů či jejich částí patřilo do třídy DNA transpozónů. Z toho 6 jich bylo zařazeno do nadrodiny Hobo-activator, 6 do nadrodiny MULE-MuDR, 1 do nadrodiny Helitronů a 1 do nadrodiny ATREP. Následně bylo nalezeno 15 repetitivních sekvencí patřící mezi malé RNA. Jednalo se o 10 sekvencí pseudogenů rRNA velké podjednotky ribozomů (LSU-rRNA-Ath) a 5 sekvencí pseudogenů rRNA malé podjednotky ribozomů (SSU-rRNA-Ath). Poslední nalezené sekvence byly různé jednoduché repetice. Těch bylo nalezeno celkem 21. Konkrétní nalezené repetice jsou zapsány v Tab. 3 na základě dostupných informací v databázi RMBase.

Tab. 3: Repetitivní sekvence nalezené v genomu *J. apiculata* programem RepeatMasker verze 4.6.0.

Typ repeticce	Nadrodina	Rodina (počet nalezených repetic)	Popis repeticce
Retrotranspozony	Ty1/Copia	ATCopia11I (1)	část vnitřní oblasti LTR-retrotransponu ATCOPIA11, ve které se mohou nacházet geny původního retroviru
		ATCopia35 LTR (1)	dlouhá terminální repeticce (LTR) retrotransponu ATCOPIA35
		ATCopia82_I-int (1)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotransponu ATCOPIA82
		Copia-3_AT-I (1)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotransponu Copia-3
		ATCopia61_I-int (1)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotransponu ATCOPIA61
		ATCopia24I (2)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotransponu ATCOPIA24
		ATCopia14I (1)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotransponu ATCOPIA14
		ATCopia18A_I (1)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotransponu ATCOPIA18A
	Gypsy/DIRS1	Athila4C_I-int (1)	část vnitřní oblasti endogenního retroviru ATHILA4C kódující polyprotein ATHILA4C1p
DNA transpozony	Hobo-Activator	hAT-N4_Aly (6)	hobo-Activator DNA transposon identifikovaný z <i>Arabidopsis lyrata</i>
	MULE-MuDR	ATMU12 (1)	MuDR-type transposon z <i>Arabidopsis thaliana</i>
		ATMU10 (1)	fosilní DNA autonomního DNA transponu ATMU10
		Vandal10 (1)	fosilie autonomního DNA transponu Vandal
		Vandal13 (2)	molekulární fosilie autonomního Vandal-like transponu nacházející se na chromosomu 3 <i>A. thaliana</i> . V sekvenci tohoto elementu jsou vloženy další 2 mobilní elementy, ATDNATA1 a ATSINE2A
		Vandal16 (1)	autonomní Vandal-like DNA transposon nacházející se na chromosomech 1 a 3 <i>A. thaliana</i>
	Helitrony	Helitron3 (1)	autonomní DNA transposon z nadrodiny Helitronů
ATREP	ATREPX1 (1)	neautonomní DNA transposon z nadrodiny ATREP transponů. V sekvenci tohoto mobilního elementu je vložen další transposon, ATHATN1	

Tab. 3: Repetitivní sekvence nalezené v genomu *J. apiculata* programem RepeatMasker verze 4.6.0.

Typ repetice	Repetice (počet nalezených)	Popis repetice
Malé RNA	LSU-rRNA_Ath (10)	pseudogen rRNA velké podjednotky ribozomů
	SSU-rRNA_Ath (5)	pseudogen rRNA malé podjednotky ribozomů
Jednoduché repetice	(TGG)n (7)	Opakující se sekvence TGG
	(CTG)n (1)	Opakující se sekvence CTG
	(CAT)n (2)	Opakující se sekvence CAT
	(ATG)n (2)	Opakující se sekvence ATG
	(CGA)n (1)	Opakující se sekvence CGA
	(ATTG)n (1)	Opakující se sekvence ATTG
	(ATGGTG)n (1)	Opakující se sekvence ATGGTG
	(TTGGGG)n (1)	Opakující se sekvence TTGGGG
	(CCCCAA)n (2)	Opakující se sekvence CCCCCA
(CACCAT)n (3)	Opakující se sekvence CACCAT	

Porovnáním sekvence DNA *J. apiculata* programem BLAST s databází ISfinder nebyly nalezeny žádné přítomné inzerční elementy. V programu CRISPRfinder byly vyhledány CRISPRy v sekvenci DNA *J. apiculata*. Bylo nalezeno celkem 16 potvrzených CRISPRů (viz Tab. 4 v Příloze) a 18 sporných CRISPRů (viz Tab. 5 v Příloze). Jako sporné byly programem označeny takové CRISPRy, které byly tvořeny pouze dvěma či třemi přímými repeticemi nebo opakovaná sekvence nebyla vždy 100% identická. Některé opakující motivy CRISPRů byly od sebe odděleny jinou sekvencí DNA. Takový segment se nazývá „spacer“. U potvrzených CRISPRů bylo nalezeno celkem 358 spacerů, u sporných CRISPRů 23 spacerů. Jak lze vidět v Tab. 4, délka CRISPRů je velmi variabilní. Nejdelší CRISPR v *J. apiculata* dosahoval délky 7 767 nukleotidů a obsahoval 105 spacerů mezi opakovanými motivy. Naopak nejkratší CRISPR byl dlouhý 251 nukleotidů s pouze třemi spacery. Délka samotné repetice byla až na 2 výjimky téměř konstantní, průměrná délka byla 36 nukleotidů. U sporných CRISPRů byl naopak poměrně velký rozptyl v délkách repetice, od 23 do 53 nukleotidů, a převážně obsahovaly jen 1 spacer (viz Tab. 5 v Příloze).

5.1.3 Anotace tRNA

V genomu *J. apiculata* se nacházelo celkově 116 sekvencí kódujících různé tRNA a 2 pseudogeny, jeden pro tRNA fenylalaninu a druhý pro tRNA leucinu. Nejvyšší nalezený počet tRNA pro určitou aminokyselinu byl 15, a to pro leucin. Naopak nejnižší nalezený počet tRNA byl 2, pro aminokyselinu histidin. Počet všech nalezených tRNA i jejich antikodonů je uveden níže v Tab. 6.

Tab. 6: Přehled molekul tRNA pro jednotlivé aminokyseliny nalezených v genomu *J. apiculata* programem tRNA scan-SE.

Aminokyselina	Počet nalezených tRNA	Antikodony (počet)				
Leucin	15	GAG (3)	CAG (4)	TAA (2)	TAG (3)	CAA (3)
Arginin	12	ACG (3)	TCT (4)	CCG (3)	CCT (2)	
Serin	10	CGA (3)	GGA (2)	GCT (2)	TGA (3)	
Glycin	8	GCC (3)	TCC (3)	CCC (2)		
Threonin	8	CGT (4)	TGT (2)	GGT (2)		
Alanin	8	CGC (2)	TGC (4)	GGC (2)		
Prolin	7	GGG (2)	CGG (2)	TGG (3)		
Methionin	7	CAT (7)				
Valin	6	CAC (2)	GAC (2)	TAC (2)		
Lysin	5	CTT (2)	TTT (3)			
Fenylalanin	4	GAA (4)				
Tyrosin	4	GTA (4)				
Glutamin	4	TTG (3)	CTG (1)			
Cystein	3	GCA (3)				
Kys. asparagová	3	GTC (3)				
Tryptofan	3	CCA (3)				
Kys. glutamová	3	TTC (3)				
Asparagin	3	GTT (3)				
Isoleucin	3	GAT (3)				
Histidin	2	GTG (2)				

5.1.4 Ontologie genů

Z celkového počtu 10 096 protein kódujících sekvencí anotovaných RASTem byla následnou GO analýzou potvrzená funkce u 1 936 genů, což činí 19,2 % celého genomu. Z celkového počtu kódujících sekvencí bylo navíc 3 832 označených pouze jako hypotetické geny. Tyto hypotetické geny pak nakonec tvoří 38 % celého genomu. GO analýza je prováděna ve 3 kategoriích. První kategorie popisuje místa, kde se produkty genů nacházejí na úrovni vnitrobuněčných struktur a makromolekulárních komplexů (viz Tab. 7). Druhá kategorie popisuje procesy, kterých se dané geny účastní (viz Tab. 8 v Příloze) a poslední kategorie popisuje konkrétní molekulární funkce genů (viz Tab. 9 v Příloze). Počet anotovaných genů i geny samotné se mohou, ale nemusí překrývat v jednotlivých kategoriích i definovaných skupinách. Geny jsou v tabulkách každé kategorie seřazeny podle nejvyššího zastoupení v jednotlivých skupinách. Nejvyšší zastoupení anotovaných genů se zpravidla týká nejobecnějších skupin. Hodnota p-value vyjadřuje pravděpodobnost výskytu „x“ genů z celkového počtu „n“ genů přiřazených k určité GO skupině. Čím blíže je tato hodnota číslu

0, tím pravděpodobnější je správnost přiřazení daných genů k určité GO skupině. Hodnota p-value koreluje se zvyšujícím se počtem přiřazených genů ke stejné skupině.

Tab. 7: Skupiny genů první kategorie anotovaných GO analýzou *J. apiculata*.

Identifikační číslo	Počet anotovaných genů	p-value	Skupina
GO:0005622	152	4.6451125169728e-122	Vnitrobuněčné
GO:0005840	92	1.12915432123107e-73	Ribozóm
GO:0016020	85	4.93865280884893e-68	Membrána
GO:0016021	29	6.66005199130635e-23	Nedílná součást membrány
GO:0005694	9	8.86087191283031e-07	Chromozóm
GO:0005737	7	3.51228691353618e-05	Cytoplazma
GO:0009289	7	3.51228691353618e-05	Pilus
GO:0005960	2	0.225443233018957	Komplex štěpení glycinu
GO:0009379	2	0.225443233018957	Komplex helikáz Hollidayova spoje
GO:0009986	2	0.225443233018957	Buněčný povrch
GO:0015934	2	0.225443233018957	Velká podjednotka ribozomů
GO:0016469	1	0.899218808659587	Proton transportující dvousektorový ATPázový komplex
GO:0019867	1	0.899218808659587	Vnější membrána
GO:0042597	1	0.899218808659587	Periplazmatický prostor

První kategorie je nejobecnější a platí pro ni, že téměř všechny geny v ní anotované, jsou anotované také v některé z dalších kategorií. Nicméně velké množství genů, u nichž byla anotována specifická funkce a procesy, kterých se účastní, není v první kategorii vůbec obsaženo. Nejobecnější a také nejvíce zastoupenou skupinou v první kategorii byla Vnitrobuněčné. Čítala 152 různých genů a každý z nich byl zastoupen v některé z dalších skupin. Druhou nejvíce zastoupenou skupinou byla skupina Ribozóm. Všech 92 genů z této skupiny bylo také zařazeno do skupiny Translace v druhé kategorii, ale ne všechny geny z kategorie Translace (110) byly zařazeny do skupiny Ribozóm. Poslední výrazně zastoupenou skupinou z první kategorie byla Membrána (85). Pouze 9 genů z této skupiny nebylo zastoupeno v některé z dalších skupin. Ve všech případech se jednalo o permeázu z nadrodiny lékových/metabolických transportérů (DMT superfamily).

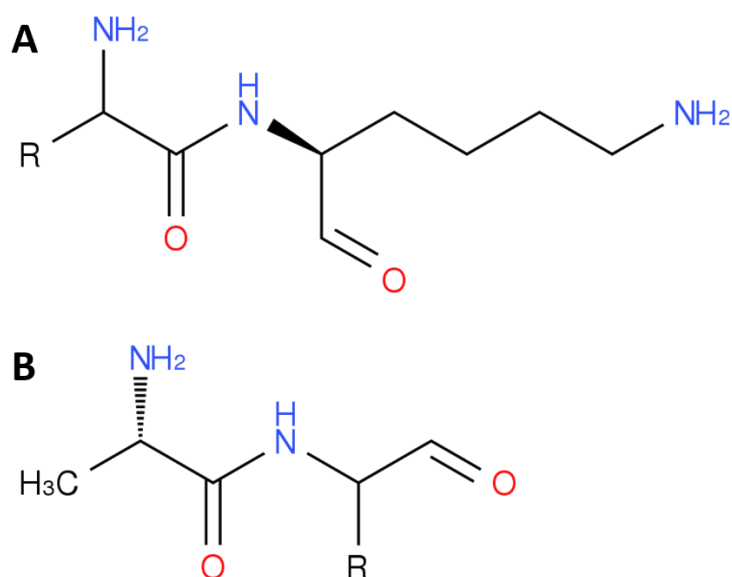
Druhá kategorie je již specifitější, a proto také obsahuje více definovaných skupin a nižší maximální počet zastoupených genů ve skupině. I zde se ovšem objevují poměrně obecné procesy jako Translace, která je se 110 geny nejvíce zastoupenou skupinou v této kategorii, Transport, který je druhý v pořadí se 77 geny a Biosyntetické procesy, které jsou

zde s pouhými 7 geny překvapivým výsledkem, protože v této kategorii je celkem 18 genů ze specifických biosyntetických procesů a ani jeden z nich není zastoupen ve skupině Biosyntetické procesy. Ve skupině Translace jsou pouze 2 geny, které nejsou zastoupeny v žádné jiné skupině. V obou případech jde o ribozóm recyklující faktory. Z dalších skupin spojených s procesem translace, jako Iniciace translace, jsou z celkem 9 genů pouze 3 zastoupeny v obecnější skupině Translace. Ze skupiny Terminace translace, která obsahuje 4 geny není dokonce ani jeden zastoupen ve skupině Translace. Stejně tomu je u skupiny Translační elongace obsahující 2 geny. Co se týče skupiny Transport, pouze 2 geny ze 77 jsou zastoupeny také v jiné skupině z této kategorie týkající se transportu, a to ve skupině Transport draselných iontů. Z dalších 14 skupin spojených s transportem čítajících celkem 68 genů nebyl ve skupině Transport zastoupen žádný gen. V této kategorii je také 12 skupin obsahujících dohromady 68 genů zapojených do skládání, úprav, transportu a metabolismu proteinů (viz Tab. 8 v Příloze).

Třetí kategorie z GO anotace zahrnuje specifické funkce produktů genů na molekulární úrovni. V této kategorii je zastoupeno nejvíce genů a nejvíce definovaných funkčních skupin, kterých je 520. Nejvyšší počet zastoupení genů má funkční skupina Vazba ATP, celkem 115 genů, druhou v pořadí je funkční skupina Strukturní složka ribozómu s 94 geny a třetí nejobsáhlejší skupinou je Adenylát cyklázová aktivita se 79 geny. Naopak nejmenší počet genů, pouze 1, má 72 různých funkčních skupin. Přehled všech funkčních skupin této kategorie je obsažen v Tab. 9 v Příloze.

5.1.5 Analýza přítomnosti genů sekundárního metabolismu

V programu antiSMASH verze 4.0 byly vyhledány geny sekundárního metabolismu. Bylo nalezeno 9 klastrů genů. Čtyři klastry se týkaly metabolismu blíže neidentifikovaných terpenů. Dva klastry obsahovaly geny metabolismu neribozomálních peptidů, což jsou peptidy, které nevznikají translací mRNA na ribosomech, ale jsou syntetizovány jinou cestou. Program antiSMASH také vytvořil předpovídanou strukturu základní jednotky obou peptidů (viz Obr. 9). Dále bylo nalezeno shodně po jednom klastru genů pro metabolismus Bakteriocinu, Resorcinolu a Ladderanu. Z každého klastru bylo následně identifikováno několik genů spojených se sekundárním metabolismem kombinací výsledků z antiSMASH, BLASTu a GO analýzy (viz Tab. 10).



Obr. 9: Předpovězená struktura základní jednotky neribozomálního peptidu z klastru č. 4 (A) a klastru č. 9 (B).

Tab. 10: Klastry genů sekundárního metabolismu nalezené v genomu *J. apiculata* programem antiSMASH verze 4.0.

Klastr (délka v nt)	Celkový počet genů v klastru	Produkty identifikovaných genů spojených se sekundárním metabolismem	GO anotace	Typ metabolitu	Funkce v metabolismu
Klastr č. 1 (20 928)	15	15-cis-fytoen desaturáza		Terpen	Biosyntéza
		Fytoen syntáza	GO:0016767	Terpen	Biosyntéza
		Serin/threonin protein kináza		Terpen	Biosyntéza
Klastr č. 2 (20 817)	13	S hopanoidy asociovaná cukerná epimeráza	GO:0045552	Terpen	Biosyntéza
		Radikální SAM protein HpnH asociovaný s biosyntézou hopanoidů		Terpen	Biosyntéza
		Fytoen syntáza		Terpen	Biosyntéza
		Diguanylát cykláza		Terpen	Biosyntéza
		3-dehydrochinát syntáza	GO:0003856	Terpen	Biosyntéza
Klastr č. 3 (21 945)	14	NAD dependentní dehydratáza		Terpen	Biosyntéza
		Glykosyl transferáza		Terpen	Biosyntéza
		CRP regulátor transkripce		Terpen	Regulace
		Polyprenol monofosfomanóza syntáza		Terpen	Biosyntéza

Tab. 10: Klastry genů sekundárního metabolismu nalezené v genomu *J. apiculata* programem antiSMASH verze 4.0.

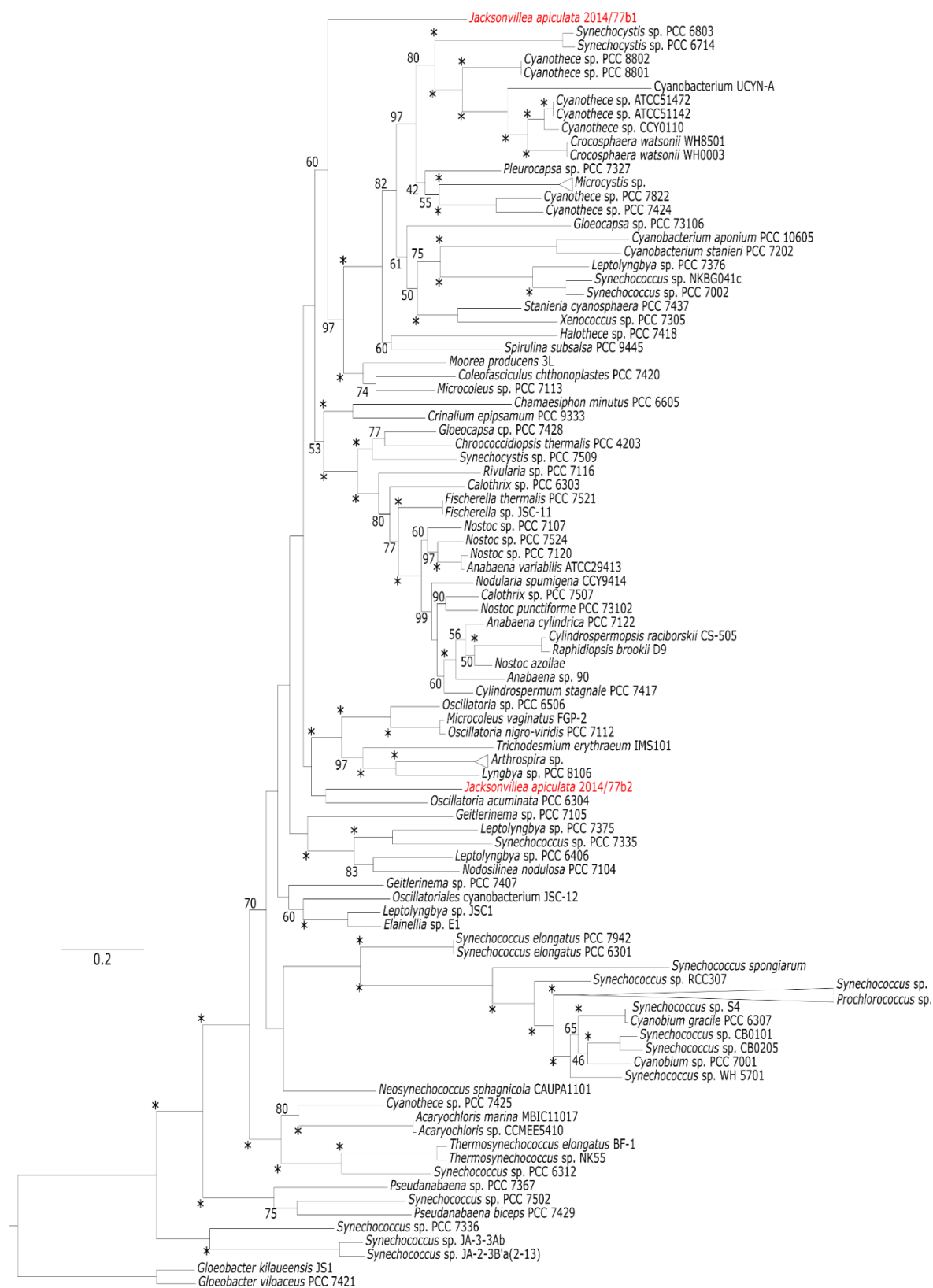
Klastr č. 3 (21 945)	14	Skvalen-hopen cykláza		Terpen	Biosyntéza
		2-sukcinylbenzoát- CoA ligáza		Terpen	Biosyntéza
Klastr č. 4 (38 540)	34	ATP vazebný protein FtsE		Neribozomální peptid	Transport
		Syntáza neribozomálního peptidu		Neribozomální peptid	Biosyntéza
		Histidin kináza		Neribozomální peptid	Regulace
		ABC transportér permeáza		Neribozomální peptid	Transport
		SAM dependentní metyltransferáza		Neribozomální peptid	Biosyntéza
		N-Acylglukosamin-6- fosfát 2-epimeráza	GO:0047465	Neribozomální peptid	Biosyntéza
		Pseudouridin syntáza		Neribozomální peptid	Biosyntéza
Klastr č. 5 (10 283)	13	Gcn5 příbuzná N- acetyltransferáza		Bakteriocin	Biosyntéza
		Aminotransferáza		Bakteriocin	Biosyntéza
		Nifl 1 podobný prekurzor vedoucího peptidu		Bakteriocin	Biosyntéza
		Oxidoreduktáza		Bakteriocin	Biosyntéza
Klastr č. 6 (20 931)	22	FMN vazebný protein s flavin reduktázovou doménou		Terpen	Biosyntéza
		15-cis-fytoen desaturáza		Terpen	Biosyntéza
		Fytoen syntáza		Terpen	Biosyntéza
		Cytochrom P450		Terpen	Biosyntéza
		Alfa-amyláza		Terpen	Biosyntéza
Klastr č. 7 (25 364)	31	3-oxoacyl-ACP syntáza		Resorcinol	Biosyntéza
		Transkripční regulátor GntR rodiny	GO:0003677, GO:0003700, GO:0005622, GO:0006813, GO:0008324, GO:0009058, GO:0030528, GO:0045449	Resorcinol	Regulace
		Glutamin vazebný lipoprotein		Resorcinol	Biosyntéza
		ATP vazebný protein ABC transportéru		Resorcinol	Transport
		Permeáza ABC transportéru		Resorcinol	Transport
Klastr č. 8 (41 165)	35	NADH dehydrogenáza	GO:0003954	Ladderan	Biosyntéza
		Transportér EamA rodiny		Ladderan	Transport
		Beta-ketoacyl syntáza		Ladderan	Biosyntéza

Tab. 10: Klastry genů sekundárního metabolismu nalezené v genomu *J. apiculata* programem antiSMASH verze 4.0.

Klastr č. 8 (41 165)	35	Fosfodiesteráza	GO:0004035	Ladderan	Biosyntéza
		MFS transportér		Ladderan	Transport
		Velká podjednotka acetolaktát syntázy	GO:0003984	Ladderan	Biosyntéza
		Fosfodiesteráza	GO:0004035	Ladderan	Biosyntéza
Klastr č. 9 (14 353)	8	Aminotransferáza		Neribozomální peptid	Biosyntéza
		Syntáza neribozomálního peptidu		Neribozomální peptid	Biosyntéza
		AMP dependentní syntáza		Neribozomální peptid	Biosyntéza

5.1.6 Fylogenomická analýza

Na základě 14 core proteinů vybraných programem AMPHORA2-master byl sestrojen fylogenomický strom metodou maximum likelihood z celkového počtu 131 sinic (viz Obr. 10). Ve fylogenomickém stromu je *J. apiculata* zvýrazněna červenou barvou. Analýzou core proteinů byla zjištěna přítomnost 2 proteinů každého typu v proteomu *J. apiculata*. Podle první sady proteinů *J. apiculata*, označené „*Jacksonvillea apiculata* 2014/77b1“ byla sinice samostatně zařazena do vývojové větve na úroveň sinic řádu Oscillatoriales s 60% podporou. Druhá sada označená „*Jacksonvillea apiculata* 2014/77b2“ také způsobila přiřazení sinice k dalším z řádu Oscillatoriales a jako nejbližšího příbuzného byla k ní přiřazena sinice *Oscillatoria acuminata* PCC 6304, ovšem s nulovou podporou. Nelze tedy přesně určit, jak moc je toto zařazení *J. apiculata* ve fylogenezi sinic správné. Podpora topologie fylogenomického stromu byla provedena bootstrapovou analýzou o 1 000 opakováních, 100% podporu získalo 67 uzlů. *J. apiculata* byla nalezena v sedimentu subtropického jezera, je tedy sladkovodní sinicí. Vzorek „*Jacksonvillea apiculata* 2014/77b1“ byl zařazen na úroveň 2 mořských a jedné půdní sinice. Vzorek „*Jacksonvillea apiculata* 2014/77b2“ byl zařazen do vývojové linie, která obsahuje zástupce obývající sladké, slané vody i půdy.



Obr. 10: Fylogenomický strom sestrojžený metodou ML na základě 14 proteinů ze 131 proteomů sinic. Číselné označení uzlů představuje procentuální podporu topologie stromu na základě bootstrapové analýzy s 1 000 opakováními, hvězdičky označují 100% podporu. Měřítko na obrázku nepředstavuje tempo evoluce. *J. apiculata* je v obrázku zvýrazněna červeně.

5.2 *Synechococcus* sp. S4

5.2.1 Obecné informace o genomu

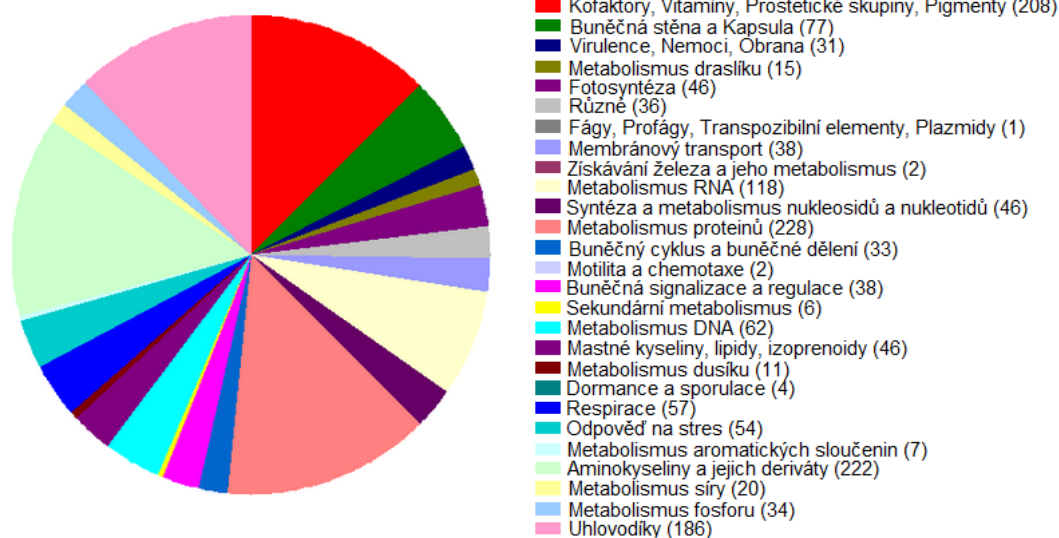
Synechococcus sp. S4 je epizoickou nebo endozoickou sinicí. Pomocí nástroje RAST byla provedena anotace kódujících sekvencí. Sekvenční data byla tvořena jedním kontigem, protože genom byl sestaven podle referenční sekvence *C. gracile* PCC 6307. Velikost genomu byla stanovena na 2 662 774 Mb, s 70,6% obsahem GC bází. V sekvenci bylo nalezeno 2 672 protein kódujících sekvencí a 64 různých RNA. Do subsystémů bylo zařazeno celkem 42 % ze všech kódujících sekvencí. Funkce byla identifikována u 1 757 sekvencí, hypotetických zůstalo 915. Celkový přehled byl zapsán do Tab. 11. Zastoupení anotovaných kódujících sekvencí v subsystémech a jejich kategorie jsou na Obr. 11. Největší počet genů byl zařazen do subsystému Metabolismus proteinů (228). Další 2 nejvíce zastoupené kategorie s více než 200 geny byly Aminokyseliny a jejich deriváty (222) a Kofaktory, Vitamíny, Prostetické skupiny, Pigmenty (208). Nejméně zastoupených subsystémů s počtem genů nižší než 10 bylo 6. Jednalo se o kategorie Metabolismus aromatických sloučenin (7), Sekundární metabolismus (6), Dormance a sporulace (4), Motilita a chemotaxe (2), Získávání železa a jeho metabolismus (2), Fágy, Profágy, Transpozibilní elementy, Plazmidy (1).

Tab. 11: Přehled zařazení protein kódujících sekvencí sinice *Synechococcus* sp. S4 nástrojem RAST do subsystémů.

Celkový počet protein kódujících sekvencí	Zařazeno do subsystému		Nezařazeno do subsystému	
	Identifikovaných	Hypotetických	Identifikovaných	Hypotetických
2 672	1 131 (42 %)		1 541 (58 %)	
	1 090	41	667	874

Distribuce v subsystémech

Kategorie subsystémů



Obr. 11: Zastoupení anotovaných genů *Synechococcus* sp. S4 v subsystémech a jejich kategorie podle RASTu.

5.2.2 Anotace repetitivních sekvencí

Pomocí programu RepeatMasker verze 4.0.6 bylo v sekvenčních datech *Synechococcus* sp. S4 nalezeno 21 různých repetitivních sekvencí na základě jejich porovnání s repetitivy zapsanými v databázi RMBASE pro *Arabidopsis thaliana*. Byly nalezeny 2 části LTR retroelementu ATCOPIA82 z nadrodiny Ty1/Copia. Dále byly nalezeny 2 DNA transpozony z nadrodiny Hobo-activator. Ze skupiny malých RNA bylo nalezeno 14 repetitivních sekvencí. Jednalo se o 11 sekvencí pseudogenů rRNA velké podjednotky ribozomů (LSU-rRNA-Ath) a 3 sekvence pseudogenů rRNA malé podjednotky ribozomů (SSU-rRNA-Ath). Nakonec byly nalezeny 3 repetice ze skupiny jednoduchých repetitiv. Konkrétní nalezené repetice jsou zapsány v Tab. 12 na základě dostupných informací v databázi RMBASE.

Tab. 12: Repetitivní sekvence nalezené v genomu *Synechococcus* sp. S4 programem RepeatMasker verze 4.6.0.

Typ repetice	Nadrodina	Rodina (počet nalezených repetic)	Popis repetice
Retrotranspozony	Ty1/Copia	ATCopia82_I-int (2)	blíže neidentifikovaná část vnitřní oblasti LTR-retrotranspononu ATCOPIA82
DNA transpozony	Hobo-Activator	hAT-N4_Aly (2)	hobo-Activator DNA transpon identifikovaný z <i>Arabidopsis lyrata</i>
Typ repetice	Repetice (počet nalezených)		Popis repetice
Malé RNA	LSU-rRNA_Ath (11)		pseudogen rRNA velké podjednotky ribozomů
	SSU-rRNA_Ath (3)		pseudogen rRNA malé podjednotky ribozomů
Jednoduché repetice	(CCG) _n (2)		Opakující se sekvence CCG
	(CGG) _n (1)		Opakující se sekvence CGG

Porovnáním sekvence DNA *Synechococcus* sp. S4 programem BLAST s databází ISfinder byly nalezeny 2 přítomné inzerční elementy produkující významnou shodu. Jednalo se o inzerční element ISShes11 původně z organismu *Shewanella* sp., a o ISAcba1 z neklasifikované bakterie z kmene Aktinobakterií (viz Tab. 13).

Tab. 13: Inzerční elementy nalezené v genomu *Synechococcus* sp. S4 programem BLAST v databázi ISfinder.

Inzerční element	Skupina	Rodina	Identifikační značka v databázi GenBank	Původní organismus
ISShes11		Tn3	NC_008573	<i>Shewanella</i> sp. ANA-3
ISAcba1	ISSod11	IS1595	MKSW01000042	Actinobacteria bacterium 69-20

Pomocí programu CRISPRfinder nebyly nalezeny žádné potvrzené CRISPRy v genomu *Synechococcus* sp. S4, bylo pouze nalezeno 6 sporných CRISPRů. Délka opakovaných motivů se pohybovala mezi 23-31 nukleotidy. Všechny tyto sporné CRISPRy obsahovaly shodně po 1 spaceru (viz Tab. 14).

Tab. 14: Sporné CRISPRy nalezené v genomu *Synechococcus* sp. S4 programem CRISPRfinder.

Opakovaný motiv	Délka přímé repetice (nt)	Délka CRISPRu (nt)	Počet spacerů mezi motivy
GGAAACGGCTGCGTTGGAAACCG	23	79	1
CTCGGGGGAGACTTGAACCTCCCGACC	26	95	1
TGGGACCCGGCTCGGAGCGGTCAG	24	93	1
TCGGAGACGGACCACANGGCGGG	23	71	1
GCGCATCTGTCCATCCCCATCTGGGAGTTC	30	117	1
GCCTTGATCTTCACGCTACTGGCCCTTTCGC	31	92	1

5.2.3 Anotace tRNA

V genomu *Synechococcus* sp. S4 bylo nalezeno 47 různých sekvencí kódujících molekuly tRNA. Nejvyšší počet nalezených sekvencí pro tRNA byl shodný po pěti pro 3 aminokyseliny, Alanin, Arginin a Leucin. Naopak nejnižší počet nalezených sekvencí byl shodný po jedné pro 9 aminokyselin (viz Tab. 15).

Tab. 15: Přehled molekul tRNA pro jednotlivé aminokyseliny nalezených v genomu *Synechococcus* sp. S4 programem tRNA scan-SE.

Aminokyselina	Počet nalezených tRNA	Antikodony (počet)				
Alanin	5	GGC (2)	CGC (1)	TGC (2)		
Arginin	5	ACG (1)	CCG (1)	CCT (1)	TCT (2)	
Leucin	5	GAG (1)	CAG (1)	TAG (1)	CAA (1)	TAA (1)
Valin	4	GAC (1)	CAC (1)	TAC (2)		
Serin	4	GGA (1)	CGA (1)	TGA (1)	GCT (1)	
Prolin	3	GGG (1)	CGG (1)	TGG (1)		
Threonin	3	GGT (1)	CGT (1)	TGT (1)		
Methionin	3	CAT (3)				
Glycin	2	CCC (1)	TCC (1)			
Lysin	2	TTT (2)				
Isoleucin	2	GAT (2)				
Fenylalanin	1	GAA (1)				
Asparagin	1	GTT (1)				
Aspartát	1	GTC (1)				
Glutamát	1	TTC (1)				
Histidin	1	GTG (1)				
Glutamin	1	TTG (1)				
Tyrosin	1	GTA (1)				
Cystein	1	GCA (1)				
Tryptofan	1	CCA (1)				

5.2.4 Ontologie genů

Ze 2 672 protein kódujících sekvencí byla GO analýzou anotována funkce u 741 genů, což je 28 % celého genomu, hypotetických genů zůstalo 801, ty tvoří 30 % genomu. Nejvíce zastoupenou skupinou v první kategorii byla skupina Vnitrobuněčné, do které bylo zařazeno 64 genů, druhou skupinou v pořadí byla skupina Ribozóm s 46 geny (viz Tab. 16). Pouze ve 2 skupinách se nacházel nějaký gen, který nebyl zastoupen v žádné další kategorii, v každé po jednom genu. Ve skupině Membrána to byl gen kódující permeázu z nadrodiny lékových/metabolických transportérů (DMT superfamily). Ve skupině Nedílná součást membrány šlo o gen pro prekursor vnějšího membránového proteinu A.

Tab. 16: Skupiny genů první kategorie anotovaných GO analýzou *Synechococcus* sp. S4.

Identifikační číslo	Počet anotovaných genů	p-value	Skupina
GO:0005622	64	2.64239426309419e-192	Vnitrobuněčné
GO:0005840	46	6.70202709840456e-137	Ribozóm
GO:0016020	20	5.17840826986123e-57	Membrána
GO:0016021	8	1.93961578323872e-20	Nedílná součást membrány
GO:0005694	4	1.15846272386253e-08	Chromozóm
GO:0005737	2	0.00248505302154037	Cytoplazma
GO:0005960	1	0.295151638420477	Komplex štěpení glycinu
GO:0009379	1	0.295151638420477	Komplex helikáz Hollidayova spoje
GO:0009986	1	0.295151638420477	Buněčný povrch

Nejvíce zastoupenou skupinou v druhé kategorii byla výrazně skupina Translace s 52 geny. Z těch bylo 46 zastoupených také ve skupině Ribozóm v první kategorii. Pouze 1 gen ze skupiny Translace nebyl zastoupen v žádné z jiných skupin i kategorií, a to gen pro ribozóm recyklující faktor. Rovněž pouze 1 gen ze skupiny Translace byl zastoupen i v jiné skupině v druhé kategorii. Šlo o gen pro Translační iniciační faktor 2 ze skupiny Iniciace translace. Všechny ostatní skupiny této kategorie obsahovaly méně než 20 genů. Dvě další skupiny v pořadí s nejvyšším počtem genů byly skupiny Transport elektronů s 15 a Transport s 13 geny. V obecnější skupině Transport byl zastoupen jen jeden gen z dalších skupin spojených s transportem, gen pro permeázu ze skupiny Transport draselných iontů. Dále je v této kategorii 14 skupin obsahujících celkem 30 genů spojených se skládáním, úpravami, transportem či metabolismem proteinů (viz Tab. 17 v Příloze).

Ve třetí kategorii GO analýzy bylo genům *Synechococcus* sp. S4 přiřazeno celkem 868 funkcí. Dvěma jasně nejvíce zastoupenými funkčními skupinami zde byly Strukturní složka

ribozómu se 46 geny a Vazba ATP s 39 geny. Ke všem ostatním funkčním skupinám bylo přiřazeno méně než 20 genů. S pouze jedním genem bylo v této kategorii zastoupeno 306 funkčních skupin (viz Tab. 18 v Příloze).

5.2.5 Analýza přítomnosti genů sekundárního metabolismu

Celkem 7 klastrů genů sekundárního metabolismu bylo nalezeno v genomu *Synechococcus* sp. S4 pomocí programu antiSMASH verze 4.0. Čtyři klastry se týkaly metabolismu Bakteriocinu, dva blíže nespécifikovaných terpenů a 1 metabolismu Polyketid syntázy typu III. Z každého klastru bylo identifikováno několik genů spojených se sekundárním metabolismem dané látky kombinací výsledků z antiSMASH, BLASTu a GO analýzy. Žádný z níže uvedených genů ovšem nebyl anotován GO analýzou (viz Tab. 19).

Tab. 19: Klastry genů sekundárního metabolismu nalezené v genomu *Synechococcus* sp. S4 programem antiSMASH verze 4.0.

Klastr (délka v nt)	Celkový počet genů v klastru	Produkty identifikovaných genů spojených se sekundárním metabolismem	GO anotace	Typ metabolitu	Funkce v metabolismu
Klastr č. 1 (41 099)	39	Peptidáza M16		T3pk	Biosyntéza
		Peptidáza M16			Biosyntéza
		MFP podjednotka effluxního transportéru RND rodiny			Transport
		ATP vazebný protein ABC transportéru DevA rodiny			Transport
		Glykosyl transferáza			Biosyntéza
		Naringenin-chalkon syntáza			Biosyntéza
		Flavin dependentní dehydrogenáza			Biosyntéza
		Geranylgeranyl reduktáza			Biosyntéza
		Translační GTPáza TypA			Biosyntéza
		ATP vazebný protein ABC transportéru			Transport
		Glutamyl-tRNA reduktáza			Biosyntéza
		Glukóza-1-fosfát adenyltransferáza			Biosyntéza
		NADP ⁺ dependentní fosfoglukonát dehydrogenáza			Biosyntéza
Klastr č. 2 (10 226)	13	Acyl-CoA syntetáza		Bakteriocin	Biosyntéza
		Protein fixující dusík			Biosyntéza

Tab. 19: Klastry genů sekundárního metabolismu nalezené v genomu *Synechococcus* sp. S4 programem antiSMASH verze 4.0.

Klastr č. 3 (10 256)	8	Riboflavin syntetizující protein RibD		Bakteriocin	Biosyntéza
		Bakteriocin			Biosyntéza
Klastr č. 4 (10 166)	11	Bakteriocin		Bakteriocin	Biosyntéza
Klastr č. 5 (20 931)	19	Regulátor transkripce rodiny LysR		Terpeny	Regulace
		15-cis-fytoen desaturáza			Biosyntéza
		Fytoen/skvalen syntetáza			Biosyntéza
		Effluxní membránový protein RND rodiny			Transport
		Alfa/beta hydroláza			Biosyntéza
		ATPáza ABC transportéru antimikrobiálních peptidů			Transport
		Permeáza ABC transportéru			Transport
Klastr č. 6 (21 258)	23	Lykopen cykláza		Terpeny	Biosyntéza
		Malát:chinon oxidoreduktáza			Biosyntéza
		Regulační protein zpětného vstřebávání Fe ²⁺ /Zn ²⁺ iontů			Regulace
		cAMP vazebný protein			Regulace
Klastr č. 7 (10 547)	18	ATPáza Mn/Zn ABC transportéru		Bakteriocin	Transport
		Sigma-70 faktor DNA řízené RNA polymerázy			Regulace
		Protein s Nif11/cupid doménou			Biosyntéza
		o-metyltransferáza			Biosyntéza

5.2.6 Analýza horizontálního přenosu genů

Horizontální přenos genů byl u *Synechococca* sp. S4 testován programem SigiHMM. Program rozděluje geny do 3 kategorií:

Normal = normální/domácí gen

Putal = cizí gen

Incon = normální/domácí gen nacházející se uvnitř genomického ostrovu

U *Synechococca* sp. S4 bylo touto analýzou odhaleno 84 potencionálních cizích genů a 2 domácí geny nacházející se uvnitř genomických ostrovů. Na základě identifikace funkce v BLASTu tyto 2 geny kódují 4-hydroxythreonin-4-fosfát dehydrogenázu a blíže neidentifikovanou transpozázu. Mimo geny byly také identifikovány skupiny bakterií, ze kterých mohou pocházet. Celkově nejvyšší počet genů byl identifikován ze skupiny Gammaproteobakterií, nejnižší naopak ze skupiny Bacilů a Betaproteobakterií (viz Tab. 20).

Tab. 20: Počet identifikovaných cizích genů a jejich dárce nalezení v genomu *Synechococcus* sp. S4 programem SigiHMM.

Počet cizích genů	Dárce
30	Gammaproteobakterie
13	Aktinobakterie
6	Sphingobakterie
7	Deltaproteobakterie
4	Nitrospira
4	Alfaproteobakterie
3	Halobakterie
2	Clostridia
2	Deinokoky
2	Metanopyri
1	Bacilli
1	Betaproteobakterie

Z 84 potencionálních cizích genů bylo možno přesně či částečně identifikovat funkci v BLASTu pouze u 24. Zbylých 60 genů je pouze hypotetických. Jediný protein, který byl identifikován více než 1 byl enzym dihydrofolát reduktáza, pokaždé od jiného potencionálního dárce. Identifikované geny a jejich potencionální dárce jsou uvedeny v Tab. 21.

Tab 21.: Identifikované cizí geny a jejich potencionální dárcovská skupina bakterií.

Cizí gen	Potencionální dárce
Vankomycin B-typ obranný protein VanW	Gammaproteobakterie
Beta-karoten ketoláza (EC 1.14.-.-)	Deltaproteobakterie
Necharakterizovaný ycf64 protein podobný monothiol glutaredoxinu	Gammaproteobakterie
Mg ²⁺ transportující protein C ATPáza	Gammaproteobakterie
Alfa-alfa-trehalózafosfát syntáza (UDP tvořící)	Sphingobakterie
Sacharózafosfát syntáza	
Možný světlem inducibilní protein	Aktinobakterie
Hlavní membránový protein I	Gammaproteobakterie
Sodík/glutamát symportér	Gammaproteobakterie
Transkripční regulátor GntR rodiny	

Tab 21.: Identifikované cizí geny a jejich potencionální dárcovská skupina bakterií.

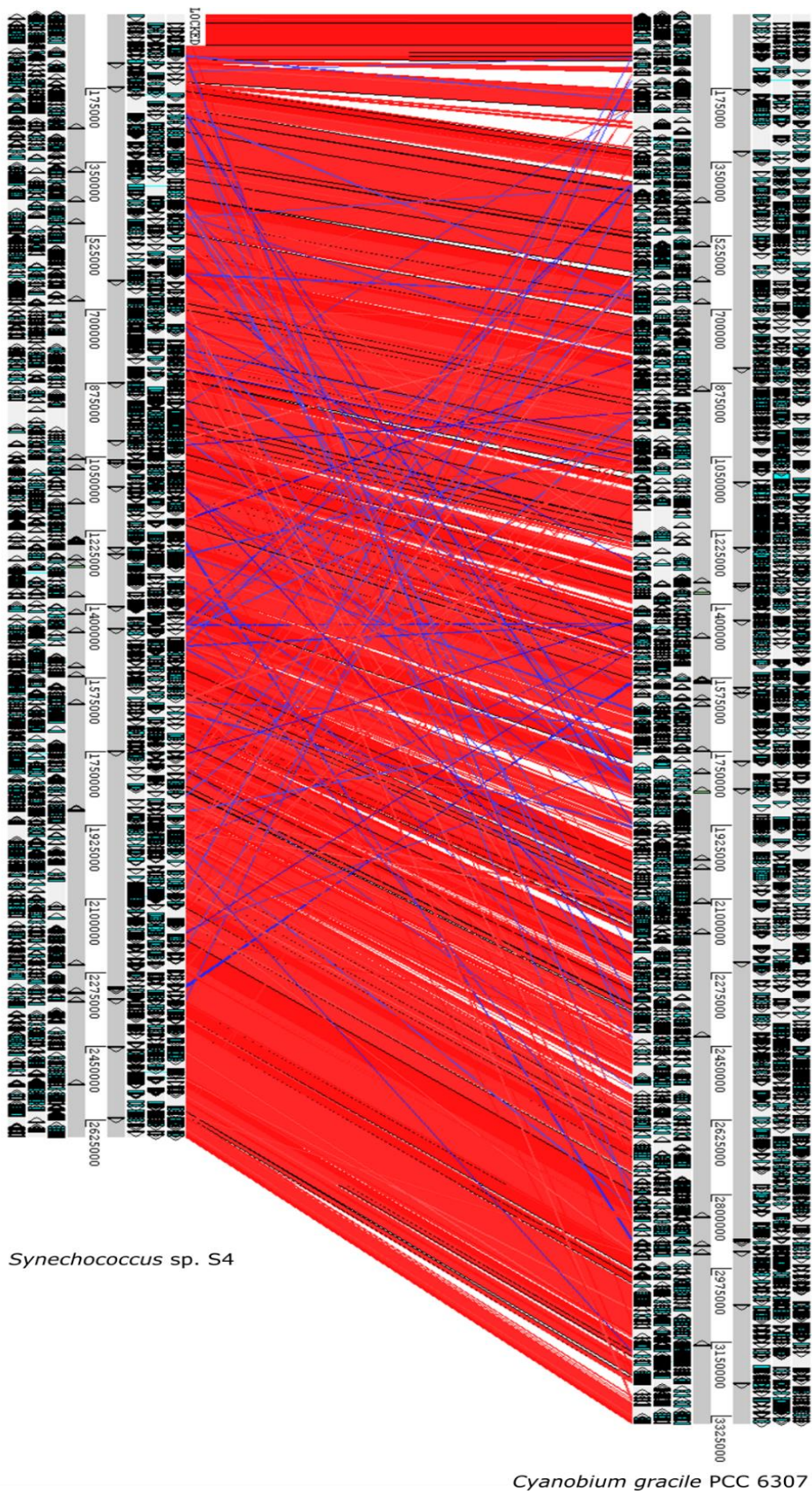
Biotin karboxyl nesoucí protein acetylCoA karboxylázy	Aktinobakterie
Alfa podjednotka receptorového proteinu FtsY (TC 3.A.5.1.1)	
Dihydrofolát reduktáza (EC 1.5.1.3)	Sphingobakterie
Dihydrofolát reduktáza (EC 1.5.1.3)	Gammaproteobakterie
Protein z rodiny dienelakton hydroláz	Gammaproteobakterie
Integráza IntIPac	Gammaproteobakterie
Dioxygenáová doména Glyoxaláza/Bleomycin obranného proteinu	Gammaproteobakterie
Membránový protein	
Oktaprenyl-difosfát syntáza (EC 2.5.1.-) / Dimetylallyltransferáza (EC 2.5.1.1) / Geranyltransferáza (farnesyldifosfát syntáza) (EC 2.5.1.10) / Geranylgeranyl pyrofosfát syntetáza (EC 2.5.1.29)	Gammaproteobakterie
Sacharopin dehydrogenáza	Gammaproteobakterie
Adenylát cykláza (EC 4.6.1.1) / Guanylát cykláza (EC 4.6.1.2)	
Protein PE rodiny	Aktinobakterie
Prekurzor hydroxyprolin bohatého glykoproteinu 1 buněčné stěny	Aktinobakterie
Olovo, kadmium, zinek a rtuť transportující ATPáza (EC 3.6.3.3) (EC 3.6.3.5); Měď-translokační P-tyt ATPáza (EC 3.6.3.4)	
Transkripční regulátor MerR rodiny	Gammaproteobakterie

5.2.7 Analýza přítomnosti delecí, inverzí

Na základě celogenomového porovnání byly v programu ACT manuálně prohledány genomy *Synechococca* sp. S4 a *C. gracile* PCC 6307 na přítomnost deletovaných a invertovaných úseků (viz Obr. 12). V genomu *Synechococca* bylo nalezeno celkem 229 delecí různé délky od desítek po desetitisíce bází. Nejdelší deletovaný úsek měl 47 444 bp a obsahoval 12 identifikovaných a 5 hypotetických genů (viz Tab. 22). Bylo nalezeno celkem 73 invertovaných úseků různé délky. Procentuální shoda nalezených inverzí se pohybovala od 72 do 100 %. Mezi těmito dvěma genomy byly také nalezeny 4 zcela identické inverze. Dále byly v genomu *Synechococca* 2x nalezeny 2 invertované sekvence, které odpovídaly pouze 1 sekvenci v genomu *C. gracile*. Procentuální shoda těchto sekvencí se pohybovala od 79 do 81 %. Naopak v genomu *C. gracile* byly také nalezeny 2 sekvence odpovídající jedné inverzi v genomu *Synechococca*. Zde byla procentuální shoda mezi 86 a 87 %.

Tab. 22: Geny z vybrané deletované oblasti genomu *C. gracile*.

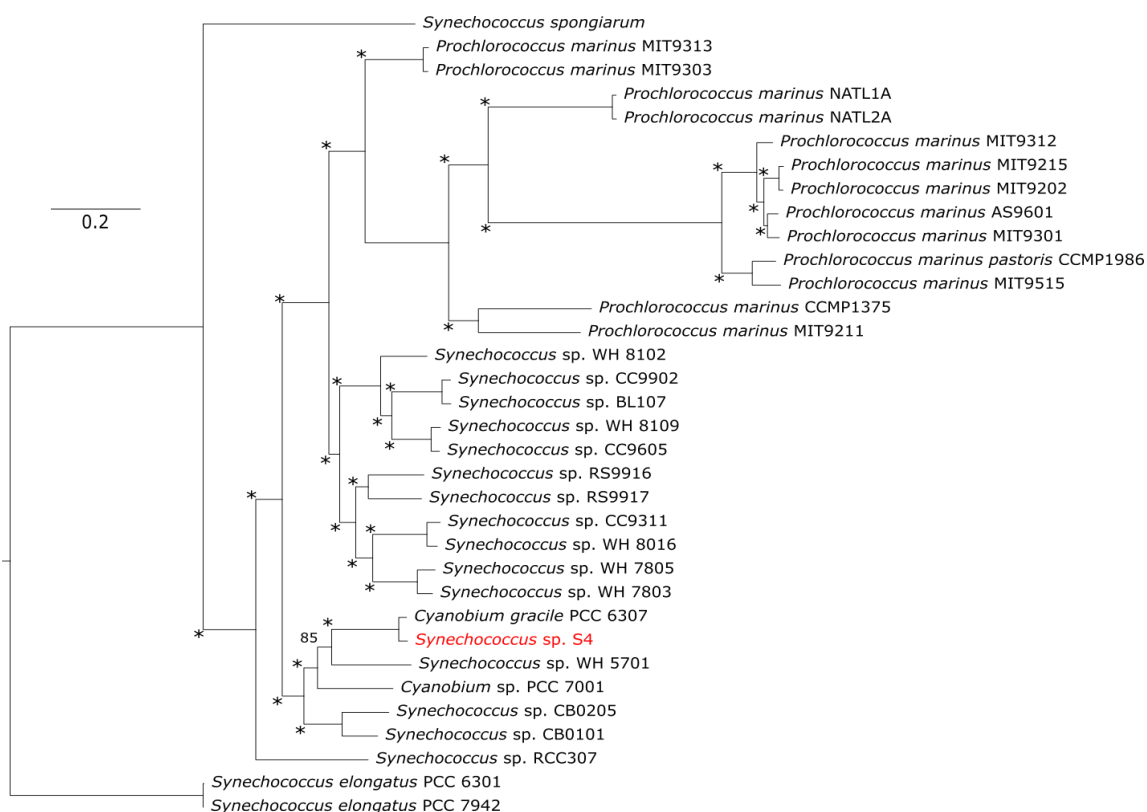
Deletovaná oblast v sekvenci <i>Cyanobium gracile</i> PCC 6307	
267869-315312 (47444 bp)	
Identifikovaný gen	Pozice a délka genu
ATP vazebný protein ABC transportéru pro methionin	272803-275154 (2352 bp)
Neidentifikovaný protein mobilního elementu	278367-279488 (1122 bp)
Neidentifikovaný protein mobilního elementu	279525-281114 (1590 bp)
Exportér ABC typu pro bakteriocin/antibiotika	284785-286647 (1863 bp)
Membrane-fusion protein	286652-287794 (1143 bp)
Exportér ABC typu pro bakteriocin/antibiotika	287921-289945 (2025 bp)
Neidentifikovaný protein mobilního elementu	Spojený gen (291518-291640,291640-291864, 291868-292623) (123+225+756 bp)
Neidentifikovaný protein mobilního elementu	292681-294306 (1626 bp)
Neidentifikovaný protein mobilního elementu	Spojený gen (294340-294600,294600-294800,294804-294878) (261+201+75 bp)
Neidentifikovaný protein mobilního elementu	Spojený gen (299132-299872,299874-300455) (741+582 bp)
Adenylát cykláza (EC 4.6.1.1)	305706-307919 (2214 bp)
Protein podobný lehkému řetězci kinesinu	310498-313443 (2946 bp)



Obr. 12: Porovnání genomů *Synechococcus* sp. S4 a *C. gracile* PCC 6307 v programu ACT. Červenou barvou jsou znázorněny shodující se úseky, modrou inverze, bílá místa reprezentují neshodující úseky.

5.2.8 Fylogenomická analýza

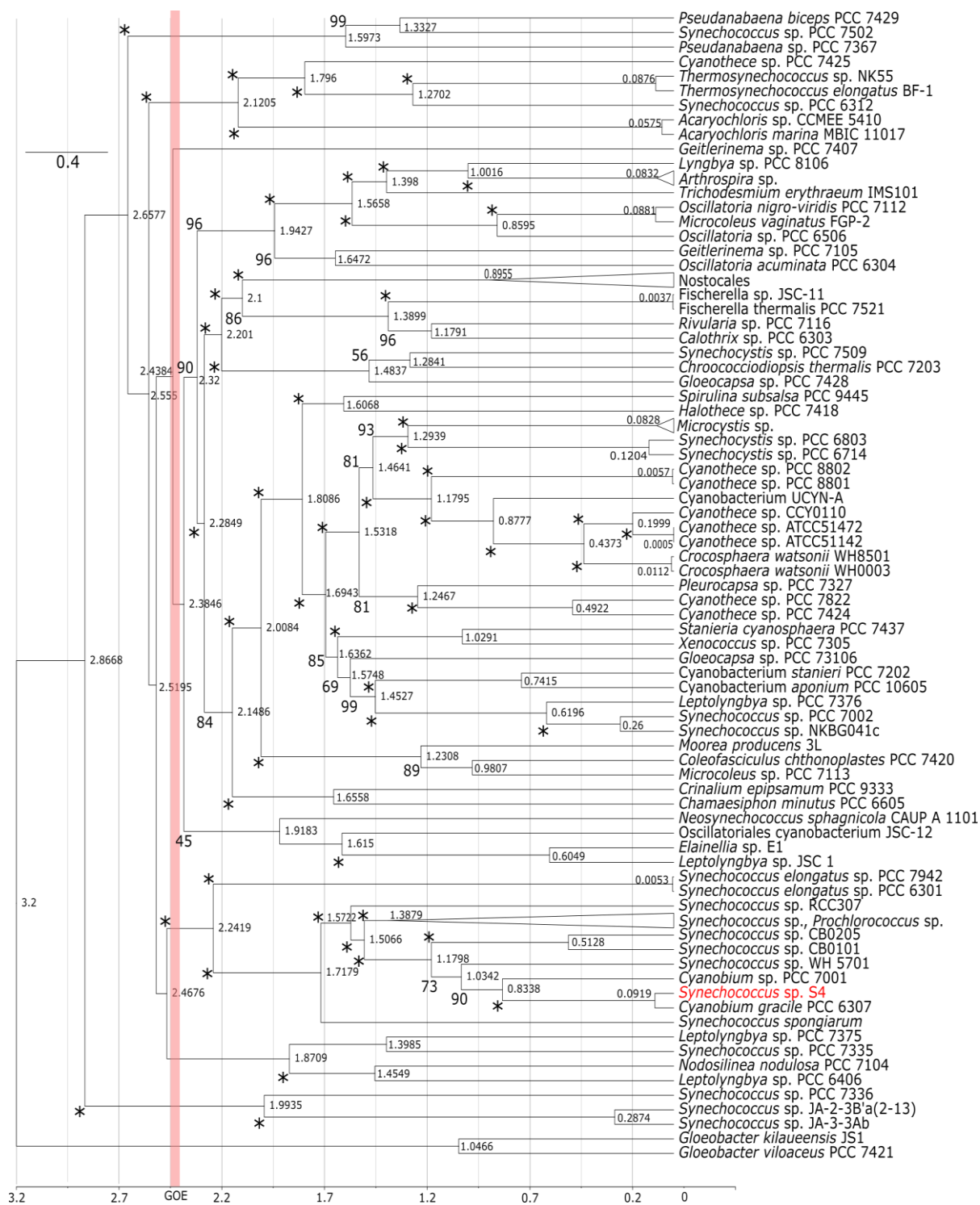
Na základě 482 ortologních genů vybraných programem Hal byl sestrojen fylogenomický strom metodou maximum likelihood z proteomů 34 sinic vybraných z řádu Synechococcales. Strom byl rozdělen do 5 klastrů a vnější skupiny, do které patřily *Synechococcus elongatus* kmeny PCC 6301 a PCC 7942. Ancestrální skupinu v tomto stromu tvořil *Synechococcus spongiarum*. V průběhu evoluce došlo celkem 3x k vývinu mořského a sladkovodního druhu *Synechococca*. Nejrecentněji odštěpenou linií byla monofyletická linie sinice *Prochlorococcus*. *Synechococcus* sp. S4 je ve fylogenomickém stromu zvýrazněn červenou barvou a jako nejbližší příbuzný druh byl k němu přiřazen *Cyanobium gracile* PCC 6307, volně žijící sladkovodní druh sinice, se 100% podporou bootstrapové analýzy. Další nejbližší sousedi *Synechococca* sp. S4 v této vývojové větvi byly mořské sinice *Synechococcus* sp. WH 5701 a *Cyanobium* sp. PCC 7001. Kromě jednoho uzlu, který měl bootstrapovou podporu 85 %, byla topologie tohoto fylogenomického stromu podpořena 100% ve všech uzlech (viz Obr. 13).



Obr. 13: Fylogenomický strom sestavený metodou ML na základě 482 ortologních genů ze 34 proteomů sinic. Číselné označení uzlů představuje procentuální podporu topologie stromu na základě bootstrapové analýzy s 1000 opakováními, hvězdičky označují 100% podporu. Měřítko na obrázku nepředstavuje tempo evoluce. *Synechococcus* sp. S4 je zvýrazněn červeně.

5.2.9 Datování fylogeneze *Synechococcus* sp. S4

Pro datovací analýzu byl použit fylogenomický strom sestrojený metodou maximum likelihood na základě 66 ortologních genů vybraných programem Hal, ze 131 sinic. Tento strom byl v programu r8s verze 1.70 přetvořen na chronogram na základě kalibrace 3 uzlů (viz kapitola Materiál a metody). Nejstarší odštěpená linie sinic *Synechococcus* sp., která se oddělila před téměř 2,9 mld. let obsahovala kmeny PCC 7336, JA-2-3B'a (2-13) a JA-3-3Ab. Asi o 200 milionů let se oddělily vláknité sinice rodu *Pseudanabaena*. Následně se objevovaly další vláknité a kokální sinice. Celý řád Nostocales tvořil monofyletickou skupinu sinic. Velká vývojová větev, která vedla k vývoji *Prochlorococců* a *Synechococců* zahrnující i zkoumaný druh *Synechococcus* sp. S4 se oddělila před 2,5 mld. let, což je těsně před GOE. Poté docházelo k rapidnějšímu odštěpování nových linií. Na základě této analýzy bylo odhadnuto stáří sinice *Synechococcus* sp. S4 na 0,09 mld., tedy 90 milionů let (viz Obr. 14).



Obr. 14: Chronogram sestavený metodou ML na základě 66 ortologních genů ze 131 proteomů sinic. Číselné označení uzlů představuje procentuální podporu topologie stromu na základě bootstrapové analýzy s 1 000 opakováními, hvězdičky označují 100% podporu. Číselné označení větvíčích míst znázorňuje jejich stáří v jednotkách miliard let. Měřítko na obrázku představuje 0,4 mld. let. *Synechococcus* sp. S4 je vyznačen červeně. GOE je vyznačeno červeným pruhem.

6 Diskuze

6.1 *Jacksonvillea apiculata*

Na základě fylogenomické analýzy nebylo možné určit přesné zařazení sinice *J. apiculata* do fylogeneze sinic. Genom obsahoval 2 sady vybraných proteinů detekovaných programem AMPHORA2-master, které patří do základní sady pro sinice. Tyto sady byly následně manuálně rozděleny. Ve fylogenezi byly obě zasazeny do řádu Oscillatoriales. Je pravděpodobné, že tento genom byl kontaminován sinicí z tohoto řádu. Bližší určení kontaminantu ovšem nebylo možné z důvodu nepřítomnosti příbuzného druhu sinice v databázi GenBank. Jednalo se tedy o kombinaci genomů 2 zcela neznámých sinic. Později byla *J. apiculata* popsána Hašlerem *et al.* (2017). Ti zkonstruovali fylogenetický strom *J. apiculata* na základě sekvencí 16S rRNA. Podle jejich výsledku fylogenetické analýzy byla *J. apiculata* zařazena k sinicím rodu *Desertifilum*, které v této práci nebyly zahrnuty, protože v době analýzy ještě nebyly do databáze GenBank přidány sekvence těchto sinic. Rod *Desertifilum* je také součástí řádu Oscillatoriales (Dadheech *et al.*, 2012). Lze tedy potvrdit, že zařazení *J. apiculata* do fylogeneze sinic v této práci bylo úspěšné alespoň na úrovni řádu.

Vzhledem k tomu, že přiřazení vzorku *Jacksonvillea apiculata* 2014/77b2 k *O. acuminata* bylo bez bootstrapové podpory, pro obrazné porovnání molekulárních vlastností genomu byly vybrány jiné sinice z řádu Oscillatoriales, a to *Moorea producens* 3L a *Microcoleus* sp. PCC 7113. V porovnání s těmito dvěma sinicemi *J. apiculata* má navrch co do velikosti genomu, která je 10,7 Mb, počtu protein kódujících sekvencí i počtu nalezených tRNA, který byl téměř dvojnásobný oproti ostatním sinicím. Tento počet vůbec neodpovídá standardu, který lze nalézt v genomech sinic (Chan *et Lowe*, 2009). Je možné předpokládat, že tento vysoký počet má na svědomí kontaminant. Velikost genomu *J. apiculata* se blíží největšímu známému genomu sinice *Calothrix* sp. PCC 7103 o velikosti 11,6 Mb (Shih *et al.*, 2012). V obsahu GC bází byly všechny porovnávané sinice téměř shodné, pohyboval se mezi 44-46 %, prostředí výskytu se u lišilo u každé sinice (viz Tab. 23).

Tab. 23: Srovnání obecných vlastností genomu *J. apiculata*, *M. producens* a *Microcoleus* sp.

Organismus	<i>Jacksonvillea apiculata</i> 2014/77b	<i>Moorea producens</i> 3L	<i>Microcoleus</i> sp. PCC 7113
Velikost genomu (Mb)	10,7	8,5	7,5
Obsah GC (%)	44	44	46
Počet protein kódujících sekvencí	10 096	9 231	6 972
tRNA	116	56	69
Prostředí výskytu	Sladká voda	Slaná voda	Půda

V programu Repeatmasker verze 4.0.6 byly porovnány nalezené repetitivní sekvence ve všech genomech. V genomech sinic byly již dříve identifikovány nejčastěji krátké tandemové repetice (Mazel *et al.*, 1990; Asayama *et al.*, 1995). Z transpozibilní elementů byly v genomech sinic identifikovány malé transpozibilní elementy s invertními repeticemi (MITE) (Zhou *et al.*, 2008) a inzerční elementy (Lin *et al.*, 2010). *J. apiculata* měla ve své sekvenci nejvyšší počet retroelementů oproti ostatním genomům. Nejvíce retroelementů bylo z nadrodiny Ty1/Copia (Kumar *et al.*, 1997). Ty byly nejčastější i v obou dalších sinicích. V počtu nalezených DNA transpozonů jasně převyšovala ostatní sinice *M. producens*, za to ale měla nejnižší počet malých RNA, kde tentokrát nejvyšší počet měl *Microcoleus*. Počet nalezených jednoduchých repetic byl téměř totožný u *J. apiculata* i *M. producens*. Obecně nejzastoupenějšími jednoduchými repeticemi u sinic bývají tetranukleotidové repetice a často jsou zasazovány do kódujících regionů. Maximální počet jednoduchých repetic u sinic byl vyhodnocen u *C. gracile* PCC 6307, 1 repetice na 0,46 kbp a minimální u *Acaryochloris marina* 1 repetice na 2,02 kbp (Kabra *et al.*, 2016). U *J. apiculata* ovšem byly nejvíce zastoupené trinukleotidové (viz Tab. 3). Zatímco genom *J. apiculata* neobsahoval žádné inzerční elementy, u ostatních sinic se počet inzerčních elementů pohyboval od 15 do 18. Nálezy inzerčních elementů v genomech sinic jsou velmi variabilní co do počtu a délky (Lin *et al.*, 2010). Mezi porovnávanými genomy nebyla nalezena žádná korelace v obsahu CRISPRů. Ty se vyskytují ve většině známých rodech sinic, výjimkou jsou třeba rody *Prochlorococcus* a *Synechococcus*. Průměrně sinice obsahují méně než 10 CRISPRů na genom (Cai *et al.*, 2013). Největší počet potvrzených CRISPRů byl nalezen v genomu *J. apiculata* s výrazně větším množstvím spacerů než u ostatních sinic. Naopak byl v tomto genomu nalezen nejnižší počet sporných CRISPRů. Zde obě další sinice výrazně převyšovala *M. producens* (viz Tab. 24).

Tab. 24: Srovnání repetitivních sekvencí nalezených v genomech *J. apiculata*, *M. producens* a *Microcoleus* sp.

Organismus	<i>Jacksonvillea apiculata</i> 2014/77b	<i>Moorea producens</i> 3L	<i>Microcoleus</i> sp. PCC 7113
Počet retroelementů	10	5	7
Počet DNA transpozonů	14	22	7
Inzerční elementy	0	15	18
Počet malých RNA	15	6	24
Počet jednoduchých repetit	21	23	14
Počet potvrzených CRISPRů	16	5	10
Počet potvrzených spacerů	358	19	72
Počet sporných CRISPRů	17	90	22
Počet sporných spacerů	24	107	27

Porovnáním výsledků analýzy ontologie genů vybraných kategorií bylo zjištěno, že *M. producens* i *Microcoleus* sp. jsou v počtech anotovaných genů celkem konstantní. Za to *J. apiculata* obě sinice výrazně převyšovala počtem genů ve všech vybraných skupinách, možná právě z důvodu kontaminace. Nejvyšší rozdíly jsou viditelné ve skupinách Translace a Vazba ATP, u kterých *J. apiculata* dosáhla počtu přes 100 genů. Naopak nejmenší rozdíl mezi genomy je ve skupině Skládání proteinů, kde je rozdíl genů obou sinic oproti *J. apiculata* zhruba třetinový (viz Tab. 25).

Tab. 25: Porovnání vybraných skupin z různých kategorií analýzy ontologie genů u *J. apiculata*, *Microcoleus* sp. a *M. producens*.

Identifikační číslo	<i>Jacksonvillea apiculata</i> 2014/77b	<i>Moorea producens</i> 3L	<i>Microcoleus</i> sp. PCC 7113	Skupina
GO:0006412	110	56	64	Translace
GO:0006810	77	45	59	Transport
GO:0006260	12	4	4	Replikace DNA
GO:0006457	17	12	11	Skládání proteinů
GO:0006281	14	7	7	Reparace DNA
GO:0005524	115	68	66	Vazba ATP
GO:0003735	94	47	49	Strukturní složka ribozómu

Analýzou sekundárního metabolismu bylo nalezeno 9 klastrů genů pro rozličné metabolity. U *Microcolea* byl nalezen shodný počet klastrů jako u *J. apiculata*, *M. producens* v tomto směru výrazně vedla s 33 nalezenými klastry. Téměř polovinu z nalezených klastrů u *J. apiculata* tvořily geny pro metabolismus terpenů, počet těchto klastrů byl téměř shodný

u obou dalších sinic. Dále byly nalezeny 2 klastry genů pro neribozomální peptidy a 1 pro polyketid syntázu. Ty slouží k syntéze mnoha jiných sekundárních metabolitů (Shih *et al.*, 2013). Zde byla *J. apiculata* co do počtu klastrů poslední, těsně za *Microcoleem*. *M. producens* zde opět obě sinice velmi převyšovala. U všech sinic byl také nalezen klastř genů pro Bakteriocin, který se používá jako konzervant či antibiotikum (Galvez *et al.*, 2008; Piper *et al.*, 2009) a jeho výskyt byl již dříve potvrzen u několika druhů sinic (Begley *et al.*, 2009; Goto *et al.*, 2010). Pouze u *J. apiculata* a *M. producens* byl nalezen klastř pro metabolismus Ladderanu, jehož funkce je u mikroorganismů zatím neznámá (Chaban *et al.*, 2014), a nakonec pouze u *J. apiculata* klastř pro metabolismus Resorcinolu, který negativně ovlivňuje růst sinic (Nakai *et al.*, 2008) (viz Tab. 26).

Tab. 26: Porovnání počtu nalezených klastrů genů sekundárního metabolismu u *J. apiculata*, *M. producens* a *Microcoleus* sp.

Sekundární metabolity	<i>Jacksonvillea apiculata</i> 2014/77b	<i>Moorea producens</i> 3L	<i>Microcoleus</i> sp. PCC 7113
Terpeny	4	4	2
Neribozomální peptidy	2	8	3
Bakteriocin	1	6	2
Ladderan	1	1	0
Resorcinol	1	0	0

6.2 *Synechococcus* sp. S4

Na základě fylogenomické analýzy byla určena jako nejbližší příbuzný tohoto druhu volně žijící sinice *Cyanobium gracile* PCC 6307, která při srovnání obecných vlastností genomu měla genom větší o necelých 700 kb. S tím souvisel i vyšší počet protein kódujících sekvencí u *C. gracile*. Obsah GC báží byl u obou genomů velmi podobný, lišil se o necelé 2 %. Co se týče tRNA sekvencí, v genomu *Synechococcus* sp. S4 se nacházelo 47 sekvencí kódujících tRNA, v genomu *C. gracile* to bylo 43 tRNA sekvencí (viz Tab. 27).

Tab. 27: Srovnání obecných vlastností genomu *C. gracile* PCC 6307 a *Synechococcus* sp. S4

Organismus	<i>Cyanobium gracile</i> PCC 6307	<i>Synechococcus</i> sp. S4
Velikost genomu (Mb)	3,3	2,7
Obsah GC (%)	68,7	70,6
Počet protein kódujících sekvencí	3 094	2 672
tRNA	43	47
Prostředí výskytu	Sladká voda	Epi/endozoické (<i>Paramecium</i> sp.)

V programu Repeatmasker verze 4.0.6 byly porovnány repetitivní sekvence obou organismů. Na rozdíl od *Synechococca* u *C. gracile* nebyl nalezen ani jeden retrotranspozon, ani jeden DNA transpozon. Bylo nalezeno 6 artefaktů inzerčního elementu IS5 a 5 artefaktů inzerčního elementu IS3, dále 12 sekvencí pseudogenů rRNA velké podjednotky ribozomů (LSU-rRNA-Ath) a 2 sekvence pseudogenů rRNA malé podjednotky ribozomů (SSU-rRNA-Ath). Dále nebyl nalezen žádný potvrzený i sporný CRISPR. Také nebyl nalezen žádný z inzerčních elementů. V obou genomech byly ovšem nalezeny stejné typy i počty jednoduchých repetic. Podle Kabra *et al.* 2016 byl ovšem u *C. gracile* PCC 6307 vyhodnocen nejvyšší počet jednoduchých repetic ze všech známých sinic, 805 v celém genomu při průměrném počtu 1 repetice na 0,46 kbp. K identifikaci používali ovšem jiný software, specializovaný na vyhledávání jednoduchých repetic, MISA (Microsatellite Identification Tool) (<http://pgrc.ipk-gatersleben.de/misa/misa.html>). Shrnutí porovnání nalezených repetitivních sekvencí je v Tab. 28.

Tab. 28: Porovnání repetitivních sekvencí nacházejících se v genomech *Synechococcus* sp. S4 a *C. gracile* PCC 6307.

Organismus	<i>Cyanobium gracile</i> PCC 6307	<i>Synechococcus</i> sp. S4
Počet retroelementů	0	2
Počet DNA transpozonů	0	2
Inzerční elementy	11	2
Počet malých RNA	14	14
Počet jednoduchých repetic	3	3
Počet potvrzených CRISPRů	0	0
Počet potvrzených spacerů	0	0
Počet sporných CRISPRů	0	6
Počet sporných spacerů	0	6

Výsledky analýzy genové ontologie u obou genomů byly porovnány za účelem zjištění nepřítomných genů, které mohly být u *Synechococca* sp. S4 deletovány v rámci redukce genomu. U *Synechococca* bylo anotováno 741 genů, u *C. gracile* jen 657 genů. U obou sinic byly nalezeny nějaké skupiny s vyšším zastoupením genů v různých kategoriích oproti té druhé. Skupin s vyšším zastoupením genů u *Synechococca* bylo pouze 17, kdežto skupin s vyšším zastoupením u *C. gracile* bylo 61. V případě vyššího zastoupení genů u *Synechococca* šlo většinou o rozdíl pouhého 1 genu, kdežto v opačném případě se rozdíl pohyboval mezi 1-4 geny, což u 61 skupin celkem čítalo rozdíl 88 genů. Nejvíce skupin

s vyšším zastoupením u *C. gracile* se týkalo translace a transportu molekul, a na druhé straně u *Synechococca* šlo o transport specifických iontů. Do skupin s vyšším zastoupením byly zahrnuty také skupiny s nulovým zastoupením genů. Těch bylo u *C. gracile* podstatně méně (4) oproti *Synechococcus*, který měl nulové zastoupení v 21 skupinách. Z těchto 21 skupin se nejvíce týkalo metabolismu a transportu molekul. Konkrétní skupiny jsou vypsány v Tab. 29 a v Tab. 30 v Příloze.

U analýzy horizontálního přenosu genů byly porovnány nalezené cizí geny v genomu *Synechococca* s cizími geny v *C. gracile*. U 10 nalezených a identifikovaných cizích genů byla potvrzena stejná funkce v obou genomech. Pouze u 3 z těchto genů byl také potvrzen stejný potenciální dárcovský organismus. Dále byla identifikována stejná funkce u 11 dalších genů, ale zatímco tyto geny byly u *Synechococca* označeny jako cizí, u *C. gracile* byly označeny jako normální/domácí geny. Těchto 11 genů bylo proto vyhledáno také v dalších 2 nejbližších sinicích v rámci fylogeneze, a to u *Synechococca* sp. CB0205 a *Cyanobium* sp. PCC 7001 (viz Obr. 13). U 2 genů byl potvrzen cizí původ i funkce u *Synechococca* sp. CB0205 a u 1 genu byl také potvrzen cizí původ i funkce u *Cyanobium* sp. PCC 7001. Nakonec bylo nalezeno 12 hypotetických genů, u kterých byl určen cizí původ jak u *Synechococca* sp. S4 tak u *C. gracile* a u 8 z těchto hypotetických genů byl také určen stejný potenciální dárcovský organismus.

V rámci diskuze deletovaných úseků v genomu *Synechococca* oproti *C. gracile* byly vyhledány geny nacházející se v těchto úsecích a na základě jejich funkce určené BLASTem byly zařazeny do několika základních kategorií. Do kategorií replikace a reparace DNA byl přiřazen shodný počet genů. Z velké části šlo v těchto kategoriích o geny kódující určitý typ helikáz a nukleotidyltransferáz. V kategorii metabolismus byly v největším zastoupení geny kódující produkty účastníci se metabolismu mastných kyselin. Zajímavým genem v této kategorii byl gen *PsaJ* kódující protein IX podjednotky reakčního centra fotosystému I. Jeho funkce je pouze teoretická, pravděpodobně interaguje s dalšími proteiny a pomáhá stabilizovat komplex fotosystému I. Nicméně bylo zjištěno, že absence tohoto genu nijak neovlivňuje efektivitu fotosyntézy (Xu *et al.*, 1994). V kategorii transport se v deletovaných úsecích nacházely především geny, jejichž funkce byla spojena s transportem určitých iontů. Bylo zde nalezeno celkem 7 genů spojených s antiportérem pro protony a sodíkové ionty. V další kategorii týkající se genů spojených s reakcí na změny podmínek se nacházely geny pro Cefalosporin hydroxylázu, histidin kinázu a hlavně geny kódující proteiny z rodiny CsbD

spojené s odpovědí bakterií na stres (Weber *et al.*, 2006). Poslední kategorie zahrnovala geny spojené s biosyntézou molekul. Zde byly nalezeny 2 geny pro glutamát syntázu a gen pro RfbJ protein spojený se syntézou lipopolysacharidů (Jiang *et al.*, 1991) (viz Tab. 29).

Tab. 29: Kategorie a počty genů v deletovaných úsecích *C. gracile* PCC 6307

Kategorie	Replikace DNA	Reparace DNA	Metabolismus	Transport	Reakce na změny podmínek	Biosyntéza
Počet genů v deletovaném úseku	13	13	25	39	13	15

Podobně jako u *Synechococca* sp. S4 byly u *C. gracile* nalezeny 4 klastry genů sekundárního metabolismu související s metabolismem Bakteriocinu, který se používá jako konzervant či antibiotikum (Galvez *et al.*, 2008; Piper *et al.*, 2009). Byly také nalezeny 2 klastry genů pro Polyketid syntázy typu III, které slouží k syntéze mnoha jiných sekundárních metabolitů (Shih *et al.*, 2013) a 2 klastry genů pro metabolismus blíže nespécifikovaných terpenů.

Všechny druhy rodu *Synechococcus* použité v této práci se podle výsledků datování vyvinuly po GOE (viz Obr. 14), což z velké části potvrzuje také Dvořák *et al.* (2014). Na základě datování fylogeneze sinic byl určen původ *Synechococca* sp. S4 do doby před přibližně 90 miliony let. Dá se s jistotou říct, že se jedná o mladý druh. Na místě je samozřejmě nejistota správného určení stáří z důvodu nedostatečných fosilních záznamů (Schopf, 2001). Nejnovější poznatky však řadí původ sinic jako takových do doby před 3,2 mld. let (Satkoski *et al.*, 2015). Datování sinic v této práci je v souladu s dřívějšími publikacemi v tom, že velká většina vývojových linií sinic se vyvinula až po GOE (Shirmermeister *et al.*, 2013, Dvořák *et al.*, 2014).

Vzhledem k faktu, že byla tato sinice izolována z jiného organismu, byly hledány vlastnosti, které by napovídaly, zda není symbiontem daného hostitele, prvoka z rodu *Paramecium*. Oproti jiným známým sinicím s redukovaným genomem, jako UCYN-A (1,4 Mb) (Tripp *et al.*, 2010) a *Prochlorococcus marinus* SS120 (1,7 Mb) (Dufresne *et al.*, 2003) má ovšem *Synechococcus* genom větší o přibližně 1 Mb. Nicméně porovnání výsledků GO analýzy ukázalo vysoký počet funkčních skupin, ve kterých měl *Synechococcus* menší počet genů než jeho nejbližší příbuzný *C. gracile* PCC 6307 (viz Tab. 30 v Příloze). Poměr G:C k A:T bázím také nesedí s informacemi o redukovaných

genomech (Ussery *et al.*, 2004; Dufresne *et al.*, 2003) jelikož u *Synechococca* sp. S4 byl obsah G:C nad 70 %. U symbiontů dochází ke ztrátě genů důležitých k reparaci DNA. U *B. aphidicola* chybí geny *recA* a *lexA*. Oba tyto geny byly v genomu analyzovaného *Synechococca* nalezeny, jejich funkce byla ale potvrzena jen BLASTem a nikoliv analýzou ontologie genů. Nebyly ovšem podobně jako u *B. aphidicola* nalezeny geny *umuCD* a *uvrABC*, které jsou také spojeny s reparací DNA (Shigenobu *et al.*, 2000). Také nebyl nalezen gen *dnaA*, který je regulátorem transkripce. Ten se sice u *B. aphidicola* vyskytuje, ale např. u symbiotické sinice *Nostoc azollae* došlo k jeho pseudogenizaci. Pseudogenizovaným genem u *N. azollae* je také gen *recQ* kódující určitý typ DNA helikázy (Ran *et al.*, 2010). Tento gen chybí například u některých studovaných kmenů *P. marinus* (Dufresne *et al.*, 2005). V genomu *Synechococca* sp. S4 byl tento gen nalezen. Na základě celkové bioinformatické analýzy nelze prokázat, zda u *Synechococca* sp. S4 došlo k redukci genomu a ztrátě některých genů typických pro symbionty z důvodu potenciální symbiózy s prvokem rodu *Paramecium*, ze kterého byl izolován. Dal by se svými vlastnostmi přirovnat k *Synechococcus* sp. WH 8102, který je srovnatelný co do velikosti genomu (2,5 Mb) a počtu protein kódujících sekvencí (2 525). Tento kmen je příbuzný rodu *Prochlorococcus*, který obsahuje volně žijící sinice se silně redukovanými genomy. Dufresne *et al.* (2005) provedli srovnání těchto sinic. *Synechococcus* sp. WH 8102 se podle této publikace také nejvíce blíží *Synechococcus* sp. S4 obsahem G:C bází, který má 59%. V této práci autoři dále hledali geny, jejichž ztráta by mohla způsobit redukci genomu a posun poměru bází směrem k A:T, šlo hlavně o geny spojené s reparací DNA. Na rozdíl od 2 silně redukovaných kmenů *P. marinus* MED4 a SS120, které neobsahovaly všechny nebo některé z identifikovaných genů, *Synechococcus* sp. WH 8102 obsahoval všechny. Dva z genů analyzovaných v této publikaci byly nalezeny i u *Synechococca* sp. S4. Šlo o geny *recQ*, kódující helikázu a *recJ*, kódující DNA specifickou jednovláknovou exonukleázu.

7 Závěr

Teoretická část této diplomové práce pojednávala o fylogenomice a využití celogenomového sekvenování ke studiu evoluce sinic.

Praktická část se zabývala kompletní anotací 2 vybraných genomů sinic od zpracování primárních sekvenačních dat, přes anotaci kódujících regionů, repetitivních sekvencí, tRNA až po následnou fylogenomickou analýzu.

První byl *de novo* sestaven a následně kompletně anotován genom sladkovodní sinice *Jacksonvillea apiculata* o velikosti 10,7 Mb a obsahem G:C bází 44 %. Poté byla u této sinice provedena anotace repetitivních sekvencí, kterých bylo nalezeno více než 60, a jejich porovnání s jinými sinicemi bylo velmi variabilní. Dále bylo nalezeno 116 sekvencí tRNA. Tímto počtem *J. apiculata* vysoce převyšovala ostatní sinice vybrané pro porovnání. Analýzou ontologie genů byla určena funkce u 1 936 genů, které tvořily 19 % celého genomu. U mnoha skupin *J. apiculata* téměř dvojnásobně převyšovala jiné sinice počtem zařazených genů. Analýzou sekundárního metabolismu bylo nalezeno 9 klastrů genů, které převážně kódovaly produkty pro metabolismus nespécifikovaných terpenů, ale byly také nalezeny geny pro 2 neribozomální peptidy sloužící k syntéze rozličných sekundárních metabolitů. Nakonec byla provedena fylogenomická analýza na základě 14 vybraných genů základních pro sinice. Ta byla úspěšná pouze na úrovni zařazení řádu, a to zřejmě z důvodu kontaminace kultury *J. apiculata* jinou neznámou sinicí. Do budoucna by bylo vhodné porovnat genom *J. apiculata* se sinicemi rodu *Desertifilum*, které byly na základě 16S rRNA přiřazeny jako nejbližší příbuzný rod v jiné práci.

Dále byl na základě referenční sekvence sestaven a kompletně anotován genom sinice *Synechococcus* sp. S4 o velikosti 2,7 Mb s 70% obsahem G:C bází. Rovněž zde byla provedena anotace repetitivních sekvencí i tRNA. Analýzou ontologie genů byla úspěšně určena funkce u 741 genů, což tvoří 28 % genomu a byla zjištěna redukce v počtu mnoha genů oproti nejbližší příbuzné sinici *Cyanobium gracile* PCC 6307. Bylo nalezeno 7 klastrů genů pro sekundární metabolismus rozličných sloučenin jako antibiotikum Bakteriocin, polyketid syntázy, sloužící k syntéze různých sekundárních metabolitů, či nespécifikované terpeny. Analýzou horizontálního přenosu genů bylo určeno na 84 cizích genů, ale funkci se podařilo identifikovat pouze u 24. Dále byly nalezeny invertované a deletované úseky v genomu této sinice oproti jejímu nejbližšímu příbuznému, sinici

Cyanobium gracile PCC 6307, v místech těchto delecí bylo nalezeno mnoho genů spojených ze základními buněčnými procesy. Blízká příbuznost těchto dvou sinic byla potvrzena fylogenomickou analýzou na základě 482 ortologních genů se 100% bootstrapovou podporou. Datovací analýzou bylo určeno stáří *Synechococca* sp. S4 na přibližně 90 milionů let. U této sinice nebyla potvrzena redukce genomu spojená se symbiózou, doprovázená ztrátami několika typických genů a posunu poměru bází směrem k A:T, což může vyvrátit domněnku, že je symbiotickým organismem i přes místo jejího nálezu.

8 **Literatura**

Akman, L.; Yamashita, A.; Watanabe, H.; Oshima, K.; Shiba, T.; Hattori, M.; Aksoy, S. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics* (2002) 32: 402-407

Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* (1990) 215: 403-410

Amard, B.; Bertrand-Sarfati, J. Microfossils in 2000 ma old cherty stromatolites of the Franceville group, Gabon. *Precambrian Research* (1997) 81: 197-221

Anbar, A. D.; Duan, Y.; Lyons, T. W.; Arnold G. L.; Kendall, B.; Creaser, R. A.; Kaufman, A. J.; Gordon, G. W.; Scott, C.; Garvin, J.; Buick, R. A whiff of oxygen before the Great Oxidation Event? *Science* (2007) 317: 1903-1906

Andersson, S. G. E.; Karlberg, O.; Canbäck, B.; Kurland, C. G. On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* (2003) 358: 165-179

Asayama, M.; Kabasawa, M.; Takahashi, I.; Aida, T.; Shirai, M. Highly repetitive sequences and characteristics of genomic DNA in unicellular cyanobacterial strains. *FEMS Microbiology Letters* (1996) 137: 175-181

Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F.; Olsen, G. J.; Olson, R.; Osterman, A. L.; Overbeek, R. A.; McNeil, L. K.; Paarmann, D.; Paczian, T.; Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.; Vonstein, V.; Wilke, A.; Zagnitko, O. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* (2008) 9: 75

Baumann, P.; Moran, N. A.; Baumann, L. The evolution and genetics of aphid endosymbionts. *BioScience* (1997) 47: 12-20

Begley, M.; Cotter, P. D.; Hill, C.; Ross, R. P. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Applied and Environmental Microbiology* (2009) 75: 5451-5460

- Beiderbeck, H.; Taraz, K.; Budzikiewicz, H.; Walsby, A. E. Anachelin, the Siderophore of the Cyanobacterium *Anabaena cylindrica* CCAP 1403/2A. *Zeitschrift für Naturforschung* (2000) 55: 681-687
- Bekker, A.; Holland, H. D.; Wang, P.-L.; Rumble III, D.; Stein, H. J.; Hannah, J. L.; Coetzee, L. L.; Beukes, N. J. Dating the rise of atmospheric oxygen. *Nature* (2004) 427: 117-120
- Bininda-Emonds, O. R. P. The evolution of supertrees. *Phylogenetics series* (2004) 19: 315-322
- Brocks, J. J.; Logan, G. A.; Buick, R.; Summons, R. E. Archaean molecular fossils and the early rise of eukaryotes. *Science* (1999) 285: 1033-1036
- Burja, A. M.; Banaigs, B.; Abou-Mansour, E.; Burgess, J. G.; Wright, P. C. Marine cyanobacteria – a prolific source of natural products. *Tetrahedron* (2001) 57: 9347-9377
- Cai, F.; Axen, S. D.; Kerfeld, C. A. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum *Cyanobacteria*. *RNA Biology* (2013) 10: 687-693
- Calteau, A.; Fewer, F. P.; Latifi, A.; Coursin, T.; Laurent, T.; Jokela, J.; Kerfeld, C. A.; Sivonen, K.; Piel, J. Gugger, M. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics* (2014) 15: 977
- Carver, T. J.; Rutherford, K. M.; Berriman, M.; Rajandream, M. A.; Barrell, B. G.; Parkhill, J. ACT: the Artemis Comparison Tool. *Bioinformatics* (2005) 21: 3422-3423
- Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* (1999) 17: 540-552
- Catling, D. C.; Claire, M. W. How Earth's atmosphere evolved to an oxic state: a status report. *Earth and Planetary Science Letters* (2005) 237: 1-20
- Chaban, V. V.; Nielsen, M. B.; Kopec, W.; Khandelia, H. Insights into the role of cyclic ladderane lipids in bacteria from computer simulations. *Chemistry and Physics of Lipids* (2014) 181: 76-82
- Chan, C. X., Ragan, M. A. Next-generation phylogenomics. *Biology Direct* (2013) 8: 1-6

- Chan, P. P.; Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research* (2009) 37: D93-D97
- Chevreur, B., Wetter, T., Suhai, S. Genome sequence assembly using trace signals and additional sequence information. (1999)
- Dadheech, P. K.; Abed, R. M. M.; Mahmoud, H.; Mohan, M. K.; Krienitz, L. Polyphasic characterization of cyanobacteria isolated from desert crusts, and the description of *Desertifilum tharensense* gen. Et sp. nov. (Oscillatoriales). *Phycologia* (2012) 51: 260-270
- Delcher, A. L.; Harmon, D.; Kasif, S.; White, O.; Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Resolutions* (1999) 27: 4636-4641
- Delsuc, F., Brinkmann, H., Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* (2005) 6: 361-375
- Dittmann, E.; Fewer, D. P.; Neilan, B. A. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiology Reviews* (2012) 37: 23-43
- Dufresne, A.; Salanoubat, M.; Partensky, F.; Artiguenave, F.; Axmann, I. M.; Barbe, V.; Duprat, S.; Galperin, M. Y.; Koonin, E. V.; Le Gall, F.; Makarova, K. S.; Ostrowski, M.; Oztas, S.; Robert, C.; Rogozin, I. B.; Scanlan, D. J.; de Marsac, N. T.; Weissenbach, J.; Wincker, P.; Wolf, Y. I. Hess, W. R. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences of the United States of America* (2003) 100: 10020-10025
- Dufresne, A.; Garczarek, L.; Partensky, F. Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biology* (2005) 6: R14
- Dvořák, P.; Casamatta, D. A.; Pouličková, A.; Hašler, P.; Ondřej, V.; Sanges, R. *Synechococcus*: 3 billion years of global dominance. *Molecular Ecology* (2014) 23: 5538-5551
- Edwards, S. V., Fertil, B., Giron, A., Deschavanne, P. J. A genomic schism in birds revealed by phylogenetic analysis of DNA strings. *Systematic Biology* (2002) 51: 599-613
- Eisen, J. A., Fraser, C. M. Phylogenomics: Intersection of Evolution and Genomics. *Science* (2003) 300: 1706-1707

- Emelyanov, V. V. Mitochondrial connection to the origin of the eukaryotic cell. *The FEBS Journal* (2003) 270: 1599-1618
- Esser, Ch., Ahmadinejad, N., Wiegand, Ch., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., Bryant, D., Steel, M. A., Lockhart, P. J., Penny, D., Martin, W. A genome phylogeny for mitochondria among α -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular Biology and Evolution* (2004) 21: 1643-1660
- Falcon, L. I.; Magallon, S.; Castillo, A. Dating the cyanobacteria ancestor of the chloroplast. *ISME Journal* (2010) 4: 777-783
- Fan, Q.; Huang, G.; Lechno-Yossef, S.; Wolk, C. P.; Kaneko, T.; Tabata, S. Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena* apod. strain PCC 7120. *Molecular Microbiology* (2005) 58: 227-243
- Flombaum, P.; Gallegos, J. L.; Gordillo, R. A.; Rincón, J.; Zabala, L. L.; Jiao, N.; Karl, D. M.; Li, W. K. W.; Lomas, M. W.; Veneziano, D.; Vera, C. S.; Vrugt, J. A.; martiny, A. C. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences* (2013) 110: 9824-9829
- Felsenstein, J. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution* (1981) 17: 368-376
- Galvez, A.; Lopez, R. L.; Abriouel, H.; Valdivia, E.; Omar, N. B. Application of bacteriocins in the control of foodborne pathogenic and spoilage bacteria. *Critical Reviews in Biotechnology* (2008) 28: 125-152
- Gee, H. Evolution: Ending inkongruence. *Nature* (2003) 425: 782
- Giovannoni, S. J.; Thrash, J. C.; Temperton, B. Implications of streamlining theory for microbial ekology. *The International Society for Microbial Ecology Journal* (2014) 8: 1553-1565
- Goto, Y.; Li, B.; Claesen, J.; Shi, Y.; Bibb, M. J.; van der Donk, W. A. Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLOS Biology* (2010) 8: e1000339

- Griese, M.; Lange, C.; Soppa, J. Ploidy in cyanobacteria. *FEMS Microbiology Letters* (2011) 323: 124-131
- Grissa, I., Vergnaud, G., Pourcel, Ch. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* (2007) 35: 52-57
- Gupta, R. S.; Mathews, D. J. Signature proteins for the major clades of Cyanobacteria. *BMC Evolutionary Biology* (2010) 10
- Gupta, R. S.; Pereira, M.; Chandrasekera, C.; Johari, V. Molecular signatures in protein sequences that are characteristic of cyanobacteria and plastid homologues. *International journal of systematic and evolutionary mikrobiology* (2003) 53: 1833-1842
- Hašler, P.; Casamatta, D.; Dvořák, P.; Poulíčková, A. *Jacksonvillea apiculata* (Oscillatoriales, Cyanobacteria) *gen. & sp. nov.*: a new genus of filamentous, epipsamic cyanobacteria from North Florida. *Phycologia* (2017) 56: 284-295
- Hess, W. R. Cyanobacterial genomics for ekology and biotechnology. *Current Opinion in Microbiology* (2011) 14: 608-614
- Honda, D.; Yokota, A.; Sugiyama, J. detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *Journal of Molecular Evolution* (1999) 48: 723-739
- Huson, D. H., Bryant, D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* (2006) 23: 254-267
- Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. Phylogenomics: the beginning of incongruence?. *TRENDS in Genetics* (2006) 22: 225-231.
- Jiang, X. M.; Neal, B.; Santiago, F.; Lee, S. J.; Romana, L. K.; Reeves, P. R. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Molecular Microbiology* (1991) 5: 695-713
- Kabra, R.; Kapil, A.; Attarwala, K.; Rai, P. K.; Shanker, A. Identification of common, unique and polymorphic microsatellites among 73 cyanobacterial genomes. *World Journal of Microbiology and Biotechnology* (2016) 32: 1-7

- Kumar, A.; Pearce, S. R.; McLean, K.; Harrison, G.; Heslop-Harrison, J. S.; Waugh, R.; Flavell, A. J. The Ty1-*copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. *Genetica* (1997) 100: 205-217
- Kumar, K.; Mella-Herrera, R. A.; Golden, J. W. Cyanobacterial Heterocysts. *Cold Spring Harbor Perspectives in Biology* (2010) 2: a000315
- Lang, B. F., Seif, E., Gray, M. W., O'Kelly, Ch. J., Burger, G. A comparative genomics approach to the evolution of eukaryotes and their mitochondria. *The Journal of Eukaryotic Microbiology* (1999) 46: 320-326
- Lartillot, N., Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* (2004) 21: 1095-1109
- Le, S. Q., Gascuel, O. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* (2008) 25: 1307-1320
- Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* (2010) 107: 10430-10435
- Lin, S.; Haas, S.; Zemojtel, T.; Xiao, P.; Vingron, M.; Li, R. Genome-wide comparison of cyanobacterial transposable elements, potential genetic diversity indicators. *Gene* (2011) 473: 139-149
- Lowe, T. M., Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 1997; 25: 955-964
- Luo, H., Sun, Z., Arndt, W., Shi, J., Friedman, R., Tang, J. Gene order phylogeny and the evolution of methanogens. *PLOS-One* (2009) 4: 1-5
- Lynch, M. Streamlining and simplification of microbial genome architecture. *The Annual Review of Microbiology* (2006) 60: 327-349
- Lyons, T. W.; Reinhard, C. T.; Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* (2014) 506: 307-315

- Mazel, D.; Houmard, J.; Castets, A. M.; Tandeau de Marsac, N. Highly repetitive DNA sequences in cyanobacterial genomes. *Journal of Bacteriology* (1990) 172: 2755-2761
- Miller, M. A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)* (2010) 1-8
- Moore, L. R.; Goericke, R.; Chisholm, S. W. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature of growth, pigments, fluorescence and absorptive properties. *Marine Ecology Progress Series* (1995) 116: 259-275
- Moore, L. R.; Post, A. F.; Rocap, G.; Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria. *Limnology and Oceanography* (2002) 47: 989-996
- Morel, A.; Ahn, Y.-H.; Partensky, F.; Vaultot, D.; Claustre, H. *Prochlorococcus* and *Synechococcus*: a comparative study of their optical properties in relation to their size and pigmentation. *Journal of Marine Research* (1993) 51: 617-649
- Moret, B. M. E., Wang, L.-S., Warnow, T., Wyman, S. K. New approaches for reconstructing phylogenies from gene order data. *Bioinformatics* (2001) 17: 165-173
- Morris, J. J.; Lenski, R. E.; Zinser, E. R. The Black Queen Hypothesis: Evolution of dependencies through adaptive gene loss. *mBio* (2012) 3: e00036-12
- Mulkidjanian, A. Y.; Koonin, E. V.; Makarova, K. S.; Mekhedov, S. L.; Sorokin, A.; Wolf, Y. I.; Dufresne, A.; Partensky, F.; Burd, H.; Kaznadzey, D.; Haselkorn, R.; Galperin, M. Y. The cyanobacterial genome core and the origin of photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* (2006) 103: 13126-13131
- Nakai, S.; Zou, G.; Song, X.; Pan, Q.; Zhou, S.; Hosomi, M. Release of anti-cyanobacterial allelochemicals from aquatic and terrestrial plants applicable for artificial floating islands. *Journal of Water and Environment Technology* (2008) 6: 55-63
- Neilands, J. B. Siderophores: Structure and function of microbial iron transport compounds. *The Journal of Biological Chemistry* (1995) 270: 26723-26726
- Nisbet, E. G.; Sleep, N. H. The habitat and nature of early life. *Nature* (2001) 409: 1083-1091

- Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N. Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics* (2005) 36: 541-562
- Philippe, H., Brinkmann, H., Lavrov, D. V., Timothy J Littlewood, D., Manuel, M., Wörheide, G., & Baurain, D. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLOS-Biology* (2011) 9: 402.
- Piper, C.; Cotter, P. D.; Ross, R. P.; Hill, C. Discovery of medically significant lantibiotics. *Current Drug Discovery Technologies* (2009) 6: 1-18
- Ran, L.; Larsson, J.; Vigil-Stenman, T.; Nylander, J. A. A.; Ininbergs, K.; Zheng, W.-W.; Lapidus, A.; Lowry, S.; Haselkorn, R.; Bergman, B. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLOS-One* (2010) 5
- Ren, F., Tanaka, H., Yang, Z. A likelihood look at the supermatrix-supertree controversy. *Gene* (2008) 441: 119-125
- Riley, M. A.; Wertz, J. E. Bacteriocins: evolution, ecology, and application. *Annual Review of Microbiology* (2002) 56: 117-137
- Rippka, R.; Deruelles, J.; Waterbury, J. B.; Herdman, M.; Stanier, R. Y. Generic assignments, strain histories and properties of pure cultures of Cyanobacteria. *Journal of General Microbiology* (1979) 111: 1-61
- Robbertse, B., Yoder, R. J., Boyd, A., Reeves, J., Spatafora, J. W. Hal: an automated pipeline for phylogenetic analyses of genomic data. *PLOS Currents* (2011) 3: RRN1213
- Robertson, B. R.; Tezuka, N.; Watanabe, M. Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. *International Journal of Systematic and Evolutionary Microbiology* (2001) 51: 861-871
- Sagan, L. On the origin of mitosing cells. *Journal of Theoretical Biology* (1967) 14: 255-274
- Sanchez-Baracaldo, P.; Hayes, P. K.; Blank, C. E. Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* (2005) 3: 145-165

- Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* (2002) 19: 101-109
- Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* (2003) 22: 301-302
- Satkoski, A. M.; Beukes, N. J.; Li, W.; Beard, B. L.; Johnson, C. M.; A redox-stratified ocean 3.2 billion years ago. *Earth and Planetary Science Letters* (2015) 15: 43-53
- Schirrneister, B. E.; de Vos, J. M.; Antonelli, A. Bagheri, H. C. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proceedings of the National Academy of Sciences of the United States of America* (2013) 110: 1791-1796
- Schirrneister, B. E.; Gugger, M.; Donoghue, P. C. J. Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* (2015) 5: 769-785
- Schopf, J. W. The fossil record: tracing the roots of the cyanobacterial lineage. V: Whitton, B. A.; Potts, M. The ecology of cyanobacteria: their diversity in time and space. *Springer, Berlin* (2002)
- Shigenobu, S.; Watanabe, H., Hattori, M.; Sakaki, Y.; Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* (2000) 407: 81-86
- Shih, P. M.; Wu, D.; Latifi, A.; Axen, S. D.; Fewer, D. P.; Talla, E.; Calteau, A.; Cai, F.; Tandeau de Marsac, N.; Rippka, R.; Herdman, M.; Sivonen, K.; Coursin, T. Laurent, T.; Goodwin, L.; Nolan, M.; Davenport, K. W.; Han, C. S.; Rubin, E. M.; Eisen, J. A.; Woyke, T.; Gugger, M.; Kerfeld, C. A. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* (2013) 110: 1053-1058
- Smit, A. F. A.; Hubley, R.; Green, P. *RepeatMasker Open-3.0*. dostupné z <<http://www.repeatmasker.org>> (1996-2010)
- Smith, A. J.; London, J.; Stanier, R. Y. Biochemical basis of obligate autotrophy in blue-green algae and thiobacilli. *Journal of Bacteriology* (1967) 94: 972-983

- Snel, B., Huynen, M. A., Dutilh, B. E. Genome Trees and the Nature of Genome Evolution. *The Annual Review of Microbiology* (2005) 59: 191-209
- Stamatakis, A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014
- Staub, R. Ernährungsphysiologisch-autökologische untersuchungen an der planktischen blualge *Oscillatoria rubescens* DC. *Schweizerische Zeitschrift für Hydrologie* (1961) 23: 83-198
- Steel, M., Rodrigo, A. Maximum Likelihood Supertrees. *Systematic Biology* (2008) 5: 243-250
- Sun, Z.; Blanchard, J. L. Strong genome-wide selection early in the evolution of *Prochlorococcus* resulted in a reduced genome through the loss of a large number of small effect genes. *PLOS-ONE* (2014) 9: e88837
- Swingley, W. D.; Blankenship, R. E.; Raymond, J. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Molecular Biology and Evolution* (2008) 25: 643-654
- Tamas, I.; Klasson, L.; Cänback, B.; Näslund, A. K.; Eriksson, A.-S.; Wernegreen, J. J.; Sandström, J. P.; Moran, N. A.; Andersson, S. G. E. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* (2002) 296: 2376-2379
- Tilmann, W.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Medema, T.; Medema, M. H. antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* (2015) 43: 237-243
- Tripp, H. J.; Bench, S. R.; Turk, K. A.; Foster, R. A.; Desany, B. A.; Niazi, F.; Affourtit, J. P.; Zehr, J. P. Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* (2010) 464: 90-94
- Ussery, D. W.; Hallin, P. F. Genome update: AT content in sequenced prokaryotic genomes. *Microbiology* (2004) 150: 749-752

- Von Mering, C.; Hugenholtz, P.; Raes, J.; Tringe, S. G.; Doerks, T.; Jensen, L. J.; Ward, N.; Bork, P. Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* (2007) 315: 1126-1130
- Waack, S.; Keller, O.; Asper, R.; Brodaq, T.; Damm, C.; Fricke, W. F.; Surovcik, K.; Meinicke, P.; Merkl, R. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* (2006) 16: 142
- Wagele, J. W., Mayer, C. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology* (2007) 7: 147
- Wang, L.-S., Warnow, T., Moret, B. M. E., Jansen, R. K., Raubeson, L. A. Distance-based genome rearrangement phylogeny. *Journal of Molecular Evolution* (2006) 63: 473-483
- Wang, H.; Fewer, D. P.; Sivonen, K. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLOS One* (2011) 6: e22384
- Wang, Z., Wu, M. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Scientific Reports* (2015) 5
- Weber, A.; Kögl, S. A.; Jung, K. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in *Escherichia coli*. *Journal of Bacteriology* (2006) 188: 7165-7175
- Wieland Brown, L. C.; Acker, M. G.; Clardy, J.; Walsh, C. T.; Fischbach, M. A. Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. *Proceedings of the National Academy of Sciences of the United States of America* (2009) 106: 2549-2553
- Woese, C. R., Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* (1977) 74: 5088-5090
- Woese, C. R., Otto Kandler, and Mark L. Wheelis. "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences* (1990) 87: 4576-4579.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L., Koonin, E. V. Genome trees constructed using five different approaches suggest new major bacterial clades. *BioMed Central Evolutionary Biology* (2001) 1

Xu, Q.; Yu, L.; Chitnis, V. P.; Chitnis, P. R. Function and organization of photosystem I in a cyanobacterial mutant strain that lacks PsaF and PsaJ subunits. *The Journal of Biological Chemistry* (1994) 269: 3205-3211

Young, K. D. Bacterial morphology: why have different shapes? *Current Opinion in Microbiology* (2007) 10: 596-600

Zhang, Z., Schwartz, S., Wagner, L., Miller, W. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* (2000) 7: 203-214

Zhou, F.; Tran, T.; Xu, Y. *Nezha*, a novel active miniature inverted-repeat transposable element in cyanobacteria. *Biochemical and Biophysical Research Communications* (2008) 365: 790-794

9 Seznam použitých zkratek a symbolů

(X) _n	opakující se sekvence daných nukleotidů
A, G, C, T	adenin, guanin, cytosin, thymin
A:T	adenin:Thymin
ABC	ATP binding cassette
ACP	acyl-carrier protein
ACT	Artemis Comparison Tool
<i>ada</i>	gen kódující 6-o-metylguanin-DNA metyltransferázu
Ala	alanin
<i>alkA</i>	gen kódující 3-metyladenin DNA glykosylázu
AMP, ADP, ATP	adenosinmono-, di-, trifosfát
AMPHORA	Automated Phylogenomic infeRence Application
antiSMASH	antibiotics and Secondary Metabolite Analysis Shell
ATCopia	transpozony nadrodiny Ty1/Copia z <i>Arabidopsis thaliana</i>
ATDNATA1	DNA transpozon <i>Arabidopsis thaliana</i> z nadrodiny DNA/Mariner
ATHATN1	DNA transpozon <i>Arabidopsis thaliana</i> z nadrodiny DNA/HAT
Athila4C	transpozon nadrodiny Gypsy/DIRS1
ATMU	DNA transpozon <i>Arabidopsis thaliana</i> z nadrodiny MULE-MuDR
ATPáza	adenosintrifosfatáza
ATREP	nadrodina neautonomních DNA transpozonů z <i>Arabidopsis thaliana</i>
ATSINE2	DNA transpozon <i>Arabidopsis thaliana</i> z nadrodiny SINE
BLAST	Basic Local Alignment Search Tool
BLASTN	Basic Local Alignment Search Tool for Nucleotides
BLASTP	Basic Local Alignment Search Tool for Proteins
Bp	páry bází
C11	molekula s jedenácti uhlíkatým řetězcem
cAMP	cyklický 3', 5'-adenosinmonofosfát
CDP, CTP	cytidindi-, trifosfát
CIPRES	CyberInfrastructure for Phylogenetic Research
Clp	rodina serin-proteáz
CoA	koenzym A
CRISPR	segment nahromaděných pravidelně rozmístěných krátkých palindromických repetit
CRP	c-reaktivní protein

<i>crp</i>	gen kódující c-reaktivní protein
CsbD	rodina proteinů spojená s odpovědí na stres
<i>cyaA</i>	gen kódující adenylát cyklázu
CyOGs	klastry ortologních skupin proteinů sinic
Cytochrom P450	skupina hem-thiolát monooxygenáz
dCMP, dCTP	deoxycytidinmono-, trifosfát
DevA	rodina ABC transportérů
DMT superfamily	Drug/metabolite transporter superfamily
DNA	deoxyribonukleová kyselina
<i>dnaA</i>	gen kódující chromozomální replikační iniciační protein DnaA
<i>dnaN</i>	gen kódující beta podjednotku DNA polymerázy III
<i>dnaX</i>	gen kódující tau podjednotku DNA polymerázy III
dTDP	deoxythymidindifosfát
EamA	transportér DMT nadrodiny
<i>exol/xseA</i>	gen kódující velkou podjednotku exonukleázy VII
FMN	Flavinmononukleotid, riboflavin -5'-fosfát
FtsE	protein pro tvorbu a stabilitu septa při buněčném dělení
FtsY	receptorový protein
G:C	Guanin:Cytosin
Gcn5	histon acetyltransferáza Gcn5
GLIMMER	Gene Locator and Interpolated Markov ModelER
GMP, GDP, GTP	guanosinmono-, di-, trifosfát
GntR	transkripční regulátor glukonát utilizačního systému GNT-I
GO	ontologie genů
GOE	Great Oxidation Event
guanin-N1-	vazba na dusík N1 guaninu
Gypsy/DIRS1	nadrodina retroelementů s dlouhými terminálními repeticemi
hAT-N4_Aly	DNA transpozon nadrodiny Hobo-activator
Helitrony	nadrodina autonomních DNA transpozonů
Hobo-activator	nadrodina DNA transpozonů
<i>holB</i>	gen kódující delta podjednotku DNA polymerázy III
HpnH	protein asociovaný se syntézou hopanoidů
IMP	inositolmonofosfát
IntIPac	integráza třídy I
IS1595	rodina transpozibilních elementů
ISAcba1	inzerční element z bakterie kmene Aktinobakterií
ISfinder	insertion-elements finder

ISShes11	inzerční element z <i>Shewanella</i> sp.
ISSod11	skupina transpozibilních elementů
Kb	kilobáze
LBA	přitažlivost dlouhých větví
<i>lexA</i>	gen kódující LexA represor transkripce
LSU-rRNA-Ath	rRNA velké podjednotky ribozomů <i>Arabidopsis thaliana</i>
LysR	rodina transkripčních regulátorů
Mb	megabáze
MCL	Markov Cluster
MerR	mercury resistance operon regulátory protein
MFP	membrane fusion protein
MFS	Major facilitator superfamily
MIRA	Mimicking Intelligent Read Assembly
MISA	Microsatellite Identification Tool
MITE	malé transpozibilní elementy s invertními repeticemi
ML	maximum-likelihood
mRNA, tRNA, rRNA	mediátorová, transferová, ribozomální ribonukleová kyselina
MRP	Matrix Representation using Parsimony
MULE-MuDR	nadrodina DNA transpozonů
MUSCLE	Multiple Sequence Comparison by Log-Expectation
<i>mutY</i>	gen kódující Adenin/Guanin-specifickou DNA glykolázu
N50	průměrná délka kontigů
NADH	redukovaný nikotinamidadenin dinukleotid
NADP+	oxidovaná forma nikotinamidadenin dinukleotidofostátu
NCBI	National Center for Biotechnology Information
Nif11	nitrogen-fixation domain
NRPS	syntázy neribozomálních peptidů
<i>pcb</i>	gen kódující Pcb protein světlosběrného komplexu
PE	Prolin-Glutamát motiv obsahující protein
PKS	polyketid syntázy
P-P	pyrofosfát
Protein-N(Pi)-fosfohistidin	protein-fosfohistidin
<i>PsaJ</i>	gen kódující protein IX podjednotky reakčního centra fotosystému I
<i>psbU</i>	gen kódující protein fotosystému II
<i>psbV</i>	gen kódující cytochrom <i>c</i> asociovaný s fotosystémem II
PUFA PKS	polyketid syntázy polynenasycených mastných kyselin
RAST	Rapid Annotation using Subsystem Technology

RAxMLHPC	Randomized Axelerated Maximum Likelihood for High Performance Computing
<i>recA</i>	gen kódující rekombinázu RecA
<i>recD</i>	gen kódující podjednotku exonukleázy V
<i>recJ</i>	gen kódující jednovláknovou DNA specifickou exonukleázu RecJ
<i>recQ</i>	gen kódující DNA helikázu RecQ
RibD	riboflavin syntetizující protein
RMbase	databáze repetitivních sekvencí pro RepeatMasker
RMBlast	RepeatMasker BLAST
RNA	ribonukleová kyselina
RND	resistance-nodulation-division
SAM	Sterile Alpha Motif
SRP	signal recognition particle
<i>ssb</i>	gen kódující jednovláknový DNA vazebný protein
SSU-rRNA-Ath	rRNA malé podjednotky ribozomů <i>Arabidopsis thaliana</i>
<i>sucA</i>	gen kódující E1 složku 2-oxoglutarát dehydrogenázového komplexu
<i>sucB</i>	gen kódující E2 složku 2-oxoglutarát dehydrogenázového komplexu
T3pk	polyketid syntáza typu III
Tn3	rodina transpozibilních elementů
Ty1/Copia	nadrodina retroelementů s dlouhými terminálními repeticemi
UDP, UTP	uridindi-, trifosfát
<i>umuCD</i>	gen kódující protein zapojený do ochrany proti ultrafialovému záření a mutacemi
<i>uvrABC</i>	gen kódující endonukleázu UvrABC
Vandal	DNA transpozon <i>Arabidopsis thaliana</i> z nadrodiny MULE-MuDR
VanW	Vankomycin B-typ obranný protein
X-Pro	aminokyselina vázaná s Prolinem
<i>xseB</i>	gen kódující malou podjednotku exonukleázy VII
ycf64	necharakterizovaný protein podobný monothiol glutaredoxinu

10 Přílohy

Tab. 1: Sinice použité ke konstrukci fylogenomických stromů.

Sinice	Přístupový kód genomu	Sinice	Přístupový kód genomu
<i>Acaryochloris</i> sp. CCMEE 5410	GCA_000238775.2	<i>Nodularia spumigena</i> CCY9414	CP007203
<i>Acaryochloris marina</i> MBIC11017	NC_009925	<i>Nostoc azollae</i>	CP002059
<i>Anabaena cylindrica</i> PCC 7122	GCA_000317695.1	<i>Nostoc</i> sp. PCC 7107	GCA_000316625.1
<i>Anabaena</i> sp. 90	NC_019427	<i>Nostoc</i> sp. PCC 7120	BA000019
<i>Anabaena variabilis</i> ATCC 29413	CP000117	<i>Nostoc</i> sp. PCC 7524	GCA_000316645.1
<i>Arthrospira maxima</i> CS-328	GCA_000173555.1	<i>Nostoc punctiforme</i> PCC 73102	CP001037
<i>Arthrospira</i> sp. PCC 8005	GCA_000973065.1	<i>Oscillatoria acuminata</i> PCC 6304	GCA_000317105.1
<i>Arthrospira platensis</i> C1	GCA_000307915.1	<i>Oscillatoria nigro-viridis</i> PCC 7112	GCA_000317475.1
<i>Arthrospira platensis</i> NIES-39	NC_016640	<i>Oscillatoria</i> sp. PCC 6506	GCA_000180455
<i>Arthrospira platensis</i> <i>Paraca</i>	GCA_000175415.3	Oscillatoriales cyanobacterium JSC-12	NZ_CM001633.1
<i>Calothrix</i> sp. PCC 6303	GCA_000317435.1	<i>Pleurocapsa</i> sp. PCC 7327	CP003590
<i>Calothrix</i> sp. PCC 7507	GCA_000316575.1	<i>Prochlorococcus</i> <i>marinus</i> AS960 ^S	GCA_000015645.1
<i>Chamaesiphon minutus</i> PCC 6605	GCA_000317145.1	<i>Prochlorococcus</i> <i>marinus</i> CCMP1375 ^S	GCA_000007925.1
<i>Chroococciopsis</i> <i>thermalis</i> PCC 4203	CP003597	<i>Prochlorococcus</i> <i>marinus</i> MIT 9202 ^S	GCA_000158595.1
<i>Crinalium epipsamum</i> PCC 9333	CP003620	<i>Prochlorococcus</i> <i>marinus</i> MIT9211 ^S	CP000878
<i>Crocospaera watsonii</i> WH0003	GCA_000235665.2	<i>Prochlorococcus</i> <i>marinus</i> MIT9215 ^S	CP000825
<i>Crocospaera watsonii</i> WH8501	GCA_000167195	<i>Prochlorococcus</i> <i>marinus</i> MIT9301 ^S	GCA_000015965.1

Tab. 1: Sinice použité ke konstrukci fylogenomických stromů.

<i>Cyanobacterium</i> sp. PCC 10605	CP003947	<i>Prochlorococcus</i> <i>marinus</i> MIT9303 ^S	GCA_000015705.1
<i>Cyanobacterium staineri</i> PCC 7202	CP003940	<i>Prochlorococcus</i> <i>marinus</i> MIT9312 ^S	GCA_000012645.1
<i>Cyanobacterium</i> UCYN-A	CP001842	<i>Prochlorococcus</i> <i>marinus</i> MIT9313 ^S	GCA_000011485.1
<i>Cyanobium gracile</i> PCC 6307 ^S	GCA_000316515	<i>Prochlorococcus</i> <i>marinus</i> MIT9515 ^S	GCA_000015665.1
<i>Cyanobium</i> sp. PCC 7001 ^S	GCA_000155635.1	<i>Prochlorococcus</i> <i>marinus</i> NATL1A ^S	GCA_000015685.1
<i>Cyanothece</i> sp. ATCC51472	GCA_000231425.3	<i>Prochlorococcus</i> <i>marinus</i> NATL2A ^S	GCA_000012465.1
<i>Cyanothece</i> sp. CCY0110	GCA_000169335.1	<i>Prochlorococcus</i> <i>marinus pastoris</i> CCMP1986 ^S	GCA_000011465.1
<i>Cyanothece</i> sp. PCC 7425	CP001344	<i>Pseudanabaena</i> sp. PCC 7429	GCA_000332215.1
<i>Cyanothece</i> sp. PCC 7822	CP002198	<i>Pseudanabaena</i> sp. PCC 7367	CP003592
<i>Cyanothece</i> sp. PCC 8801	CP001287	<i>Raphidiopsis brookii</i> D9	GCA_000175855.1
<i>Cyanothece</i> sp. PCC 8802	CP001701	<i>Rivularia</i> sp. PCC 7116	GCA_000316665.1
<i>Cyanothece</i> sp. ATCC51142	CP000806	<i>Spirulina subsalsa</i> PCC 9445	GCA_000314005.1
<i>Cyanothece</i> sp. PCC 7424	CP001291	<i>Stanieria cyanosphaera</i> PCC 7437	GCA_000317575.1
<i>Cylindrospermopsis</i> <i>raciborskii</i> CS-505	GCA_000175835	<i>Synechococcus</i> sp. BL107 ^S	GCA_000153805.1
<i>Cylindrospermum</i> <i>stagnale</i> PCC 7417	GCA_000317535.1	<i>Synechococcus</i> sp. CB0101 ^S	GCA_000179235.1
<i>Fischerella</i> sp. JSC-11	GCA_000231365.2	<i>Synechococcus</i> sp. CB0205 ^S	GCA_000179255.1
<i>Fischerella thermalis</i> PCC 7521	GCA_000317225.1	<i>Synechococcus</i> sp. CC9311 ^S	CP000435
<i>Geitlerinema</i> sp. PCC 7105	GCA_000332355.1	<i>Synechococcus</i> sp. CC9605 ^S	CP000110

Tab. 1: Sinice použité ke konstrukci fylogenomických stromů.

<i>Geitlerinema</i> sp. PCC 7407	CP003591	<i>Synechococcus</i> sp. CC9902 ^S	CP000097
<i>Gloeobacter kilauensis</i> JS1	GCA_000484535.1	<i>Synechococcus elongatus</i> PCC 6301 ^S	GCA_000010065.1
<i>Gloeobacter viloaceus</i> PCC 7421	BA000045	<i>Synechococcus elongatus</i> PCC 7942 ^S	CP000100
<i>Gloeocapsa</i> sp. PCC 73106	GCA_000332035	<i>Synechococcus</i> sp. JA-2- 3B'a(2-13)	CP000240
<i>Gloeocapsa</i> sp. PCC 7428	CP003646	<i>Synechococcus</i> sp. JA-3- 3Ab	CP000239
<i>Halothece</i> sp. PCC 7418	CP003945	<i>Synechococcus</i> sp. PCC 7335	GCA_000155595
<i>Leptolyngbia</i> sp. JSC-1	GCA_000733415.1	<i>Synechococcus</i> sp. PCC 7336	GCA_000332275.1
<i>Leptolyngbya</i> sp. PCC 6406	GCA_000332095.2	<i>Synechococcus</i> sp. PCC 6312	CP003558
<i>Leptolyngbya</i> sp. PCC 7375	GCA_000316115	<i>Synechococcus</i> sp. PCC 7002	CP000951
<i>Leptolyngbya</i> sp. PCC 7376	CP003946	<i>Synechococcus</i> sp. PCC 7502	CP003594
<i>Moorea producens</i> 3L	GCA_000211815.1	<i>Synechococcus</i> sp. RCC307 ^S	CT978603
<i>Lyngbya</i> sp. PCC 8106	GCA_000169095	<i>Synechococcus</i> sp. RS9916 ^S	GCA_000153825
<i>Coleofasciculus chthonoplastes</i> PCC 7420	GCA_000155555	<i>Synechococcus</i> sp. RS9917 ^S	BX548020
<i>Microcoleus</i> sp. PCC 7113	CP003630	<i>Synechococcus</i> sp. NKBG041c	CP006882
<i>Microcoleus vaginatus</i> FGP-2	GCA_000214075.2	<i>Candidatus Synechococcus spongiarum</i> ^S	GCA_000586015.1
<i>Microcystis aeruginosa</i> DIANCHI905	GCA_000332585.1	<i>Synechococcus</i> sp. WH 5701 ^S	GCA_000153045
<i>Microcystis aeruginosa</i> NIES-843	GCA_000010625	<i>Synechococcus</i> sp. WH 7803 ^S	NC_009481
<i>Microcystis aeruginosa</i> PCC 7941	GCA_000312205.1	<i>Synechococcus</i> sp. WH 7805 ^S	CT971583

Tab. 1: Sinice použité ke konstrukci fylogenomických stromů.

<i>Microcystis aeruginosa</i> PCC 9432	GCA_000307995.2	<i>Synechococcus</i> sp. WH 8016 ^S	GCA_000230675
<i>Microcystis aeruginosa</i> PCC 9443	GCA_000312185.1	<i>Synechococcus</i> sp. WH 8102 ^S	GCA_000195975.1
<i>Microcystis aeruginosa</i> PCC 9701	GCA_000312285.1	<i>Synechococcus</i> sp. WH 8109 ^S	GCA_000161795.2
<i>Microcystis aeruginosa</i> PCC 9717	GCA_000312165.1	<i>Synechocystis</i> sp. PCC 6714	GCA_000478825.2
<i>Microcystis aeruginosa</i> PCC 9806	GCA_000312725.1	<i>Synechocystis</i> sp. PCC 6803	BA000022
<i>Microcystis aeruginosa</i> PCC 9807	GCA_000312225.1	<i>Synechocystis</i> sp. PCC 7509	GCA_000332075
<i>Microcystis aeruginosa</i> PCC 9808	GCA_000312245.1	<i>Thermosynechococcus</i> <i>elongatus</i> BF-1	BA000039
<i>Microcystis aeruginosa</i> PCC 9809	GCA_000312265.1	<i>Thermosynechococcus</i> sp. NK55	GCA_000505665.1
<i>Microcystis aeruginosa</i> TAIHU98	GCA_000330925.1	<i>Trichodesmium</i> <i>erythraeum</i> IMS101	CP000393
<i>Microcystis</i> sp. T1 4	GCA_000297435.1	<i>Xenococcus</i> sp. PCC 7305	GCA_000332055
<i>Neosynechococcus</i> <i>sphagnicola</i> CAUP A 1101	JJML00000000	<i>Elainellia</i> sp. E1	LUGL00000000
<i>Nodosilinea nodulosa</i> PCC 7104	GCA_000309385.1		

^S – sinice použité ke konstrukci fylogenomického stromu pro *Synechococcus* sp. S4

Tab. 4: Potvrzené CRISPRy nalezené v genomu *J. apiculata* programem CRISPRfinder.

Opakovaný motiv	Délka přímé repetice (nt)	Délka CRISPRu (nt)	Počet spacerů mezi motivy
CTTCAACTAATAGAATCCCGGCAACGGGACTGAAAC	37	7767	105
GTTTCAGTCCCGTTGCCGGGATTCTATTAGTTGAAAG	37	5572	75
GTTTCCAATAATCCGAATTAACCCAATCGGTGGGG	36	3287	44
CCCTACCGGCTGGGTCAAACCGGACTAATTGGAAAC	36	2014	25
GTCGCAAAACCCAAATCCCGGTCAACGGGACTGAAAC	37	1350	18
GTTTCAGTCCCCTTGCGGGGATTTGGTTATGAAAG	35	1288	17
GTCGCAAAACCCAAATCCCGGTCAACGGGACTGAAAC	37	910	12
TCTTTCATAACCAAATCCCGCAAGGGGACTGAAAC	36	856	11
CTTTCATAACCAAATCCCGCAAGGGGACTGAAAC	35	831	11
GTTTCCAATTAGTCCGGTTTGACCCAGCCGGTAGGG	36	743	9
GTTTCAGTCCCCTTGCGGGGATTTGGTTATGAAAG	35	565	7
GTTTCATCAACCAAATCCCGCAAGGGGACTGAAA	34	543	7
GTCGCAAAACCCAAATCCCGGTCAACGGGACTGAAAC	37	476	6
TCAATCCCTGAAAGGGTTTTATGTGAGTTTCGAT	34	401	5
TTAAATTAATAGCCCCACCATTTTCCT	26	256	3
TTAAATTAATAGCCCCACCATTTTCCT	26	251	3

Tab. 5: Sporné CRISPRy nalezené v genomu *J. apiculata* programem CRISPRfinder

Opakovaný motiv	Délka přímé repetice (nt)	Délka CRISPRu (nt)	Počet spacerů mezi motivy
TGTTTAGCGTTGTTCAAGTGTAAATCAAAGTCGATTT AGCGTTTTAGAGCGTT	53	234	2
TTTCAATCCCTAACAGGGTGCAGGTGGTAAGTCTTGC GGACTTGGACTGGGAGATGGACTGGGAGA	36	177	2
TAGCTTCCCTCTGGGTATCTTTATCATCAGAAATTGTC TAAAAGGGAGAT	29	172	3
CGCCGTGACGAATGGCGCGATGACCGCCGTGACGAC CGGAG	49	150	1
TCAACAGTCGCCAAGTCTAAATCATCCCAAGCCTCA AAAGCGATCGCCCTGAGTTGAAACAGAGAAAAGGC GATC	41	124	1
CGCGGCTCGACTGAGCGATCGCCGAACGT	36	119	1
GTAACAGGACAGCTTCTCCCAAAGCCAGATGATATT G	39	118	1
ATCAAGGACAGCTTGATTGTAGCGAT AATTTGCTCAAAGTGGTGAGGAAAA	29	109	1
ATCAAGGACAGCTTGATTGTAGCGAT AATTTGCTCAAAGTGGTGAGGAAAA	37	108	1
ATCAAGGACAGCTTGATTGTAGCGAT AATTTGCTCAAAGTGGTGAGGAAAA	26	106	1
ATCAAGGACAGCTTGATTGTAGCGAT AATTTGCTCAAAGTGGTGAGGAAAA	25	105	1
CCCTTAGCATGGGCTGCCTTTCC GATCGCAAGAGTTCCCACCAATC	23	97	1
CCCTTAGCATGGGCTGCCTTTCC GATCGCAAGAGTTCCCACCAATC	23	96	1
GGCGATCGCAGGTGTTCTCGGC TCTAACTGCTGGGGCAATATTA	23	86	1
GGCGATCGCAGGTGTTCTCGGC TCTAACTGCTGGGGCAATATTA	23	172	2
TTTTTCTATTTGTAACAATATATGCGATTAATGGATA G	38	134	1
AAAAGCGTTCACGAAGTGTCTCTG TACCACAGGAAATTGAACCAATG	24	74	1
AAAAGCGTTCACGAAGTGTCTCTG TACCACAGGAAATTGAACCAATG	23	85	1

Tab. 8: Skupiny genů druhé kategorie anotovaných GO analýzou *J. apiculata*

Identifikační číslo	Počet anotovaných genů	p-value	Skupina
GO:0006412	110	3.50339726600817e-88	Translace
GO:0006810	77	1.37942154647719e-61	Transport
GO:0006118	36	1.50900796909205e-28	Transport elektronů
GO:0006457	17	3.19779182071762e-13	Skládání proteinů
GO:0006281	14	8.4037686950889e-11	Reparace DNA
GO:0006260	12	3.43604126085766e-09	Replikace DNA
GO:0008152	11	2.19258274627617e-08	Metabolické procesy
GO:0006304	10	1.39605657316643e-07	Modifikace DNA
GO:0007155	10	1.39605657316643e-07	Adheze buňky
GO:0006413	9	8.86087191283031e-07	Iniciace translace
GO:0051276	9	8.86087191283031e-07	Organizace chromozómu
GO:0009058	7	3.51228691353618e-05	Biosyntetické procesy
GO:0009236	7	3.51228691353618e-05	Biosyntéza kobalaminu

Tab. 8: Skupiny genů druhé kategorie anotovaných GO analýzou *J. apiculata*

GO:0006298	6	0.000218029302038746	Oprava chybného párování
GO:0006310	6	0.000218029302038746	Rekombinace DNA
GO:0006464	6	0.000218029302038746	Modifikace buněčných proteinů
GO:0006508	6	0.000218029302038746	Proteolýza
GO:0006813	6	0.000218029302038746	Transport draselných iontů
GO:0045449	6	0.000218029302038746	DNA závislá regulace transkripce
GO:0015031	5	0.00133058970720838	Transport proteinů
GO:0019538	5	0.00133058970720838	Metabolismus proteinů
GO:0006289	4	0.00789080024386337	Nukleotidová excizní oprava
GO:0006364	4	0.00789080024386337	Zpracování rRNA
GO:0006415	4	0.00789080024386337	Terminace translace
GO:0006461	4	0.00789080024386337	Skládání proteinových komplexů
GO:0006479	4	0.00789080024386337	Metylace proteinů
GO:0006605	4	0.00789080024386337	Cílení proteinů
GO:0006614	4	0.00789080024386337	SRP závislé kotranslační cílení proteinů do membrány
GO:0006807	4	0.00789080024386337	Metabolismus dusíkatých sloučenin
GO:0006814	4	0.00789080024386337	Transport sodných iontů
GO:0006821	4	0.00789080024386337	Transport chloridů
GO:0009306	4	0.00789080024386337	Sekrece proteinů
GO:0044267	4	0.00789080024386337	Metabolismus buněčných proteinů
GO:0005975	3	0.0444286286975741	Metabolismus uhlovodíků
GO:0006352	3	0.0444286286975741	DNA závislá iniciace transkripce
GO:0007205	3	0.0444286286975741	Signální dráha protein kinázy C aktivující receptor spřažený s G-proteinem
GO:0009228	3	0.0444286286975741	Biosyntéza thiaminu
GO:0017004	3	0.0444286286975741	Složení cytochromového komplexu
GO:0000105	2	0.225443233018957	Biosyntéza histidinu
GO:0000271	2	0.225443233018957	Biosyntéza polysacharidů
GO:0006259	2	0.225443233018957	Metabolismus DNA
GO:0006284	2	0.225443233018957	Bázová excizní oprava
GO:0006396	2	0.225443233018957	Zpracování RNA
GO:0006414	2	0.225443233018957	Translační elongace
GO:0006546	2	0.225443233018957	Katabolismus glycinu
GO:0006629	2	0.225443233018957	Metabolismus lipidů
GO:0006790	2	0.225443233018957	Metabolismus sloučenin síry
GO:0006808	2	0.225443233018957	Regulace využití dusíku
GO:0006811	2	0.225443233018957	Transport iontů
GO:0006812	2	0.225443233018957	Transport kationtů
GO:0006817	2	0.225443233018957	Transport fosfátových iontů
GO:0006886	2	0.225443233018957	Transport vnitrobuněčných proteinů
GO:0006935	2	0.225443233018957	Chemotaxe
GO:0007046	2	0.225443233018957	Biogeneze ribozómů
GO:0007165	2	0.225443233018957	Transdukce signálu

Tab. 8: Skupiny genů druhé kategorie anotovaných GO analýzou *J. apiculata*

GO:0008616	2	0.225443233018957	Biosyntéza queuosinu
GO:0015684	2	0.225443233018957	Transport železitých iontů
GO:0015727	2	0.225443233018957	Transport laktátu
GO:0015746	2	0.225443233018957	Transport citrátu
GO:0016036	2	0.225443233018957	Buněčná odpověď na nedostatek fosfátů
GO:0017038	2	0.225443233018957	Import proteinů
GO:0019627	2	0.225443233018957	Metabolismus močoviny
GO:0042254	2	0.225443233018957	Biogeneze ribozómů
GO:0042773	2	0.225443233018957	Transport elektronů spojený se syntézou ATP
GO:0006633	1	0.899218808659587	Biosyntéza mastných kyselin
GO:0006950	1	0.899218808659587	Odpověď na stres
GO:0009073	1	0.899218808659587	Biosyntéza aromatických aminokyselin
GO:0015711	1	0.899218808659587	Transport organických aniontů
GO:0015986	1	0.899218808659587	Transport protonů spojený se syntézou ATP
GO:0016310	1	0.899218808659587	Fosforylace
GO:0030001	1	0.899218808659587	Transport kovových iontů

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

Identifikační číslo	Počet anotovaných genů	p-value	Skupina
GO:0005524	115	3.26155858806372e-92	Vazba ATP
GO:0003735	94	2.76037666250223e-75	Strukturní složka ribozómu
GO:0004016	79	3.3746051654564e-63	Adenylát cyklázová aktivita
GO:0005554	52	1.91785081885529e-41	Molekulární funkce
GO:0016887	46	1.31216312297369e-36	ATPázová aktivita
GO:0005215	42	2.19731987867729e-33	Transportérová aktivita
GO:0004035	38	3.6836807944015e-30	Alkalicko fosfatázová aktivita
GO:0005506	37	2.35758517476216e-29	Vazba iontů železa
GO:0003677	36	1.50900796909205e-28	Vazba s DNA
GO:0003676	34	6.18390452486006e-27	Vazba nukleové kyseliny
GO:0005525	32	2.53517856391373e-25	Vazba s GTP
GO:0009055	32	2.53517856391373e-25	Přenos elektronů
GO:0003824	31	1.62349307803033e-24	Katalytická aktivita
GO:0003723	29	6.66005199130635e-23	Vazba s RNA
GO:0003700	28	4.26642535713405e-22	Sekvenčně specifický DNA vazebný transkripční faktor
GO:0008080	22	2.95491757300003e-17	N-acetyltransferázová aktivita
GO:0000156	21	1.89416259779545e-16	Dvousložkový regulátor odpovědi
GO:0003887	17	3.19779182071762e-13	DNA řízená DNA polymerázová aktivita
GO:0016491	17	3.19779182071762e-13	Oxidoreduktázová aktivita
GO:0004386	14	8.4037686950889e-11	Helikázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0005515	13	5.37647688423942e-10	Vazba proteinů
GO:0004322	12	3.43604126085766e-09	Ferroxidázová aktivita
GO:0046933	12	3.43604126085766e-09	Rotační mechanismus ATP syntázy, přenos vodíkových iontů
GO:0046961	12	3.43604126085766e-09	Rotační mechanismus ATPázy, přenos protonů
GO:0004316	11	2.19258274627617e-08	3-oxoacyl-[acyl-nesoucí-protein] reduktázová aktivita (NADPH)
GO:0008270	11	2.19258274627617e-08	Vazba iontů zinku
GO:0051082	11	2.19258274627617e-08	Vazba nesložených proteinů
GO:0003899	10	1.39605657316643e-07	DNA řízená RNA polymerázová aktivita
GO:0004721	10	1.39605657316643e-07	Fosfoprotein fosfatázová aktivita
GO:0008026	10	1.39605657316643e-07	ATP-dependentní helikázová aktivita
GO:0008565	10	1.39605657316643e-07	Transport proteinů
GO:0003743	9	8.86087191283031e-07	Translační iniciační faktor
GO:0004056	9	8.86087191283031e-07	Argininosukcinát lyázová aktivita
GO:0003896	8	5.59806054901103e-06	DNA primázová aktivita
GO:0003978	8	5.59806054901103e-06	UDP-glukóza 4-epimerázová aktivita
GO:0004069	8	5.59806054901103e-06	L-aspartát:2-oxoglutarát aminotransferázová aktivita
GO:0004129	8	5.59806054901103e-06	Cytochrom C oxidázová aktivita
GO:0008324	8	5.59806054901103e-06	Transmembránový transport kationtů
GO:0008462	8	5.59806054901103e-06	Clp endopeptidázová aktivita
GO:0016151	8	5.59806054901103e-06	Vazba kationtů niklu
GO:0016851	8	5.59806054901103e-06	Chelatace hořčíku
GO:0047631	8	5.59806054901103e-06	ADP-ribóza difosfatázová aktivita
GO:0004619	7	3.51228691353618e-05	Fosfoglycerát mutázová aktivita
GO:0008460	7	3.51228691353618e-05	dTDP-glukóza 4,6-dehydratázová aktivita
GO:0008955	7	3.51228691353618e-05	Peptidoglykan glykosyltransferázová aktivita
GO:0008983	7	3.51228691353618e-05	Protein-glutamát O-metyltransferázová aktivita
GO:0031071	7	3.51228691353618e-05	Cystein desulfurázová aktivita
GO:0031072	7	3.51228691353618e-05	Vazba heat shock proteinů
GO:0003697	6	0.000218029302038746	Vazba jednovláknové DNA
GO:0003848	6	0.000218029302038746	2-amino-4-hydroxy-6-hydroxymetyldihydropteridin difosfokinázová aktivita
GO:0003918	6	0.000218029302038746	DNA topoizomerázová (ATP-hydrolyzující) aktivita
GO:0003984	6	0.000218029302038746	Acetolaktát syntázová aktivita
GO:0004124	6	0.000218029302038746	Cystein syntázová aktivita
GO:0004364	6	0.000218029302038746	Glutathion transferázová aktivita
GO:0004642	6	0.000218029302038746	Fosforibosylformylglycinamidin syntázová aktivita
GO:0008556	6	0.000218029302038746	Draslík transportující ATPázová aktivita
GO:0008986	6	0.000218029302038746	Pyruvát, voda dikinázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0009002	6	0.000218029302038746	Serin-tyt D-Ala-D-Ala karboxypeptidázová aktivita
GO:0009039	6	0.000218029302038746	Ureázová aktivita
GO:0030528	6	0.000218029302038746	Regulátor transkripce
GO:0047360	6	0.000218029302038746	Undekaprenyl-fosfát galaktóza fosfotransferázová aktivita
GO:0048529	6	0.000218029302038746	Mg-protoporfyrin IX monometyl ester (oxidativní) cyklázová aktivita
GO:0004022	5	0.00133058970720838	Alkohol dehydrogenázová (NAD) aktivita
GO:0004089	5	0.00133058970720838	Karbonát dehydratázová aktivita
GO:0004143	5	0.00133058970720838	Diacylglycerol kinázová aktivita
GO:0004239	5	0.00133058970720838	Methionyl aminopeptidázová aktivita
GO:0004356	5	0.00133058970720838	Glutamát-amoniak ligázová aktivita
GO:0004739	5	0.00133058970720838	Pyruvát dehydrogenázová (acetyl-přenašející) aktivita
GO:0004851	5	0.00133058970720838	Uroporfyrin-III C-metyltransferázová aktivita
GO:0005488	5	0.00133058970720838	Vazba
GO:0008236	5	0.00133058970720838	Peptidáza serinového typu
GO:0008750	5	0.00133058970720838	NAD(P)+ transhydrogenázová (AB-specifická) aktivita
GO:0008830	5	0.00133058970720838	dTDP-4-dehydroramnóza 3,5-epimerázová aktivita
GO:0008831	5	0.00133058970720838	dTDP-4-dehydroramnóza reduktázová aktivita
GO:0008855	5	0.00133058970720838	Exodeoxyribonukleáza VII
GO:0008908	5	0.00133058970720838	Izochorismatázová aktivita
GO:0008966	5	0.00133058970720838	Fosfoglukosamin mutázová aktivita
GO:0015446	5	0.00133058970720838	Transmembránová arzen transportující ATPázová aktivita
GO:0016740	5	0.00133058970720838	Transferázová aktivita
GO:0016787	5	0.00133058970720838	Hydrolázová aktivita
GO:0047985	5	0.00133058970720838	Dehydrogenázová aktivita
GO:0000104	4	0.00789080024386337	Sukcinát dehydrogenázová aktivita
GO:0003678	4	0.00789080024386337	DNA helikázová aktivita
GO:0003684	4	0.00789080024386337	Vazba poškozené DNA
GO:0003747	4	0.00789080024386337	Translační uvolňující faktor
GO:0003841	4	0.00789080024386337	1-acylglycerol-3-fosfát O-acyltransferázová aktivita
GO:0003861	4	0.00789080024386337	3-isopropylmalát dehydratázová aktivita
GO:0003879	4	0.00789080024386337	ATP fosforibosyltransferázová aktivita
GO:0003938	4	0.00789080024386337	IMP dehydrogenázová aktivita
GO:0003939	4	0.00789080024386337	L-itol 2-dehydrogenázová aktivita
GO:0003951	4	0.00789080024386337	NAD+ kinázová aktivita
GO:0003954	4	0.00789080024386337	NADH dehydrogenázová aktivita
GO:0003989	4	0.00789080024386337	Acetyl-CoA karboxylázová aktivita
GO:0003991	4	0.00789080024386337	Acetylglutamát kinázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0003992	4	0.00789080024386337	N2-acetyl-L-ornitin:2-oxoglutarát 5-aminotransferázová aktivita
GO:0004008	4	0.00789080024386337	Cu-exportní ATPázová aktivita
GO:0004017	4	0.00789080024386337	Adenylát kinázová aktivita
GO:0004048	4	0.00789080024386337	Antranilát fosforibozyltransferázová aktivita
GO:0004049	4	0.00789080024386337	Antranilát syntázová aktivita
GO:0004088	4	0.00789080024386337	Karbamoyl-fosfát syntázová (glutamin-hydrolyzující) aktivita
GO:0004131	4	0.00789080024386337	Cytosin deaminázová aktivita
GO:0004141	4	0.00789080024386337	Dethiobiotin syntázová aktivita
GO:0004148	4	0.00789080024386337	Dihydrolipoyl dehydrogenázová aktivita
GO:0004151	4	0.00789080024386337	Dihydroorotázová aktivita
GO:0004325	4	0.00789080024386337	Ferrochelatazová aktivita
GO:0004332	4	0.00789080024386337	Fruktóza-bisfosfát aldolázová aktivita
GO:0004350	4	0.00789080024386337	Glutamát-5-semialdehyd dehydrogenázová aktivita
GO:0004375	4	0.00789080024386337	Glycin dehydrogenázová (dekarboxylační) aktivita
GO:0004392	4	0.00789080024386337	Hem oxygenázová (decyklizační) aktivita
GO:0004412	4	0.00789080024386337	Homoserin dehydrogenázová aktivita
GO:0004416	4	0.00789080024386337	Hydroxyacylglutathion hydrolázová aktivita
GO:0004478	4	0.00789080024386337	Methionin adenosyltransferázová aktivita
GO:0004518	4	0.00789080024386337	Nukleázová aktivita
GO:0004523	4	0.00789080024386337	Ribonukleáza H
GO:0004525	4	0.00789080024386337	Ribonukleáza III
GO:0004638	4	0.00789080024386337	Fosforibozylaminoimidazol karboxylázová aktivita
GO:0004733	4	0.00789080024386337	Pyridoxamin-fosfát oxidázová aktivita
GO:0004743	4	0.00789080024386337	Pyruvát kinázová aktivita
GO:0004775	4	0.00789080024386337	Sukcinát-CoA ligázová (ADP-formující) aktivita
GO:0004784	4	0.00789080024386337	Superoxid dismutázová aktivita
GO:0004795	4	0.00789080024386337	Threonin syntázová aktivita
GO:0004820	4	0.00789080024386337	Glycin-tRNA ligázová aktivita
GO:0004826	4	0.00789080024386337	Fenylalanin-tRNA ligázová aktivita
GO:0004834	4	0.00789080024386337	Tryptofan syntázová aktivita
GO:0004845	4	0.00789080024386337	Uracil fosforibozyltransferázová aktivita
GO:0005247	4	0.00789080024386337	Napětím řízený chloridový kanál
GO:0008276	4	0.00789080024386337	Protein metyltransferázová aktivita
GO:0008451	4	0.00789080024386337	X-Pro aminopeptidázová aktivita
GO:0008508	4	0.00789080024386337	Žlučová kyselina:sodík symportérová aktivita
GO:0008703	4	0.00789080024386337	5-amino-6-(5-fosforibozylamino)uracil reduktázová aktivita
GO:0008774	4	0.00789080024386337	Acetaldehyd dehydrogenázová (acetylační) aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0008781	4	0.00789080024386337	N-acylneuraminát cytidylyltransferázová aktivita
GO:0008817	4	0.00789080024386337	Kob(I)yrinová kyselina a,c-diamid adenozylyltransferázová aktivita
GO:0008915	4	0.00789080024386337	Lipid-A-disacharid syntázová aktivita
GO:0008984	4	0.00789080024386337	Protein-glutamát metylesterázová aktivita
GO:0009011	4	0.00789080024386337	Syntáza škrobu
GO:0009378	4	0.00789080024386337	Four-way junction helikázová aktivita
GO:0015450	4	0.00789080024386337	P-P-vazbou-hydrolyzou-řízený proteinový transmembránový transportér
GO:0016984	4	0.00789080024386337	Ribulóza-bisfosfát karboxylázová aktivita
GO:0016993	4	0.00789080024386337	Prekordin-8X metylmutázová aktivita
GO:0020037	4	0.00789080024386337	Vazba Hemu
GO:0030151	4	0.00789080024386337	Vazba iontů Molybdenu
GO:0042132	4	0.00789080024386337	Fruktóza 1,6-bisfosfát 1-fosfatázová aktivita
GO:0042586	4	0.00789080024386337	Peptid deformylázová aktivita
GO:0047244	4	0.00789080024386337	N-acetylglukosaminyldifosfoundekaprenol N-acetyl-beta-D-mannózaminyltransferázová aktivita
GO:0000774	3	0.0444286286975741	Adenyl-nukleotid výměnný faktor
GO:0003852	3	0.0444286286975741	2-isopropylmalát syntázová aktivita
GO:0003856	3	0.0444286286975741	3-dehydrochinát syntázová aktivita
GO:0003917	3	0.0444286286975741	DNA topoizoizomeráza typu I
GO:0003922	3	0.0444286286975741	GMP syntázová (glutamin-hydrolyzující) aktivita
GO:0003934	3	0.0444286286975741	GTP cyklohydroláza I
GO:0003935	3	0.0444286286975741	GTP cyklohydroláza II
GO:0003979	3	0.0444286286975741	UDP-glukóza 6-dehydrogenázová aktivita
GO:0004029	3	0.0444286286975741	Aldehyd dehydrogenázová (NAD) aktivita
GO:0004067	3	0.0444286286975741	Asparaginázová aktivita
GO:0004134	3	0.0444286286975741	4-alfa-glukanotransferázová aktivita
GO:0004340	3	0.0444286286975741	Glukokinázová aktivita
GO:0004345	3	0.0444286286975741	Glukóza-6-fosfát dehydrogenázová aktivita
GO:0004414	3	0.0444286286975741	Homoserin O-acetyltransferázová aktivita
GO:0004452	3	0.0444286286975741	Izopentenyl-difosfát delta-izomerázová aktivita
GO:0004462	3	0.0444286286975741	Laktoylglutathion lyázová aktivita
GO:0004515	3	0.0444286286975741	Nikotinát-nukleotid adenilyltransferázová aktivita
GO:0004558	3	0.0444286286975741	Alfa-glukosidázová aktivita
GO:0004640	3	0.0444286286975741	Fosforibozylantranilát izomerázová aktivita
GO:0004645	3	0.0444286286975741	Fosforylázová aktivita
GO:0004655	3	0.0444286286975741	Porfobilinogen syntázová aktivita
GO:0004730	3	0.0444286286975741	Pseudouridy lát syntázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0004735	3	0.0444286286975741	Pyrrolin-5-karboxylát reduktázová aktivita
GO:0004747	3	0.0444286286975741	Ribokinázová aktivita
GO:0004748	3	0.0444286286975741	Ribonukleosid-difosfát reduktázová aktivita, thioredoxin disulfide jako akceptor
GO:0004791	3	0.0444286286975741	Thioredoxin-disulfid reduktázová aktivita
GO:0004794	3	0.0444286286975741	L-threonin amoniak lyázová aktivita
GO:0004798	3	0.0444286286975741	Tymidilát kinázová aktivita
GO:0004801	3	0.0444286286975741	Sedoheptulóza-7-fosfát:D-glyceraldehyd-3-fosfát glyceron transferázová aktivita
GO:0004802	3	0.0444286286975741	Transketolázová aktivita
GO:0004825	3	0.0444286286975741	Methionin-tRNA ligázová aktivita
GO:0004852	3	0.0444286286975741	Uroporfyrinogen-III syntázová aktivita
GO:0004872	3	0.0444286286975741	Receptorová aktivita
GO:0008168	3	0.0444286286975741	Metyltransferázová aktivita
GO:0008685	3	0.0444286286975741	2-C-metyl-D-erytritol 2,4-cyklodifosfát syntázová aktivita
GO:0008783	3	0.0444286286975741	Agmatinázová aktivita
GO:0008800	3	0.0444286286975741	Beta-laktamázová aktivita
GO:0008801	3	0.0444286286975741	Beta-fosfoglukomutázová aktivita
GO:0008836	3	0.0444286286975741	Diaminopimelát dekarboxylázová aktivita
GO:0008909	3	0.0444286286975741	Izochorismát syntázová aktivita
GO:0008964	3	0.0444286286975741	Fosfoenolpyruvát karboxylázová aktivita
GO:0008976	3	0.0444286286975741	Polyfosfát kinázová aktivita
GO:0009001	3	0.0444286286975741	Serin O-acetyltransferázová aktivita
GO:0009010	3	0.0444286286975741	Sorbitol-6-fosfát 2-dehydrogenázová aktivita
GO:0016630	3	0.0444286286975741	Protochlorofylid reduktázová aktivita
GO:0016779	3	0.0444286286975741	Nukleotidyltransferázová aktivita
GO:0016987	3	0.0444286286975741	Sigma faktor
GO:0042803	3	0.0444286286975741	Protein homodimerizační aktivita
GO:0043365	3	0.0444286286975741	[formát-C-acetyltransferáza]-aktivační enzym
GO:0051087	3	0.0444286286975741	Vazba chaperonů
GO:0000286	2	0.225443233018957	Alanin dehydrogenázová aktivita
GO:0000310	2	0.225443233018957	Xantin fosforibozyltransferázová aktivita
GO:0001882	2	0.225443233018957	Vazba nukleosidů
GO:0003711	2	0.225443233018957	Regulátor transkripční elongace
GO:0003746	2	0.225443233018957	Translační elongační faktor
GO:0003840	2	0.225443233018957	Gama-glutamyltransferázová aktivita
GO:0003842	2	0.225443233018957	1-pyrrolin-5-karboxylát dehydrogenázová aktivita
GO:0003862	2	0.225443233018957	3-izopropylmalát dehydrogenázová aktivita
GO:0003864	2	0.225443233018957	3-metyl-2-oxobutanoát hydroxymetyltransferázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0003868	2	0.225443233018957	4-hydroxyfenylpyruvát dioxygenázová aktivita
GO:0003872	2	0.225443233018957	6-fosfofruktokinázová aktivita
GO:0003883	2	0.225443233018957	CTP syntázová aktivita
GO:0003905	2	0.225443233018957	Alkylbáze DNA N-glykosylázová aktivita
GO:0003906	2	0.225443233018957	DNA-(apurinové nebo apyrimidinové místo) lyázová aktivita
GO:0003908	2	0.225443233018957	Metylovaný-DNA-[protein]-cystein S-metyltransferázová aktivita
GO:0003911	2	0.225443233018957	DNA ligázová (NAD ⁺) aktivita
GO:0003913	2	0.225443233018957	DNA fotolyázová aktivita
GO:0003919	2	0.225443233018957	FMN adenylyltransferázová aktivita
GO:0003937	2	0.225443233018957	IMP cyklohydrolázová aktivita
GO:0003942	2	0.225443233018957	N-acetyl-gama-glutamyl-fosfát reduktázová aktivita
GO:0003949	2	0.225443233018957	1-(5-fosforibosyl)-5-[(5-Fosforibosylamino)metylidenamino]imidazol-4-karboxamid izomerázová aktivita
GO:0003977	2	0.225443233018957	UDP-N-acetylglucosamin difosforylázová aktivita
GO:0003987	2	0.225443233018957	Acetát-CoA ligázová aktivita
GO:0003994	2	0.225443233018957	Akonitát hydratázová aktivita
GO:0003999	2	0.225443233018957	Adenin fosforibosyltransferázová aktivita
GO:0004000	2	0.225443233018957	Adenosin deaminázová aktivita
GO:0004013	2	0.225443233018957	Adenosylhomocysteinázová aktivita
GO:0004018	2	0.225443233018957	N ⁶ -(1,2-dikarboxyetyl)AMP AMP-lyázová (fumarát-tvořící) aktivita
GO:0004019	2	0.225443233018957	Adenylosukcinát syntázová aktivita
GO:0004020	2	0.225443233018957	Adenylylsulfát kinázová aktivita
GO:0004044	2	0.225443233018957	Amidofosforibosyltransferázová aktivita
GO:0004045	2	0.225443233018957	Aminoacyl-tRNA hydrolázová aktivita
GO:0004055	2	0.225443233018957	Argininosukcinát syntázová aktivita
GO:0004070	2	0.225443233018957	Aspartát karbamoyltransferázová aktivita
GO:0004072	2	0.225443233018957	Aspartát kinázová aktivita
GO:0004073	2	0.225443233018957	Aspartát-semialdehyd dehydrogenázová aktivita
GO:0004074	2	0.225443233018957	Biliverdin reduktázová aktivita
GO:0004075	2	0.225443233018957	Biotin karboxylázová aktivita
GO:0004084	2	0.225443233018957	Větvená-řetěžená-aminokyselina transaminační aktivita
GO:0004107	2	0.225443233018957	Chorismát syntázová aktivita
GO:0004108	2	0.225443233018957	Citrát (Si)-syntázová aktivita
GO:0004109	2	0.225443233018957	Koproporfyrinogen oxidázová aktivita
GO:0004132	2	0.225443233018957	dCMP deaminázová aktivita
GO:0004146	2	0.225443233018957	Dihydrofolát reduktázová aktivita
GO:0004150	2	0.225443233018957	Dihydroneopterin aldolázová aktivita
GO:0004156	2	0.225443233018957	Dihydropteroát syntázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0004158	2	0.225443233018957	Dihydroorotát oxidázová aktivita
GO:0004160	2	0.225443233018957	Dihydroxy-kyselina dehydratázová aktivita
GO:0004178	2	0.225443233018957	Leucyl aminopeptidázová aktivita
GO:0004309	2	0.225443233018957	Exopolyfosfatázová aktivita
GO:0004314	2	0.225443233018957	[acyl-nesoucí-protein] S-malonyltransferázová aktivita
GO:0004318	2	0.225443233018957	Enoyl-[acyl-nesoucí-protein] reduktázová (NADH) aktivita
GO:0004324	2	0.225443233018957	Ferredoxin-NADP+ reduktázová aktivita
GO:0004326	2	0.225443233018957	Tetrahydrofolylpolyglutamát syntázová aktivita
GO:0004328	2	0.225443233018957	Foramidázová aktivita
GO:0004333	2	0.225443233018957	Fumarát hydratázová aktivita
GO:0004347	2	0.225443233018957	Glukóza-6-fosfát izomerázová aktivita
GO:0004349	2	0.225443233018957	Glutamát 5-kinázová aktivita
GO:0004354	2	0.225443233018957	Glutamát dehydrogenázová (NADP+) aktivita
GO:0004355	2	0.225443233018957	Glutamát syntázová (NADPH) aktivita
GO:0004358	2	0.225443233018957	Glutamát N-acetyltransferázová aktivita
GO:0004360	2	0.225443233018957	Glutamin-fruktóza-6-fosfát transaminázová (izomerizační) aktivita
GO:0004363	2	0.225443233018957	Glutathion syntázová aktivita
GO:0004365	2	0.225443233018957	Glyceraldehyd-3-fosfát dehydrogenázová (NAD+) (fosforylační) aktivita
GO:0004372	2	0.225443233018957	Glycin hydroxymetyltransferázová aktivita
GO:0004385	2	0.225443233018957	Guanylát kinázová aktivita
GO:0004399	2	0.225443233018957	Histidinol dehydrogenázová aktivita
GO:0004400	2	0.225443233018957	Histidinol-fosfát transaminázová aktivita
GO:0004401	2	0.225443233018957	Histidinol-fosfatázová aktivita
GO:0004413	2	0.225443233018957	Homoserin kinázová aktivita
GO:0004418	2	0.225443233018957	Hydroxymetyl bilan syntázová aktivita
GO:0004424	2	0.225443233018957	Imidazolglycerol-fosfát dehydratázová aktivita
GO:0004425	2	0.225443233018957	Indol-3-glycerol-fosfát syntázová aktivita
GO:0004427	2	0.225443233018957	Anorganická difosfatázová aktivita
GO:0004450	2	0.225443233018957	Izocitrát dehydrogenázová (NADP+) aktivita
GO:0004455	2	0.225443233018957	Ketol-kyselina reduktiozomerázová aktivita
GO:0004467	2	0.225443233018957	Mastná kyselina s dlouhým řetězcem-CoA ligázová aktivita
GO:0004473	2	0.225443233018957	Malát dehydrogenázová (oxaloacetát-dekarboxylační) (NADP+) aktivita
GO:0004476	2	0.225443233018957	manóza-6-fosfát izomerázová aktivita
GO:0004477	2	0.225443233018957	metenyltetrahydrofolát cyklohydrolázová aktivita
GO:0004479	2	0.225443233018957	Methionyl-tRNA formyltransferázová aktivita
GO:0004488	2	0.225443233018957	Metylenetetrahydrofolát dehydrogenázová (NADP+) aktivita
GO:0004491	2	0.225443233018957	Methylmalonát-semialdehyd dehydrogenázová (acylační) aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0004514	2	0.225443233018957	Nikotinát-nukleotid difosforylázová (karboxylační) aktivita
GO:0004516	2	0.225443233018957	Nikotinát fosforibosyltransferázová aktivita
GO:0004526	2	0.225443233018957	Ribonukleáza P
GO:0004540	2	0.225443233018957	Ribonukleázová aktivita
GO:0004550	2	0.225443233018957	Nukleosid difosfát kinázová aktivita
GO:0004563	2	0.225443233018957	Beta-N-acetylhexosaminidázová aktivita
GO:0004565	2	0.225443233018957	Beta-galaktosidázová aktivita
GO:0004585	2	0.225443233018957	Ornitin karbamoyltransferázová aktivita
GO:0004588	2	0.225443233018957	Orotát fosforibosyltransferázová aktivita
GO:0004590	2	0.225443233018957	Orotidin-5'-fosfát dekarboxylázová aktivita
GO:0004592	2	0.225443233018957	Pantoát-beta-alanin ligázová aktivita
GO:0004595	2	0.225443233018957	Pantethein-fosfát adenyltransferázová aktivita
GO:0004604	2	0.225443233018957	Fosfoadenyl-sulfát reduktázová (thioredoxin) aktivita
GO:0004605	2	0.225443233018957	Fosfatidát cytidyltransferázová aktivita
GO:0004615	2	0.225443233018957	fosfomanomutázová aktivita
GO:0004616	2	0.225443233018957	Fosfoglukonát dehydrogenáza (dekarboxylační) aktivita
GO:0004617	2	0.225443233018957	Fosfoglycerát dehydrogenázová aktivita
GO:0004618	2	0.225443233018957	Fosfoglycerát kinázová aktivita
GO:0004632	2	0.225443233018957	Fosfopantothenát--cystein lligázová aktivita
GO:0004633	2	0.225443233018957	Fosfopantothenoylcystein dekarboxylázová aktivita
GO:0004634	2	0.225443233018957	Fosfopyruvát hydratázová aktivita
GO:0004635	2	0.225443233018957	Fosforibosyl-AMP cyklohydrolázová aktivita
GO:0004636	2	0.225443233018957	Fosforibosyl-ATP difosfatázová aktivita
GO:0004637	2	0.225443233018957	Fosforibosylamin-glycin ligázová aktivita
GO:0004639	2	0.225443233018957	Fosforibosylaminoimidazolsukcinokarboxamid syntázová aktivita
GO:0004641	2	0.225443233018957	Fosforibosylformylglycinamidin cyklo-ligázová aktivita
GO:0004643	2	0.225443233018957	Fosforibosylaminoimidazolkarboxamid formyltransferázová aktivita
GO:0004644	2	0.225443233018957	Fosforibosylglycinamid formyltransferázová aktivita
GO:0004654	2	0.225443233018957	Polyribonukleotid nukleotidyltransferázová aktivita
GO:0004657	2	0.225443233018957	Prolin dehydrogenázová aktivita
GO:0004659	2	0.225443233018957	Prenyltransferázová aktivita
GO:0004664	2	0.225443233018957	Prefenát dehydratázová aktivita
GO:0004725	2	0.225443233018957	Protein tyrosin fosfatázová aktivita
GO:0004742	2	0.225443233018957	Dihydrolipoyllysin-reziduum acetyltransferázová aktivita
GO:0004749	2	0.225443233018957	Ribóza fosfát difosfokinázová aktivita
GO:0004750	2	0.225443233018957	Ribulóza-fosfát 3-epimerázová aktivita
GO:0004751	2	0.225443233018957	Ribóza-5-fosfát izomerázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0004756	2	0.225443233018957	Selenid, voda dikinázová aktivita
GO:0004777	2	0.225443233018957	Sukcinát-semialdehyd dehydrogenázová (NAD ⁺) aktivita
GO:0004781	2	0.225443233018957	Sulfát adenyltransferázová (ATP) aktivita
GO:0004789	2	0.225443233018957	Thiamin-fosfát difosforylázová aktivita
GO:0004807	2	0.225443233018957	Trióza-fosfát izomerázová aktivita
GO:0004813	2	0.225443233018957	Alanin-tRNA ligázová aktivita
GO:0004814	2	0.225443233018957	Arginin-tRNA ligázová aktivita
GO:0004815	2	0.225443233018957	Aspartát-tRNA ligázová aktivita
GO:0004816	2	0.225443233018957	Asparagin-tRNA ligázová aktivita
GO:0004817	2	0.225443233018957	Cystein-tRNA ligázová aktivita
GO:0004818	2	0.225443233018957	Glutamát-tRNA ligázová aktivita
GO:0004821	2	0.225443233018957	Histidin-tRNA ligázová aktivita
GO:0004822	2	0.225443233018957	Isoleucin-tRNA ligázová aktivita
GO:0004823	2	0.225443233018957	Leucin-tRNA ligázová aktivita
GO:0004824	2	0.225443233018957	Lysin-tRNA ligázová aktivita
GO:0004828	2	0.225443233018957	Serin-tRNA ligázová aktivita
GO:0004829	2	0.225443233018957	Threonin-tRNA ligázová aktivita
GO:0004830	2	0.225443233018957	Tryptofan-tRNA ligázová aktivita
GO:0004831	2	0.225443233018957	Tyrosin-tRNA ligázová aktivita
GO:0004832	2	0.225443233018957	Valin-tRNA ligázová aktivita
GO:0004849	2	0.225443233018957	Uridin kinázová aktivita
GO:0004853	2	0.225443233018957	Uroporphyrinogen dekarboxylázová aktivita
GO:0004871	2	0.225443233018957	Transdukce signálu
GO:0005315	2	0.225443233018957	Transmembránový přenos anorganického fosfátu
GO:0008097	2	0.225443233018957	Vazba 5S rRNA
GO:0008124	2	0.225443233018957	4-alfa-hydroxytetrahydrobiopterin dehydratázová aktivita
GO:0008173	2	0.225443233018957	RNA metyltransferázová aktivita
GO:0008176	2	0.225443233018957	tRNA (guanin-N7-)-methyltransferázová aktivita
GO:0008444	2	0.225443233018957	CDP-diacylglycerol-glycerol-3-fosfát 3-fosfatidyltransferázová aktivita
GO:0008446	2	0.225443233018957	GDP-manóza 4,6-dehydratázová aktivita
GO:0008448	2	0.225443233018957	N-acetylglukosamin-6-fosfát deacetylázová aktivita
GO:0008479	2	0.225443233018957	Queuin tRNA-ribosyltransferázová aktivita
GO:0008531	2	0.225443233018957	Riboflavin kinázová aktivita
GO:0008534	2	0.225443233018957	Oxidovaná léze purinové báze DNA N-glykosylázová aktivita
GO:0008661	2	0.225443233018957	1-deoxy-D-xylulóza-5-fosfát syntázová aktivita
GO:0008679	2	0.225443233018957	2-hydroxy-3-oxopropionát reduktázová aktivita
GO:0008705	2	0.225443233018957	Methionin syntázová aktivita
GO:0008716	2	0.225443233018957	D-alanin-D-alanin ligázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0008718	2	0.225443233018957	D-aminokyselina dehydrogenázová aktivita
GO:0008720	2	0.225443233018957	D-laktát dehydrogenázová aktivita
GO:0008734	2	0.225443233018957	L-aspartát oxidázová aktivita
GO:0008756	2	0.225443233018957	o-sukcinylbenzoát-CoA ligázová aktivita
GO:0008760	2	0.225443233018957	UDP-N-acetylglukosamin 1-karboxyvinyltransferázová aktivita
GO:0008761	2	0.225443233018957	UDP-N-acetylglukosamin 2-epimerázová aktivita
GO:0008762	2	0.225443233018957	UDP-N-acetylmuramát dehydrogenázová aktivita
GO:0008763	2	0.225443233018957	UDP-N-acetylmuramát-L-alanin ligázová aktivita
GO:0008764	2	0.225443233018957	UDP-N-acetylmuramoylalanin-D-glutamát ligázová aktivita
GO:0008765	2	0.225443233018957	UDP-N-acetylmuramoylalanin-D-glutamát-2,6-diaminopimelát lligázová aktivita
GO:0008776	2	0.225443233018957	Acetát kinázová aktivita
GO:0008780	2	0.225443233018957	Acyl-[acyl-nesoucí-protein]-UDP-N-acetylglukosamin O-acyltransferázová aktivita
GO:0008784	2	0.225443233018957	Alanin racemázová aktivita
GO:0008792	2	0.225443233018957	Arginin dekarboxylázová aktivita
GO:0008795	2	0.225443233018957	NAD ⁺ syntázová aktivita
GO:0008811	2	0.225443233018957	Chloramfenikol O-acetyltransferázová aktivita
GO:0008824	2	0.225443233018957	kyanát hydratázová aktivita
GO:0008835	2	0.225443233018957	Diaminohydroxyfosforibosylaminopyrimidin deaminázová aktivita
GO:0008837	2	0.225443233018957	Diaminopimelát epimerázová aktivita
GO:0008841	2	0.225443233018957	Dihydrofolát syntázová aktivita
GO:0008853	2	0.225443233018957	Exodeoxyribonukleáza III
GO:0008859	2	0.225443233018957	Exoribonukleáza II
GO:0008864	2	0.225443233018957	Formyltetrahydrofolát deformylázová aktivita
GO:0008865	2	0.225443233018957	Fruktokinázová aktivita
GO:0008878	2	0.225443233018957	Glukóza-1-fosfát adenylyltransferázová aktivita
GO:0008879	2	0.225443233018957	Glukóza-1-fosfát tymidyltransferázová aktivita
GO:0008881	2	0.225443233018957	Glutamát racemázová aktivita
GO:0008883	2	0.225443233018957	Glutamyl-tRNA reduktázová aktivita
GO:0008888	2	0.225443233018957	Glycerol dehydrogenázová aktivita
GO:0008889	2	0.225443233018957	Glycerofosfodiester fosfodiesterázová aktivita
GO:0008892	2	0.225443233018957	Guanin deaminázová aktivita
GO:0008893	2	0.225443233018957	Guanosin-3',5'-bis(difosfát) 3'-difosfatázová aktivita
GO:0008897	2	0.225443233018957	Holo-[acyl-nesoucí-protein] syntázová aktivita
GO:0008901	2	0.225443233018957	Ferredoxin hydrogenázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0008914	2	0.225443233018957	Leucyltransferázová aktivita
GO:0008925	2	0.225443233018957	Maltóza O-acetyltransferázová aktivita
GO:0008928	2	0.225443233018957	Manóza-1-fosfát guanylyltransferázová (GDP) aktivita
GO:0008935	2	0.225443233018957	1,4-dihydroxy-2-naftoyl-CoA syntázová aktivita
GO:0008936	2	0.225443233018957	Nikotinamidázová aktivita
GO:0008939	2	0.225443233018957	Nikotinát-nukleotid-dimethylbenzimidazol fosforibosyltransferázová aktivita
GO:0008963	2	0.225443233018957	Fosfo-N-acetylmuramoyl-pentapeptid-transferázová aktivita
GO:0008974	2	0.225443233018957	Fosforibulokinázová aktivita
GO:0008982	2	0.225443233018957	Protein-N(PI)-fosfohistidin-cukr fosfotransferázová aktivita
GO:0009007	2	0.225443233018957	Místně specifická DNA-metyltransferázová (adenin-specifická) aktivita
GO:0009019	2	0.225443233018957	tRNA (guanin-N1-)-metyltransferázová aktivita
GO:0009022	2	0.225443233018957	tRNA nukleotidyltransferázová aktivita
GO:0009030	2	0.225443233018957	Thiamin-fosfát kinázová aktivita
GO:0009042	2	0.225443233018957	Valin-pyruvát transaminázová aktivita
GO:0015075	2	0.225443233018957	Transmembránový transportér iontů
GO:0015093	2	0.225443233018957	Transmembránový transportér iontů železa
GO:0015105	2	0.225443233018957	Transmembránový transportér arzenu
GO:0015129	2	0.225443233018957	Transmembránový transportér laktátu
GO:0015137	2	0.225443233018957	Transmembránový transportér citrátu
GO:0016041	2	0.225443233018957	Glutamát syntázová (ferredoxin) aktivita
GO:0016163	2	0.225443233018957	Nitrogenázová aktivita
GO:0016223	2	0.225443233018957	Beta-alanin-pyruvát transaminázová aktivita
GO:0016767	2	0.225443233018957	Geranylgeranyl-difosfát geranylgeranyltransferázová aktivita
GO:0016852	2	0.225443233018957	Sirohydrochlorin kobaltochelátázová aktivita
GO:0016994	2	0.225443233018957	Prekorrin-6A reduktázová aktivita
GO:0017061	2	0.225443233018957	S-metyl-5-thioadenosin fosforylázová aktivita
GO:0017103	2	0.225443233018957	UTP:galaktóza-1-fosfát uridylyltransferázová aktivita
GO:0017168	2	0.225443233018957	5-oxoprolinázová (ATP-hydrolyzující) aktivita
GO:0018849	2	0.225443233018957	Mukonát cykloizomerázová aktivita
GO:0019134	2	0.225443233018957	Glukosamin-1-fosfát N-acetyltransferázová aktivita
GO:0030060	2	0.225443233018957	L-malát dehydrogenázová aktivita
GO:0030234	2	0.225443233018957	Regulátor enzymů
GO:0030272	2	0.225443233018957	5-formyltetrahydrofolát cyklo-ligázová aktivita
GO:0030604	2	0.225443233018957	1-deoxy-D-xylulóza-5-fosfát reduktiozomerázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0030788	2	0.225443233018957	Prekorin-2 C20-metyltransferázová aktivita
GO:0042286	2	0.225443233018957	Glutamát-1-semialdehyd 2,1-aminomutázová aktivita
GO:0043169	2	0.225443233018957	Vazba kationtů
GO:0046026	2	0.225443233018957	Prekorin-4 C11-metyltransferázová aktivita
GO:0046406	2	0.225443233018957	Mg protoporfyrin IX metyltransferázová aktivita
GO:0046523	2	0.225443233018957	S-metyl-5-thioribóza-1-fosfát izomerázová aktivita
GO:0046872	2	0.225443233018957	Vazba kovových iontů
GO:0047100	2	0.225443233018957	Glyceraldehyd-3-fosfát dehydrogenázová (NADP+) (fosforylační) aktivita
GO:0047465	2	0.225443233018957	N-acylglukosamin-6-fosfát 2-epimerázová aktivita
GO:0047905	2	0.225443233018957	fruktóza-6-fosfát fosfoketolázová aktivita
GO:0047952	2	0.225443233018957	Glycerol-3-fosfát dehydrogenázová [NAD(P)+] aktivita
GO:0048037	2	0.225443233018957	Vazba kofaktorů
GO:0048307	2	0.225443233018957	Ferredoxin-nitrit reduktázová aktivita
GO:0048472	2	0.225443233018957	Threonin-fosfát dekarboxylázová aktivita
GO:0050193	2	0.225443233018957	fosfoketolázová aktivita
GO:0050278	2	0.225443233018957	Sedoheptulóza-bisfosfatázová aktivita
GO:0050281	2	0.225443233018957	Serin-glyoxylát transaminázová aktivita
GO:0050311	2	0.225443233018957	Sulfit reduktázová (ferredoxin) aktivita
GO:0050380	2	0.225443233018957	Undekaprenyl-difosfatázová aktivita
GO:0050511	2	0.225443233018957	Undekaprenyldifosfo-muramoylpentapeptid beta-N-acetylglukosaminyltransferázová aktivita
GO:0050515	2	0.225443233018957	4-(cytidin 5'-difosfo)-2-C-metyl-D-erytritol kinázová aktivita
GO:0050518	2	0.225443233018957	2-C-metyl-D-erythritol 4-fosfát cytidyltransferázová aktivita
GO:0050532	2	0.225443233018957	2-fosfosulfolaktát fosfatázová aktivita
GO:0050570	2	0.225443233018957	4-hydroxythreonin-4-fosfát dehydrogenázová aktivita
GO:0050620	2	0.225443233018957	Fykocyanobilin:ferredoxin oxidoreduktázová aktivita
GO:0051745	2	0.225443233018957	4-hydroxy-3-metylbut-2-en-1-yl difosfát reduktázová aktivita
GO:0051989	2	0.225443233018957	Koproporfyrinogen dehydrogenázová aktivita
GO:0000287	1	0.899218808659587	Vazba iontů hořčiku
GO:0003844	1	0.899218808659587	1,4-alfa-glukan větvicí enzymatická aktivita
GO:0003857	1	0.899218808659587	3-hydroxyacyl-CoA dehydrogenázová aktivita
GO:0003858	1	0.899218808659587	3-hydroxybutyrát dehydrogenázová aktivita
GO:0003874	1	0.899218808659587	6-pyruvoyltetrahydropterin syntázová aktivita
GO:0003961	1	0.899218808659587	O-acetylhomoserin aminokarboxypropyltransferázová aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0003963	1	0.899218808659587	RNA-3'-fosfát cyklázová aktivita
GO:0003998	1	0.899218808659587	Acylfosfatázová aktivita
GO:0004066	1	0.899218808659587	Asparagin syntázová (glutamin-hydrolyzující) aktivita
GO:0004076	1	0.899218808659587	Biotin syntázová aktivita
GO:0004085	1	0.899218808659587	Butyryl-CoA dehydrogenázová aktivita
GO:0004096	1	0.899218808659587	Katalázová aktivita
GO:0004139	1	0.899218808659587	Deoxyribóza-fosfát aldolázová aktivita
GO:0004140	1	0.899218808659587	Defosfo-CoA kinázová aktivita
GO:0004157	1	0.899218808659587	Dihydropyrimidinázová aktivita
GO:0004170	1	0.899218808659587	dUTP difosfatázová aktivita
GO:0004300	1	0.899218808659587	Enoyl-CoA hydratázová aktivita
GO:0004308	1	0.899218808659587	Exo-alfa-sialidázová aktivita
GO:0004311	1	0.899218808659587	Farnesyltranstransferázová aktivita
GO:0004335	1	0.899218808659587	Galaktokinázová aktivita
GO:0004342	1	0.899218808659587	Glukosamin-6-fosfát deaminázová aktivita
GO:0004410	1	0.899218808659587	Homocitrát syntázová aktivita
GO:0004420	1	0.899218808659587	Hydroxymetylglutaryl-CoA reduktázová (NADPH) aktivita
GO:0004556	1	0.899218808659587	Alfa-amylázová aktivita
GO:0004568	1	0.899218808659587	Chitinázová aktivita
GO:0004594	1	0.899218808659587	Pantothenát kinázová aktivita
GO:0004601	1	0.899218808659587	Peroxidázová aktivita
GO:0004612	1	0.899218808659587	fosfoenolpyruvát karboxykinázová (ATP) aktivita
GO:0004614	1	0.899218808659587	Fosfoglukomutázová aktivita
GO:0004757	1	0.899218808659587	Sepiapterin reduktázová aktivita
GO:0004766	1	0.899218808659587	Spermidin syntázová aktivita
GO:0004799	1	0.899218808659587	Tymidylát syntázová aktivita
GO:0005507	1	0.899218808659587	Vazba iontů mědi
GO:0005529	1	0.899218808659587	Vazba uhlovodíků
GO:0008233	1	0.899218808659587	Peptidázová aktivita
GO:0008477	1	0.899218808659587	Purin nukleosidázová aktivita
GO:0008514	1	0.899218808659587	Transmembránový transport organických aniontů
GO:0008662	1	0.899218808659587	1-fosfofruktokinázová aktivita
GO:0008675	1	0.899218808659587	2-dehydro-3-deoxy-fosfoglukonát aldolázová aktivita
GO:0008677	1	0.899218808659587	2-dehydropantoát 2-reduktázová aktivita
GO:0008700	1	0.899218808659587	4-hydroxy-2-oxoglutarát aldolázová aktivita
GO:0008710	1	0.899218808659587	8-amino-7-oxononanoát syntázová aktivita
GO:0008726	1	0.899218808659587	Alkanesulfonát monooxygenázová aktivita
GO:0008752	1	0.899218808659587	FMN reduktázová aktivita
GO:0008767	1	0.899218808659587	UDP-galaktopyranóza mutázová aktivita
GO:0008794	1	0.899218808659587	Arzeničnan reduktázová (glutaredoxin) aktivita

Tab. 9: Skupiny genů třetí kategorie anotovaných GO analýzou *J. apiculata*

GO:0008804	1	0.899218808659587	Karbamát kinázová aktivita
GO:0008847	1	0.899218808659587	Enterobacter ribonukleázová aktivita
GO:0008861	1	0.899218808659587	Formiát C-acetyltransferázová aktivita
GO:0008870	1	0.899218808659587	Galaktosid O-acetyltransferázová aktivita
GO:0008923	1	0.899218808659587	Lysin dekarboxylázová aktivita
GO:0008929	1	0.899218808659587	Metylglyoxal syntázová aktivita
GO:0008965	1	0.899218808659587	Fosfoenolpyruvát-protein fosfotransferázová aktivita
GO:0008992	1	0.899218808659587	Represor LexA
GO:0010181	1	0.899218808659587	Vazba FMN
GO:0015423	1	0.899218808659587	Maltózu-transportující ATPázová aktivita
GO:0016301	1	0.899218808659587	Kinázová aktivita
GO:0018658	1	0.899218808659587	Salicylát 1-monooxygenázová aktivita
GO:0030976	1	0.899218808659587	Vazba Thiamin pyrofosfátu
GO:0031176	1	0.899218808659587	Endo-1,4-beta-xylanázová aktivita
GO:0031216	1	0.899218808659587	Neopullulanázová aktivita
GO:0042972	1	0.899218808659587	Licheninázová aktivita
GO:0045552	1	0.899218808659587	Dihydrokaempferol 4-reduktázová aktivita
GO:0046316	1	0.899218808659587	Glukonokinázová aktivita
GO:0047330	1	0.899218808659587	Polyfosfát-glukóza fosfotransferázová aktivita
GO:0047470	1	0.899218808659587	(1,4)-alfa-D-glukan 1-alfa-D-glukosylmutázová aktivita
GO:0047656	1	0.899218808659587	alfa, alfa-trehalóza fosforylázová aktivita
GO:0047733	1	0.899218808659587	CDP-glukóza 4,6-dehydratázová aktivita
GO:0047936	1	0.899218808659587	Glukóza 1-dehydrogenázová [NAD(P)] aktivita
GO:0050082	1	0.899218808659587	Maltóza fosforylázová aktivita
GO:0050262	1	0.899218808659587	Ribosylnikotinamid kinázová aktivita
GO:0050342	1	0.899218808659587	Tokoferol O-metyltransferázová aktivita

Tab. 17: Skupiny genů druhé kategorie anotovaných GO analýzou *Synechococcus* sp. S4

Identifikační číslo	Počet anotovaných genů	p-value	Skupina
GO:0006412	52	2.31251232466336e-155	Translace
GO:0006118	15	1.04875939979971e-41	Transport elektronů
GO:0006810	13	1.3375567289395e-35	Transport
GO:0006457	8	1.93961578323872e-20	Skládání proteinů
GO:0006281	6	1.80019164179392e-14	Reparace DNA
GO:0006413	4	1.15846272386253e-08	Iniciace translace

Tab. 17: Skupiny genů druhé kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0008152	4	1.15846272386253e-08	Metabolické procesy
GO:0051276	4	1.15846272386253e-08	Organizace chromozómu
GO:0006260	3	6.81110831874442e-06	Replikace DNA
GO:0006310	3	6.81110831874442e-06	Rekombinace DNA
GO:0006352	3	6.81110831874442e-06	DNA závislá iniciace transkripce
GO:0006464	3	6.81110831874442e-06	Modifikace buněčných proteinů
GO:0009058	3	6.81110831874442e-06	Biosyntetické procesy
GO:0009236	3	6.81110831874442e-06	Biosyntéza kobalaminu
GO:0006289	2	0.00248505302154037	Nukleotidová excizní oprava
GO:0006298	2	0.00248505302154037	Oprava chybného párování
GO:0006304	2	0.00248505302154037	Modifikace DNA
GO:0006364	2	0.00248505302154037	Zpracování rRNA
GO:0006415	2	0.00248505302154037	Terminace translace
GO:0006461	2	0.00248505302154037	Skládání proteinových komplexů
GO:0006508	2	0.00248505302154037	Proteolýza
GO:0006605	2	0.00248505302154037	Cílení proteinů
GO:0006614	2	0.00248505302154037	SRP závislé kotranslační cílení proteinů do membrány
GO:0006629	2	0.00248505302154037	Metabolismus lipidů
GO:0006807	2	0.00248505302154037	Metabolismus dusíkatých sloučenin
GO:0006808	2	0.00248505302154037	Regulace využití dusíku
GO:0006812	2	0.00248505302154037	Transport kationtů
GO:0006813	2	0.00248505302154037	Transport draselných iontů
GO:0006814	2	0.00248505302154037	Transport sodných iontů
GO:0006821	2	0.00248505302154037	Transport chloridů
GO:0009306	2	0.00248505302154037	Sekrece proteinů
GO:0015684	2	0.00248505302154037	Transport železitých iontů
GO:0019538	2	0.00248505302154037	Metabolismus proteinů
GO:0044267	2	0.00248505302154037	Metabolismus buněčných proteinů
GO:0045449	2	0.00248505302154037	DNA závislá regulace transkripce
GO:0000105	1	0.295151638420477	Biosyntéza histidinu
GO:0005975	1	0.295151638420477	Metabolismus uhlovodíků
GO:0006259	1	0.295151638420477	Metabolismus DNA
GO:0006284	1	0.295151638420477	Bázová excizní oprava
GO:0006396	1	0.295151638420477	Zpracování RNA
GO:0006414	1	0.295151638420477	Translační elongace
GO:0006424	1	0.295151638420477	Aminoacylace glutamyl-tRNA

Tab. 17: Skupiny genů druhé kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0006479	1	0.295151638420477	Metylace proteinů
GO:0006546	1	0.295151638420477	Katabolismus glycinu
GO:0006633	1	0.295151638420477	Biosyntéza mastných kyselin
GO:0006790	1	0.295151638420477	Metabolismus sloučenin síry
GO:0006885	1	0.295151638420477	Regulace pH
GO:0006886	1	0.295151638420477	Transport vnitrobuněčných proteinů
GO:0007046	1	0.295151638420477	Biogeneze ribozómů
GO:0007155	1	0.295151638420477	Adheze buňky
GO:0008033	1	0.295151638420477	Zpracování tRNA
GO:0008616	1	0.295151638420477	Biosyntéza queuosinu
GO:0009228	1	0.295151638420477	Biosyntéza thiaminu
GO:0015031	1	0.295151638420477	Transport proteinů
GO:0015727	1	0.295151638420477	Transport laktátu
GO:0015746	1	0.295151638420477	Transport citrátu
GO:0016036	1	0.295151638420477	Buněčná odpověď na nedostatek fosfátů
GO:0017004	1	0.295151638420477	Složení cytochromového komplexu
GO:0017038	1	0.295151638420477	Import proteinů
GO:0019627	1	0.295151638420477	Metabolismus močoviny
GO:0042254	1	0.295151638420477	Biogeneze ribozómů

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

Identifikační číslo	Počet anotovaných genů	p-value	Skupina
GO:0003735	46	6.70202709840456e-137	Strukturní složka ribozómu
GO:0005524	39	2.27286753842611e-115	Vazba ATP
GO:0005506	17	8.01086631298662e-48	Vazba iontů železa
GO:0003824	15	1.04875939979971e-41	Katalytická aktivita
GO:0003723	14	1.18907871895112e-38	Vazba RNA
GO:0005525	14	1.18907871895112e-38	Vazba GTP
GO:0005554	14	1.18907871895112e-38	Molekulární funkce
GO:0009055	14	1.18907871895112e-38	Přenos elektronů
GO:0003676	13	1.3375567289395e-35	Vazba nukleové kyseliny
GO:0016887	13	1.3375567289395e-35	ATPázová aktivita
GO:0003677	12	1.48987286494731e-32	Vazba DNA

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0003700	9	1.88172930115133e-23	Sekvenčně specifický DNA vazebný transkripční faktor
GO:0005215	8	1.93961578323872e-20	Transportérová aktivita
GO:0008080	7	1.92230311556036e-17	N-acetyltransferázová aktivita
GO:0003887	6	1.80019164179392e-14	DNA řízená DNA polymerázová aktivita
GO:0004016	6	1.80019164179392e-14	Adenylát cyklázová aktivita
GO:0004129	6	1.80019164179392e-14	Cytochrom C oxidázová aktivita
GO:0004386	6	1.80019164179392e-14	Helikázová aktivita
GO:0005515	6	1.80019164179392e-14	Vazba proteinů
GO:0016491	6	1.80019164179392e-14	Oxidoreduktázová aktivita
GO:0046933	6	1.80019164179392e-14	Rotační mechanismus ATP syntázy, přenos vodíkových iontů
GO:0046961	6	1.80019164179392e-14	Rotační mechanismus ATPázy, přenos protonů
GO:0051082	6	1.80019164179392e-14	Vazba nesložených proteinů
GO:0003899	5	1.54730321835931e-11	DNA řízená RNA polymerázová aktivita
GO:0004008	5	1.54730321835931e-11	Cu-exportní ATPázová aktivita
GO:0016151	5	1.54730321835931e-11	Vazba kationtů niklu
GO:0047985	5	1.54730321835931e-11	Dehydrogenázová aktivita
GO:0003743	4	1.15846272386253e-08	Translační iniciační faktor
GO:0003978	4	1.15846272386253e-08	UDP-glukóza 4-epimerázová aktivita
GO:0004069	4	1.15846272386253e-08	L-aspartát:2-oxoglutarát aminotransferázová aktivita
GO:0004089	4	1.15846272386253e-08	Karbonát dehydratázová aktivita
GO:0004179	4	1.15846272386253e-08	Alanyl aminopeptidázová aktivita
GO:0008026	4	1.15846272386253e-08	ATP-dependentní helikázová aktivita
GO:0008270	4	1.15846272386253e-08	Vazba iontů zinku
GO:0008324	4	1.15846272386253e-08	Transmembránový transport kationtů
GO:0008462	4	1.15846272386253e-08	Clp endopeptidázová aktivita
GO:0008854	4	1.15846272386253e-08	Exodeoxyribonukleáza V
GO:0031072	4	1.15846272386253e-08	Vazba heat shock proteinů
GO:0003918	3	6.81110831874442e-06	DNA topoizomerázová (ATP-hydrolyzující) aktivita
GO:0003954	3	6.81110831874442e-06	NADH dehydrogenázová aktivita
GO:0004048	3	6.81110831874442e-06	Antranilát fosforibozyltransferázová aktivita
GO:0004124	3	6.81110831874442e-06	Cystein syntázová aktivita
GO:0004146	3	6.81110831874442e-06	Dihydrofolát reduktázová aktivita
GO:0004356	3	6.81110831874442e-06	Glutamát-amoniak ligázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0004642	3	6.81110831874442e-06	Fosforibosylformylglycinamidin syntázová aktivita
GO:0004645	3	6.81110831874442e-06	Fosforylázová aktivita
GO:0004721	3	6.81110831874442e-06	Fosfoprotein fosfatázová aktivita
GO:0004834	3	6.81110831874442e-06	Tryptofan syntázová aktivita
GO:0005488	3	6.81110831874442e-06	Vazba
GO:0008750	3	6.81110831874442e-06	NAD(P)+ transhydrogenázová (AB-specifická) aktivita
GO:0008817	3	6.81110831874442e-06	Kob(I)yrinová kyselina a,c-diamid adenozylntransferázová aktivita
GO:0008955	3	6.81110831874442e-06	Peptidoglykan glykosyltransferázová aktivita
GO:0009001	3	6.81110831874442e-06	Serin O-acetyltransferázová aktivita
GO:0009002	3	6.81110831874442e-06	Serin-typ D-Ala-D-Ala karboxypeptidázová aktivita
GO:0009039	3	6.81110831874442e-06	Ureázová aktivita
GO:0016787	3	6.81110831874442e-06	Hydrolázová aktivita
GO:0016851	3	6.81110831874442e-06	Chelatace hořčiku
GO:0016987	3	6.81110831874442e-06	Sigma faktor
GO:0031071	3	6.81110831874442e-06	Cystein desulfurázová aktivita
GO:0000104	2	0.00248505302154037	Sukcinát dehydrogenázová aktivita
GO:0001882	2	0.00248505302154037	Vazba nukleosidů
GO:0003678	2	0.00248505302154037	DNA helikázová aktivita
GO:0003684	2	0.00248505302154037	Vazba poškozené DNA
GO:0003697	2	0.00248505302154037	Vazba jednovláknové DNA
GO:0003747	2	0.00248505302154037	Translační uvolňující faktor
GO:0003825	2	0.00248505302154037	Alfa,alfa-trehalóza-fosfát syntázová (UDP-tvořící) aktivita
GO:0003861	2	0.00248505302154037	3-isopropylmalát dehydratázová aktivita
GO:0003879	2	0.00248505302154037	ATP fosforibosyltransferázová aktivita
GO:0003922	2	0.00248505302154037	GMP syntázová (glutamin-hydrolyzující) aktivita
GO:0003938	2	0.00248505302154037	IMP dehydrogenázová aktivita
GO:0003951	2	0.00248505302154037	NAD+ kinázová aktivita
GO:0003984	2	0.00248505302154037	Acetolaktát syntázová aktivita
GO:0003989	2	0.00248505302154037	Acetyl-CoA karboxylázová aktivita
GO:0003992	2	0.00248505302154037	N2-acetyl-L-ornitin:2-oxoglutarát 5-aminotransferázová aktivita
GO:0003994	2	0.00248505302154037	Akonitát hydratázová aktivita
GO:0004035	2	0.00248505302154037	Alkalicko fosfatázová aktivita
GO:0004049	2	0.00248505302154037	Antranilát syntázová aktivita
GO:0004088	2	0.00248505302154037	Karbamoyl-fosfát syntázová (glutamin-hydrolyzující) aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0004148	2	0.00248505302154037	Dihydrolipoyl dehydrogenázová aktivita
GO:0004151	2	0.00248505302154037	Dihydroorotázová aktivita
GO:0004311	2	0.00248505302154037	Farnesyltranstransferázová aktivita
GO:0004316	2	0.00248505302154037	3-oxoacyl-[acyl-nesoucí-protein] reductázová aktivita (NADPH)
GO:0004322	2	0.00248505302154037	Ferroxidázová aktivita
GO:0004332	2	0.00248505302154037	Fruktóza-bisfosfát aldolázová aktivita
GO:0004365	2	0.00248505302154037	Glyceraldehyd-3-fosfát dehydrogenázová (NAD+) (fosforylační) aktivita
GO:0004392	2	0.00248505302154037	Hem oxygenázová (decyklizační) aktivita
GO:0004414	2	0.00248505302154037	Homoserin O-acetyltransferázová aktivita
GO:0004467	2	0.00248505302154037	Mastná kyselina s dlouhým řetězcem-CoA ligázová aktivita
GO:0004479	2	0.00248505302154037	Methionyl-tRNA formyltransferázová aktivita
GO:0004518	2	0.00248505302154037	Nukleázová aktivita
GO:0004588	2	0.00248505302154037	Orotát fosforibosyltransferázová aktivita
GO:0004590	2	0.00248505302154037	Orotidin-5'-fosfát dekarboxylázová aktivita
GO:0004619	2	0.00248505302154037	Fosfoglycerát mutázová aktivita
GO:0004638	2	0.00248505302154037	Fosforibozylaminoimidazol karboxylázová aktivita
GO:0004733	2	0.00248505302154037	Pyridoxamin-fosfát oxidázová aktivita
GO:0004739	2	0.00248505302154037	Pyruvát dehydrogenázová (acetyl-přenášející) aktivita
GO:0004766	2	0.00248505302154037	Spermidin syntázová aktivita
GO:0004784	2	0.00248505302154037	Superoxid dismutázová aktivita
GO:0004795	2	0.00248505302154037	Threonin syntázová aktivita
GO:0004801	2	0.00248505302154037	Sedoheptulóza-7-fosfát:D-glyceraldehyd-3-fosfát glyceron transferázová aktivita
GO:0004818	2	0.00248505302154037	Glutamát-tRNA ligázová aktivita
GO:0004820	2	0.00248505302154037	Glycin-tRNA ligázová aktivita
GO:0004826	2	0.00248505302154037	Fenylalanin-tRNA ligázová aktivita
GO:0004845	2	0.00248505302154037	Uracil fosforibozyltransferázová aktivita
GO:0005247	2	0.00248505302154037	Napětím řízený chloridový kanál
GO:0008124	2	0.00248505302154037	4-alfa-hydroxytetrahydrobiopterin dehydratázová aktivita
GO:0008236	2	0.00248505302154037	Peptidáza serinového typu
GO:0008460	2	0.00248505302154037	dTDP-glukóza 4,6-dehydratázová aktivita
GO:0008556	2	0.00248505302154037	Draslík transportující ATPázová aktivita
GO:0008679	2	0.00248505302154037	2-hydroxy-3-oxopropionát reductázová aktivita
GO:0008703	2	0.00248505302154037	5-amino-6-(5-fosforibozylamino)uracil reductázová aktivita
GO:0008783	2	0.00248505302154037	Agmatinázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0008800	2	0.00248505302154037	Beta-laktamázová aktivita
GO:0008829	2	0.00248505302154037	dCTP deaminázová aktivita
GO:0008865	2	0.00248505302154037	Fruktokinázová aktivita
GO:0008915	2	0.00248505302154037	Lipid-A-disacharid syntázová aktivita
GO:0008928	2	0.00248505302154037	Manóza-1-fosfát guanylyltransferázová (GDP) aktivita
GO:0008939	2	0.00248505302154037	Nikotinát-nukleotid-dimethylbenzimidazol fosforibosyltransferázová aktivita
GO:0008966	2	0.00248505302154037	Fosfoglukosamin mutázová aktivita
GO:0009378	2	0.00248505302154037	Four-way junction helikázová aktivita
GO:0015093	2	0.00248505302154037	Transmembránový transportér iontů železa
GO:0015450	2	0.00248505302154037	P-P-vazbou-hydrolyzou-řízený proteinový transmembránový transportér
GO:0016630	2	0.00248505302154037	Protochlorofylid reduktázová aktivita
GO:0016984	2	0.00248505302154037	Ribulóza-bisfosfát karboxylázová aktivita
GO:0018658	2	0.00248505302154037	Salicylát 1-monooxygenázová aktivita
GO:0020037	2	0.00248505302154037	Vazba Hemu
GO:0030234	2	0.00248505302154037	Regulátor enzymů
GO:0030528	2	0.00248505302154037	Regulátor transkripce
GO:0042586	2	0.00248505302154037	Peptid deformylázová aktivita
GO:0047360	2	0.00248505302154037	Undekaprenyl-fosfát galaktóza fosfotransferázová aktivita
GO:0047631	2	0.00248505302154037	ADP-ribóza difosfatázová aktivita
GO:0048529	2	0.00248505302154037	Mg-protoporfyrin IX monometyl ester (oxidativní) cyklázová aktivita
GO:0000156	1	0.295151638420477	Dvousložkový regulátor odpovědi
GO:0000286	1	0.295151638420477	Alanin dehydrogenázová aktivita
GO:0000287	1	0.295151638420477	Vazba iontů hořčíku
GO:0000309	1	0.295151638420477	Nickotinamid-nukleotid adenylyltransferázová aktivita
GO:0000774	1	0.295151638420477	Adenyl-nukleotid výměnný faktor
GO:0003711	1	0.295151638420477	Regulátor transkripční elongace
GO:0003746	1	0.295151638420477	Translační elongační faktor
GO:0003840	1	0.295151638420477	Gama-glutamyltransferázová aktivita
GO:0003841	1	0.295151638420477	1-acylglycerol-3-fosfát O-acyltransferázová aktivita
GO:0003856	1	0.295151638420477	3-dehydrochinát syntázová aktivita
GO:0003858	1	0.295151638420477	3-hydroxybutyrát dehydrogenázová aktivita
GO:0003862	1	0.295151638420477	3-izopropylmalát dehydrogenázová aktivita
GO:0003864	1	0.295151638420477	3-metyl-2-oxobutanoát hydroxymetyltransferázová aktivita
GO:0003874	1	0.295151638420477	6-pyruvoyltetrahydropterin syntázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0003883	1	0.295151638420477	CTP syntázová aktivita
GO:0003896	1	0.295151638420477	DNA primázová aktivita
GO:0003905	1	0.295151638420477	Alkylbáze DNA N-glykosylázová aktivita
GO:0003906	1	0.295151638420477	DNA-(apurinové nebo apyrimidinové místo) lyázová aktivita
GO:0003911	1	0.295151638420477	DNA ligázová (NAD+) aktivita
GO:0003913	1	0.295151638420477	DNA fotolyázová aktivita
GO:0003917	1	0.295151638420477	DNA topoizoimeráza typu I
GO:0003919	1	0.295151638420477	FMN adenyltransferázová aktivita
GO:0003934	1	0.295151638420477	GTP cyklohydroláza I
GO:0003935	1	0.295151638420477	GTP cyklohydroláza II
GO:0003937	1	0.295151638420477	IMP cyklohydrolázová aktivita
GO:0003942	1	0.295151638420477	N-acetyl-gama-glutamyl-fosfát reduktázová aktivita
GO:0003949	1	0.295151638420477	1-(5-fosforibosyl)-5-[(5-Fosforibosylamino)metylidenamino]imidazol-4-karboxamid izomerázová aktivita
GO:0003961	1	0.295151638420477	O-acetylhomoserin aminokarboxypropyltransferázová aktivita
GO:0003977	1	0.295151638420477	UDP-N-acetylglucosamin difosforylázová aktivita
GO:0003979	1	0.295151638420477	UDP-glukóza 6-dehydrogenázová aktivita
GO:0003987	1	0.295151638420477	Acetát-CoA ligázová aktivita
GO:0003991	1	0.295151638420477	Acetylglutamát kinázová aktivita
GO:0003998	1	0.295151638420477	Acylfosfatázová aktivita
GO:0003999	1	0.295151638420477	Adenin fosforibosyltransferázová aktivita
GO:0004013	1	0.295151638420477	Adenosylhomocysteinázová aktivita
GO:0004015	1	0.295151638420477	Adenosylmethionin-8-amino-7-oxononanoát transaminázová aktivita
GO:0004017	1	0.295151638420477	Adenylát kinázová aktivita
GO:0004018	1	0.295151638420477	N6-(1,2-dikarboxyetyl)AMP AMP-lyázová (fumarát-tvořící) aktivita
GO:0004019	1	0.295151638420477	Adenylosukcinát syntázová aktivita
GO:0004020	1	0.295151638420477	Adenylsulfát kinázová aktivita
GO:0004022	1	0.295151638420477	Alkohol dehydrogenázová (NAD) aktivita
GO:0004029	1	0.295151638420477	Aldehyd dehydrogenázová (NAD) aktivita
GO:0004044	1	0.295151638420477	Amidofosforibosyltransferázová aktivita
GO:0004045	1	0.295151638420477	Aminoacyl-tRNA hydrolázová aktivita
GO:0004055	1	0.295151638420477	Argininosukcinát syntázová aktivita
GO:0004056	1	0.295151638420477	Argininosukcinát lyázová aktivita
GO:0004065	1	0.295151638420477	Arylsulfatázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0004066	1	0.295151638420477	Asparagin syntázová (glutamin-hydrolyzující) aktivita
GO:0004070	1	0.295151638420477	Aspartát karbamoyltransferázová aktivita
GO:0004072	1	0.295151638420477	Aspartát kinázová aktivita
GO:0004073	1	0.295151638420477	Aspartát-semialdehyd dehydrogenázová aktivita
GO:0004075	1	0.295151638420477	Biotin karboxylázová aktivita
GO:0004076	1	0.295151638420477	Biotin syntázová aktivita
GO:0004084	1	0.295151638420477	Větvená-řetěžená-aminokyselina transaminační aktivita
GO:0004096	1	0.295151638420477	Katalázová aktivita
GO:0004107	1	0.295151638420477	Chorismát syntázová aktivita
GO:0004108	1	0.295151638420477	Citrát (Si)-syntázová aktivita
GO:0004109	1	0.295151638420477	Koproporfyrinogen oxidázová aktivita
GO:0004131	1	0.295151638420477	Cytosin deaminázová aktivita
GO:0004134	1	0.295151638420477	4-alfa-glukanotransferázová aktivita
GO:0004139	1	0.295151638420477	Deoxyribóza-fosfát aldolázová aktivita
GO:0004140	1	0.295151638420477	Defosfo-CoA kinázová aktivita
GO:0004141	1	0.295151638420477	Dethiobiotin syntázová aktivita
GO:0004143	1	0.295151638420477	Diacylglycerol kinázová aktivita
GO:0004145	1	0.295151638420477	Diamin N-acetyltransferázová aktivita
GO:0004150	1	0.295151638420477	Dihydroneopterin aldolázová aktivita
GO:0004156	1	0.295151638420477	Dihydropteroát syntázová aktivita
GO:0004158	1	0.295151638420477	Dihydroorotát oxidázová aktivita
GO:0004160	1	0.295151638420477	Dihydroxy-kyselina dehydratázová aktivita
GO:0004178	1	0.295151638420477	Leucyl aminopeptidázová aktivita
GO:0004239	1	0.295151638420477	Methionyl aminopeptidázová aktivita
GO:0004309	1	0.295151638420477	Exopolyfosfatázová aktivita
GO:0004314	1	0.295151638420477	[acyl-nesoucí-protein] S-malonyltransferázová aktivita
GO:0004318	1	0.295151638420477	Enoyl-[acyl-nesoucí-protein] reduktázová (NADH) aktivita
GO:0004324	1	0.295151638420477	Ferredoxin-NADP ⁺ reduktázová aktivita
GO:0004325	1	0.295151638420477	Ferrochelátázová aktivita
GO:0004326	1	0.295151638420477	Tetrahydrofolylpolyglutamát syntázová aktivita
GO:0004333	1	0.295151638420477	Fumarát hydratázová aktivita
GO:0004340	1	0.295151638420477	Glukokinázová aktivita
GO:0004342	1	0.295151638420477	Glukosamin-6-fosfát deaminázová aktivita
GO:0004345	1	0.295151638420477	Glukóza-6-fosfát dehydrogenázová aktivita
GO:0004347	1	0.295151638420477	Glukóza-6-fosfát izomerázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0004349	1	0.295151638420477	Glutamát 5-kinázová aktivita
GO:0004350	1	0.295151638420477	Glutamát-5-semialdehyd dehydrogenázová aktivita
GO:0004358	1	0.295151638420477	Glutamát N-acetyltransferázová aktivita
GO:0004360	1	0.295151638420477	Glutamin-fruktóza-6-fosfát transaminázová (izomerizační) aktivita
GO:0004362	1	0.295151638420477	Glutathion-disulfid reduktázová aktivita
GO:0004363	1	0.295151638420477	Glutathion syntázová aktivita
GO:0004372	1	0.295151638420477	Glycin hydroxymetyltransferázová aktivita
GO:0004375	1	0.295151638420477	Glycin dehydrogenázová (dekarboxylační) aktivita
GO:0004385	1	0.295151638420477	Guanylát kinázová aktivita
GO:0004399	1	0.295151638420477	Histidinol dehydrogenázová aktivita
GO:0004400	1	0.295151638420477	Histidinol-fosfát transaminázová aktivita
GO:0004401	1	0.295151638420477	Histidinol-fosfatázová aktivita
GO:0004412	1	0.295151638420477	Homoserin dehydrogenázová aktivita
GO:0004413	1	0.295151638420477	Homoserin kinázová aktivita
GO:0004416	1	0.295151638420477	Hydroxyacylglutathion hydrolázová aktivita
GO:0004418	1	0.295151638420477	Hydroxymetylbilan syntázová aktivita
GO:0004424	1	0.295151638420477	Imidazolglycerol-fosfát dehydratázová aktivita
GO:0004425	1	0.295151638420477	Indol-3-glycerol-fosfát syntázová aktivita
GO:0004427	1	0.295151638420477	Anorganická difosfatázová aktivita
GO:0004450	1	0.295151638420477	Izocitrát dehydrogenázová (NADP+) aktivita
GO:0004455	1	0.295151638420477	Ketol-kyselina reduktiozomerázová aktivita
GO:0004459	1	0.295151638420477	L-laktát dehydrogenázová aktivita
GO:0004462	1	0.295151638420477	Laktoylglutathion lyázová aktivita
GO:0004477	1	0.295151638420477	metenyltetrahydrofolát cyklohydrolázová aktivita
GO:0004478	1	0.295151638420477	Methionin adenosyltransferázová aktivita
GO:0004488	1	0.295151638420477	Metylenetetrahydrofolát dehydrogenázová (NADP+) aktivita
GO:0004514	1	0.295151638420477	Nikotinát-nukleotid difosforylázová (karboxylační) aktivita
GO:0004515	1	0.295151638420477	Nikotinát-nukleotid adenyltransferázová aktivita
GO:0004523	1	0.295151638420477	Ribonukleáza H
GO:0004525	1	0.295151638420477	Ribonukleáza III
GO:0004526	1	0.295151638420477	Ribonukleáza P
GO:0004550	1	0.295151638420477	Nukleosid difosfát kinázová aktivita
GO:0004556	1	0.295151638420477	Alfa-amylázová aktivita
GO:0004563	1	0.295151638420477	Beta-N-acetylhexosaminidázová aktivita
GO:0004585	1	0.295151638420477	Ornitin karbamoyltransferázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0004592	1	0.295151638420477	Pantoát-beta-alanin ligázová aktivita
GO:0004594	1	0.295151638420477	Pantothenát kinázová aktivita
GO:0004595	1	0.295151638420477	Pantethein-fosfát adenyltransferázová aktivita
GO:0004601	1	0.295151638420477	Peroxidázová aktivita
GO:0004604	1	0.295151638420477	Fosfoadenyl-sulfát reduktázová (thioredoxin) aktivita
GO:0004605	1	0.295151638420477	Fosfatidát cytidyltransferázová aktivita
GO:0004614	1	0.295151638420477	Fosfoglukomutázová aktivita
GO:0004616	1	0.295151638420477	Fosfoglukonát dehydrogenáza (dekarboxylační) aktivita
GO:0004617	1	0.295151638420477	Fosfoglycerát dehydrogenázová aktivita
GO:0004618	1	0.295151638420477	Fosfoglycerát kinázová aktivita
GO:0004632	1	0.295151638420477	Fosfopantothenát--cystein lligázová aktivita
GO:0004633	1	0.295151638420477	Fosfopantothenoylcystein dekarboxylázová aktivita
GO:0004634	1	0.295151638420477	Fosfopyruvát hydratázová aktivita
GO:0004635	1	0.295151638420477	Fosforibosyl-AMP cyklohydrolázová aktivita
GO:0004636	1	0.295151638420477	Fosforibosyl-ATP difosfatázová aktivita
GO:0004637	1	0.295151638420477	Fosforibosylamin-glycin ligázová aktivita
GO:0004639	1	0.295151638420477	Fosforibosylaminoimidazolsukcinokarboxamid syntázová aktivita
GO:0004640	1	0.295151638420477	Fosforibozylantranilát izomerázová aktivita
GO:0004641	1	0.295151638420477	Fosforibosylformylglycinamidin cyklo-ligázová aktivita
GO:0004643	1	0.295151638420477	Fosforibosylaminoimidazolkarboxamid formyltransferázová aktivita
GO:0004644	1	0.295151638420477	Fosforibosylglycinamid formyltransferázová aktivita
GO:0004654	1	0.295151638420477	Polyribonukleotid nukleotidyltransferázová aktivita
GO:0004655	1	0.295151638420477	Porfobilinogen syntázová aktivita
GO:0004664	1	0.295151638420477	Prefenát dehydratázová aktivita
GO:0004725	1	0.295151638420477	Protein tyrosin fosfatázová aktivita
GO:0004730	1	0.295151638420477	Pseudouridylát syntázová aktivita
GO:0004735	1	0.295151638420477	Pyrrolin-5-karboxylát reduktázová aktivita
GO:0004742	1	0.295151638420477	Dihydrolipoyllysin-reziduum acetyltransferázová aktivita
GO:0004743	1	0.295151638420477	Pyruvát kinázová aktivita
GO:0004747	1	0.295151638420477	Ribokinázová aktivita
GO:0004748	1	0.295151638420477	Ribonukleosid-difosfát reduktázová aktivita, thioredoxin disulfide jako akceptor
GO:0004749	1	0.295151638420477	Ribóza fosfát difosfokinázová aktivita
GO:0004750	1	0.295151638420477	Ribulóza-fosfát 3-epimerázová aktivita
GO:0004751	1	0.295151638420477	Ribóza-5-fosfát izomerázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0004756	1	0.295151638420477	Selenid, voda dikinázová aktivita
GO:0004757	1	0.295151638420477	Sepiapterin reduktázová aktivita
GO:0004760	1	0.295151638420477	Serin-pyruvát transaminázová aktivita
GO:0004781	1	0.295151638420477	Sulfát adenyltransferázová (ATP) aktivita
GO:0004789	1	0.295151638420477	Thiamin-fosfát difosforylázová aktivita
GO:0004794	1	0.295151638420477	L-threonin amoniak lyázová aktivita
GO:0004798	1	0.295151638420477	Tymidilát kinázová aktivita
GO:0004802	1	0.295151638420477	Transketolázová aktivita
GO:0004807	1	0.295151638420477	Trióza-fosfát izomerázová aktivita
GO:0004813	1	0.295151638420477	Alanin-tRNA ligázová aktivita
GO:0004814	1	0.295151638420477	Arginin-tRNA ligázová aktivita
GO:0004815	1	0.295151638420477	Aspartát-tRNA ligázová aktivita
GO:0004817	1	0.295151638420477	Cystein-tRNA ligázová aktivita
GO:0004821	1	0.295151638420477	Histidin-tRNA ligázová aktivita
GO:0004822	1	0.295151638420477	Isoleucin-tRNA ligázová aktivita
GO:0004823	1	0.295151638420477	Leucin-tRNA ligázová aktivita
GO:0004824	1	0.295151638420477	Lysin-tRNA ligázová aktivita
GO:0004825	1	0.295151638420477	Methionin-tRNA ligázová aktivita
GO:0004828	1	0.295151638420477	Serin-tRNA ligázová aktivita
GO:0004829	1	0.295151638420477	Threonin-tRNA ligázová aktivita
GO:0004830	1	0.295151638420477	Tryptofan-tRNA ligázová aktivita
GO:0004831	1	0.295151638420477	Tyrosin-tRNA ligázová aktivita
GO:0004832	1	0.295151638420477	Valin-tRNA ligázová aktivita
GO:0004851	1	0.295151638420477	Uroporfyrin-III C-metyltransferázová aktivita
GO:0004852	1	0.295151638420477	Uroporfyrinogen-III syntázová aktivita
GO:0004853	1	0.295151638420477	Uroporfyrinogen dekarboxylázová aktivita
GO:0005216	1	0.295151638420477	Iontový kanál
GO:0008168	1	0.295151638420477	Metyltransferázová aktivita
GO:0008173	1	0.295151638420477	RNA metyltransferázová aktivita
GO:0008176	1	0.295151638420477	tRNA (guanin-N7-)-metyltransferázová aktivita
GO:0008276	1	0.295151638420477	Protein metyltransferázová aktivita
GO:0008422	1	0.295151638420477	Beta-glukosidázová aktivita
GO:0008446	1	0.295151638420477	GDP-manóza 4,6-dehydratázová aktivita
GO:0008448	1	0.295151638420477	N-acetylglukosamin-6-fosfát deacetylázová aktivita
GO:0008451	1	0.295151638420477	X-Pro aminopeptidázová aktivita
GO:0008453	1	0.295151638420477	Alanin-glyoxylát transaminázová aktivita
GO:0008479	1	0.295151638420477	Queuin tRNA-ribosyltransferázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0008508	1	0.295151638420477	Žlučová kyselina:sodík symportérová aktivita
GO:0008531	1	0.295151638420477	Riboflavin kinázová aktivita
GO:0008534	1	0.295151638420477	Oxidovaná léze purinové báze DNA N-glykosylázová aktivita
GO:0008565	1	0.295151638420477	Transport proteinů
GO:0008661	1	0.295151638420477	1-deoxy-D-xylulóza-5-fosfát syntázová aktivita
GO:0008675	1	0.295151638420477	2-dehydro-3-deoxy-fosfoglukonát aldolázová aktivita
GO:0008685	1	0.295151638420477	2-C-metyl-D-erytritol 2,4-cyklodifosfát syntázová aktivita
GO:0008696	1	0.295151638420477	4-amino-4-deoxychorismát lyázová aktivita
GO:0008705	1	0.295151638420477	Methionin syntázová aktivita
GO:0008710	1	0.295151638420477	8-amino-7-oxononanoát syntázová aktivita
GO:0008716	1	0.295151638420477	D-alanin-D-alanin ligázová aktivita
GO:0008718	1	0.295151638420477	D-aminokyselina dehydrogenázová aktivita
GO:0008734	1	0.295151638420477	L-aspartát oxidázová aktivita
GO:0008756	1	0.295151638420477	o-sukcinylbenzoát-CoA ligázová aktivita
GO:0008760	1	0.295151638420477	UDP-N-acetylglukosamin 1-karboxyvinyltransferázová aktivita
GO:0008761	1	0.295151638420477	UDP-N-acetylglukosamin 2-epimerázová aktivita
GO:0008762	1	0.295151638420477	UDP-N-acetylmuramát dehydrogenázová aktivita
GO:0008763	1	0.295151638420477	UDP-N-acetylmuramát-L-alanin ligázová aktivita
GO:0008764	1	0.295151638420477	UDP-N-acetylmuramoylalanin-D-glutamát ligázová aktivita
GO:0008765	1	0.295151638420477	UDP-N-acetylmuramoylalanin-D-glutamát-2,6-diaminopimelát lligázová aktivita
GO:0008767	1	0.295151638420477	UDP-galaktopyranóza mutázová aktivita
GO:0008776	1	0.295151638420477	Acetát kinázová aktivita
GO:0008780	1	0.295151638420477	Acyl-[acyl-nesoucí-protein]-UDP-N-acetylglukosamin O-acyltransferázová aktivita
GO:0008781	1	0.295151638420477	N-acylneuraminát cytidyltransferázová aktivita
GO:0008784	1	0.295151638420477	Alanin racemázová aktivita
GO:0008792	1	0.295151638420477	Arginin dekarboxylázová aktivita
GO:0008794	1	0.295151638420477	Arzeničan reduktázová (glutaredoxin) aktivita
GO:0008795	1	0.295151638420477	NAD ⁺ syntázová aktivita
GO:0008830	1	0.295151638420477	dTDP-4-dehydroramnóza 3,5-epimerázová aktivita
GO:0008831	1	0.295151638420477	dTDP-4-dehydroramnóza reduktázová aktivita
GO:0008832	1	0.295151638420477	dGTPázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0008835	1	0.295151638420477	Diaminohydroxyfosforibosylaminopyrimidin deaminázová aktivita
GO:0008836	1	0.295151638420477	Diaminopimelát dekarboxylázová aktivita
GO:0008837	1	0.295151638420477	Diaminopimelát epimerázová aktivita
GO:0008841	1	0.295151638420477	Dihydrofolát syntázová aktivita
GO:0008853	1	0.295151638420477	Exodeoxyribonukleáza III
GO:0008855	1	0.295151638420477	Exodeoxyribonukleáza VII
GO:0008859	1	0.295151638420477	Exoribonukleáza II
GO:0008864	1	0.295151638420477	Formyltetrahydrofolát deformylázová aktivita
GO:0008878	1	0.295151638420477	Glukóza-1-fosfát adenyltransferázová aktivita
GO:0008879	1	0.295151638420477	Glukóza-1-fosfát tymidyltransferázová aktivita
GO:0008881	1	0.295151638420477	Glutamát racemázová aktivita
GO:0008883	1	0.295151638420477	Glutamyl-tRNA reduktázová aktivita
GO:0008888	1	0.295151638420477	Glycerol dehydrogenázová aktivita
GO:0008893	1	0.295151638420477	Guanosin-3',5'-bis(difosfát) 3'-difosfatázová aktivita
GO:0008897	1	0.295151638420477	Holo-[acyl-nesoucí-protein] syntázová aktivita
GO:0008908	1	0.295151638420477	Izochorismatázová aktivita
GO:0008909	1	0.295151638420477	Izochorismát syntázová aktivita
GO:0008923	1	0.295151638420477	Lysin dekarboxylázová aktivita
GO:0008929	1	0.295151638420477	Methylglyoxal syntázová aktivita
GO:0008935	1	0.295151638420477	1,4-dihydroxy-2-naftoyl-CoA syntázová aktivita
GO:0008963	1	0.295151638420477	Fosfo-N-acetylmuramoyl-pentapeptid-transferázová aktivita
GO:0008964	1	0.295151638420477	Fosfoenolpyruvát karboxylázová aktivita
GO:0008974	1	0.295151638420477	Fosforibulokinázová aktivita
GO:0008976	1	0.295151638420477	Polyfosfát kinázová aktivita
GO:0008983	1	0.295151638420477	Protein-glutamát O-metyltransferázová aktivita
GO:0008986	1	0.295151638420477	Pyruvát, voda dikinázová aktivita
GO:0008992	1	0.295151638420477	Represor LexA
GO:0009010	1	0.295151638420477	Sorbitol-6-fosfát 2-dehydrogenázová aktivita
GO:0009011	1	0.295151638420477	Syntáza škrobu
GO:0009018	1	0.295151638420477	Sacharóza fosforylázová aktivita
GO:0009019	1	0.295151638420477	tRNA (guanin-N1-)-metyltransferázová aktivita
GO:0009022	1	0.295151638420477	tRNA nukleotidyltransferázová aktivita
GO:0009384	1	0.295151638420477	N-acylmanózamin kinázová aktivita
GO:0015129	1	0.295151638420477	Transmembránový transportér laktátu

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0015137	1	0.295151638420477	Transmembránový transportér citrátu
GO:0015423	1	0.295151638420477	Maltózu-transportující ATPázová aktivita
GO:0016041	1	0.295151638420477	Glutamát syntázová (ferredoxin) aktivita
GO:0016619	1	0.295151638420477	Malát dehydrogenázová (oxaloacetát-dekarboxylující) aktivita
GO:0016740	1	0.295151638420477	Transferázová aktivita
GO:0016767	1	0.295151638420477	Geranylgeranyl-difosfát geranylgeranyltransferázová aktivita
GO:0016779	1	0.295151638420477	Nukleotidyltransferázová aktivita
GO:0016852	1	0.295151638420477	Sirohydrochlorin kobaltochelatazová aktivita
GO:0016993	1	0.295151638420477	Prekorin-8X metylmutázová aktivita
GO:0016994	1	0.295151638420477	Prekorin-6A reduktázová aktivita
GO:0017061	1	0.295151638420477	S-metyl-5-thioadenosin fosforylázová aktivita
GO:0017168	1	0.295151638420477	5-oxoprolinázová (ATP-hydrolyzující) aktivita
GO:0018849	1	0.295151638420477	Mukonát cykloizomerázová aktivita
GO:0019133	1	0.295151638420477	Cholin monoxygenázová aktivita
GO:0019134	1	0.295151638420477	Glukosamin-1-fosfát N-acetyltransferázová aktivita
GO:0030151	1	0.295151638420477	Vazba iontů Molybdenu
GO:0030604	1	0.295151638420477	1-deoxy-D-xylulóza-5-fosfát redukoizomerázová aktivita
GO:0030788	1	0.295151638420477	Prekorin-2 C20-metyltransferázová aktivita
GO:0030976	1	0.295151638420477	Vazba Thiamin pyrofosfátu
GO:0031176	1	0.295151638420477	Endo-1,4-beta-xylanázová aktivita
GO:0042132	1	0.295151638420477	Fruktóza 1,6-bisfosfát 1-fosfatázová aktivita
GO:0042286	1	0.295151638420477	Glutamát-1-semialdehyd 2,1-aminomutázová aktivita
GO:0042803	1	0.295151638420477	Protein homodimerizační aktivita
GO:0043169	1	0.295151638420477	Vazba kationtů
GO:0046026	1	0.295151638420477	Prekorin-4 C11-metyltransferázová aktivita
GO:0046406	1	0.295151638420477	Mg protoporfyrin IX metyltransferázová aktivita
GO:0046523	1	0.295151638420477	S-metyl-5-thioribóza-1-fosfát izomerázová aktivita
GO:0046872	1	0.295151638420477	Vazba kovových iontů
GO:0047100	1	0.295151638420477	Glyceraldehyd-3-fosfát dehydrogenázová (NADP+) (fosforylační) aktivita
GO:0047280	1	0.295151638420477	Nikotinamid fosforibosyltransferázová aktivita
GO:0047465	1	0.295151638420477	N-acylglukosamin-6-fosfát 2-epimerázová aktivita
GO:0047588	1	0.295151638420477	5-aminopentanamidázová aktivita

Tab. 18: Skupiny genů třetí kategorie anotovaných GO analýzou *Synechococcus* sp. S4

GO:0047733	1	0.295151638420477	CDP-glukóza 4,6-dehydratázová aktivita
GO:0047905	1	0.295151638420477	fruktóza-6-fosfát fosfoketolázová aktivita
GO:0047952	1	0.295151638420477	Glycerol-3-fosfát dehydrogenázová [NAD(P)+] aktivita
GO:0048037	1	0.295151638420477	Vazba kofaktorů
GO:0048307	1	0.295151638420477	Ferredoxin-nitrit reduktázová aktivita
GO:0048472	1	0.295151638420477	Threonin-fosfát dekarboxylázová aktivita
GO:0050193	1	0.295151638420477	fosfoketolázová aktivita
GO:0050278	1	0.295151638420477	Sedoheptulóza-bisfosfatázová aktivita
GO:0050281	1	0.295151638420477	Serin-glyoxylát transaminázová aktivita
GO:0050311	1	0.295151638420477	Sulfit reduktázová (ferredoxin) aktivita
GO:0050380	1	0.295151638420477	Undekaprenyl-difosfatázová aktivita
GO:0050511	1	0.295151638420477	Undekaprenyldifosfo-muramoylpentapeptid beta-N-acetylglukosaminyltransferázová aktivita
GO:0050515	1	0.295151638420477	4-(cytidin 5'-difosfo)-2-C-metyl-D-erytritol kinázová aktivita
GO:0050518	1	0.295151638420477	2-C-metyl-D-erythritol 4-fosfát cytidilyltransferázová aktivita
GO:0050570	1	0.295151638420477	4-hydroxythreonin-4-fosfát dehydrogenázová aktivita
GO:0050577	1	0.295151638420477	GDP-L-fukóza syntázová aktivita
GO:0050620	1	0.295151638420477	Fykocyanobilin:ferredoxin oxidoreduktázová aktivita
GO:0050660	1	0.295151638420477	Vazba flavin adenin dinukleotidu
GO:0051087	1	0.295151638420477	Vazba chaperonů
GO:0051745	1	0.295151638420477	4-hydroxy-3-metylbut-2-en-1-yl difosfát reduktázová aktivita
GO:0051989	1	0.295151638420477	Koproporfyrinogen dehydrogenázová aktivita

Tab. 29: Skupiny s vyšším zastoupením genů u *Synechococca* sp. S4.

Identifikační číslo	<i>Synechococcus</i> sp. S4	<i>C. gracile</i> PCC 6307	p-value	Skupina
GO:0004179	4	2	0.530003636490779	Alanyl aminopeptidázová aktivita
GO:0008270	4	3	0.813884583866962	Vazba iontů zinku
GO:0051276	4	3	0.813884583866962	Organizace chromozómu
GO:0005488	3	2	0.842316660653026	Vazba
GO:0009001	3	2	0.842316660653026	Serin O-acetyltransferázová aktivita
GO:0009236	3	2	0.842316660653026	Biosyntéza kobalaminu
GO:0016787	3	2	0.842316660653026	Hydrolázová aktivita
GO:0001882	2	1	0.87755795074152	Vazba nukleosidů

Tab. 29: Skupiny s vyšším zastoupením genů u *Synechococca* sp. S4.

GO:0003994	2	1	0.87755795074152	Akonitát hydratázová aktivita
GO:0006814	2	1	0.87755795074152	Transport sodných iontů
GO:0008939	2	1	0.87755795074152	Nikotinát-nukleotid-dimethylbenzimidazol fosforibosyltransferázová aktivita
GO:0015093	2	1	0.87755795074152	Transmembránový transportér iontů železa
GO:0015684	2	1	0.87755795074152	Transport železitých iontů
GO:0003864	1	0	0.929133530942721	3-metyl-2-oxobutanoát hydroxymetyltransferázová aktivita
GO:0004372	1	0	0.929133530942721	Glycin hydroxymetyltransferázová aktivita
GO:0006885	1	0	0.929133530942721	Regulace pH
GO:0009019	1	0	0.929133530942721	tRNA (guanin-N1-)-metyltransferázová aktivita

Tab. 30: Skupiny s vyšším zastoupením genů u *C. gracile* PCC 6307.

Identifikační číslo	<i>Synechococcus</i> sp. S4	<i>C. gracile</i> PCC 6307	p-value	Skupina
GO:0003735	46	47	0.512812158263257	Strukturní složka ribozómu
GO:0004355	0	2	0.559064352627293	Glutamát syntázová (NADPH) aktivita
GO:0016034	0	2	0.559064352627293	Maleylacetoacetát izomerázová aktivita
GO:0005524	39	41	0.644264392871147	Vazba ATP
GO:0006412	52	56	0.656204216105998	Translace
GO:0003743	4	7	0.755992615545138	Translační iniciační faktor
GO:0003978	4	7	0.755992615545138	UDP-glukóza 4-epimerázová aktivita
GO:0006413	4	7	0.755992615545138	Iniciace translace
GO:0004035	2	4	0.847641946068765	Alkalicko fosfatázová aktivita
GO:0005506	17	18	0.851461764098669	Vazba iontů železa
GO:0016491	6	9	0.86167837024102	Oxidoreduktázová aktivita
GO:0003824	15	16	0.89273168270244	Katalytická aktivita
GO:0005215	8	11	0.941701630497396	Transportérová aktivita
GO:0006810	13	17	0.949794600048318	Transport
GO:0003848	0	1	0.999999999999994	2-amino-4-hydroxy-6-hydroxymetyldihydropteridin difosfokinázová aktivita
GO:0004085	0	1	0.999999999999994	Butyryl-CoA dehydrogenázová aktivita
GO:0004328	0	1	0.999999999999994	Foramidázová aktivita

Tab. 30: Skupiny s vyšším zastoupením genů u *C. gracile* PCC 6307.

GO:0004540	0	1	0.999999999999994	Ribonukleázová aktivita
GO:0005315	0	1	0.999999999999994	Transmembránový přenos anorganického fosfátu
GO:0005507	0	1	0.999999999999994	Vazba iontů mědi
GO:0006811	0	1	0.999999999999994	Transport iontů
GO:0006817	0	1	0.999999999999994	Transport fosfatových iontů
GO:0008444	0	1	0.999999999999994	CDP-diacylglycerol-glycerol-3-fosfát 3-fosfatidyltransferázová aktivita
GO:0008892	0	1	0.999999999999994	Guanin deaminázová aktivita
GO:0009030	0	1	0.999999999999994	Thiamin-fosfát kinázová aktivita
GO:0010181	0	1	0.999999999999994	Vazba FMN
GO:0015075	0	1	0.999999999999994	Transmembránový transportér iontů
GO:0018489	0	1	0.999999999999994	Vanilát monoxygenázová aktivita
GO:0030272	0	1	0.999999999999994	5-formyltetrahydrofolát cyklo-ligázová aktivita
GO:0042773	0	1	0.999999999999994	Transport elektronů spojený se syntézou ATP
GO:0047753	0	1	0.999999999999994	Cholin sulfatázová aktivita
GO:0003913	1	2	0.999999999999995	DNA fotolyázová aktivita
GO:0004022	1	2	0.999999999999995	Alkohol dehydrogenázová (NAD) aktivita
GO:0004065	1	2	0.999999999999995	Arylsulfatázová aktivita
GO:0004096	1	2	0.999999999999995	Katalázová aktivita
GO:0004140	1	2	0.999999999999995	Defosfo-CoA kinázová aktivita
GO:0004412	1	2	0.999999999999995	Homoserin dehydrogenázová aktivita
GO:0007155	1	2	0.999999999999995	Adheze buňky
GO:0008565	1	2	0.999999999999995	Transport proteinů
GO:0008830	1	2	0.999999999999995	dTDP-4-dehydroramnóza 3,5-epimerázová aktivita
GO:0016740	1	2	0.999999999999995	Transferázová aktivita
GO:0030151	1	2	0.999999999999995	Vazba iontů Molybdenu
GO:0046872	1	2	0.999999999999995	Vazba kovových iontů
GO:0047280	1	2	0.999999999999995	Nikotinamid fosforibosyltransferázová aktivita
GO:0047588	1	2	0.999999999999995	5-aminopentanamidázová aktivita
GO:0003954	3	4	0.999999999999995	NADH dehydrogenázová aktivita
GO:0004721	3	4	0.999999999999995	Fosfoprotein fosfatázová aktivita

Tab. 30: Skupiny s vyšším zastoupením genů u *C. gracile* PCC 6307.

GO:0005525	14	16	0.999999999999995	Vazba GTP
GO:0009055	14	16	0.999999999999995	Přenos elektronů
GO:0031071	3	4	0.999999999999995	Cystein desulfurázová aktivita
GO:0004008	5	6	0.999999999999996	Cu-exportní ATPázová aktivita
GO:0016887	13	16	0.999999999999997	ATPázová aktivita
GO:0008080	7	8	0.999999999999997	N-acetyltransferázová aktivita
GO:0005554	14	17	0.999999999999998	Molekulární funkce
GO:0004016	6	7	0.999999999999999	Adenylát cyklázová aktivita
GO:0006281	6	7	0.999999999999999	Reparace DNA
GO:0008124	2	3	0.999999999999999	4-alfa-hydroxytetrahydrobiopterin dehydratázová aktivita
GO:0008556	2	3	0.999999999999999	Draslík transportující ATPázová aktivita
GO:0009306	2	3	0.999999999999999	Sekrece proteinů
GO:0006118	15	17	1	Transport elektronů
GO:0008152	4	5	1	Metabolické procesy