

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra biochemie



**Fylogenetická analýza sekvencí vybraných rodin
aldehyddehydrogenas z nižších rostlin**

BAKALÁŘSKÁ PRÁCE

Autor:	Šimon Pavlů
Studijní program:	B1406 Biochemie
Studijní obor:	Bioinformatika
Forma studia:	Prezenční
Vedoucí práce:	Mgr. David Kopečný, Ph.D.
Rok:	2019

Prohlašuji, že jsem bakalářskou práci vypracoval/a samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním bakalářské práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl/a jsem seznámen/a s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne *podpis bakaláře*

Poděkování:

Chtěl bych poděkovat vedoucímu své bakalářské práce Mgr. Davidu Kopečnému PhD., za odborné vedení, čas, ochotu a jeho cenné rady během psaní této Bakalářské práce.

Bibliografická identifikace

Jméno a příjmení autora	Šimon Pavlů
Název práce	Fylogenetická analýza sekvencí z vybraných rodin aldehyddehydrogenas z nižších rostlin
Typ práce	Bakalářská
Pracoviště	Katedra biochemie
Vedoucí práce	Mgr. David Kopečný, Ph.D.
Rok obhajoby práce	2019

Abstrakt

Tato bakalářská práce se zabývá analýzou sekvencí aldehyddehydrogenas (ALDH) v nižších rostlinách – řasách a kapradinách. Teoretická část je zaměřena na rešerši současných poznatků o nadrodině rostlinných aldehyddehydrogenas. Výsledková část zahrnuje vyhledávání pomocí softwaru BLAST a identifikaci *ALDH* genů v genomech řas *Klebsormidium nitens*, *Coccomyxa subellipsoidea* a *Chromochloris zofingiensis* a genomu kapradiny *Azolla filiculoides*. Sekvence kódovaných proteinů byly rozřazeny dle schválené klasifikace do rodin a podrodin a byla provedena fylogenetická analýza příbuznosti ALDH z vybraných rodin s členy z vyšších rostlin, pomocí softwaru PhyML. Bylo zjištěno, že většina *ALDH* genů je v genomech řas přítomna v jedné kopii a u dvou řas byla zjištěna absence silně konzervovaného genu *ALDH7*. Dále byla zjištěna přítomnost *ALDH* genu, o kterém se předpokládalo, že je přítomen pouze ve vyšších rostlinách.

Klíčová slova	aldehyddehydrogenasa, BLAST, PhyML, metabolismus aldehydů, řasa, kapradina, fylogeneze
Počet stran	67
Počet příloh	0
Jazyk	Český

Bibliographical identification

Autor's first name and surname	Šimon Pavlů
Title	Phylogenetic analysis of sequences of selected aldehyde dehydrogenase families from lower plants
Type of thesis	Bachelor
Department	Department of biochemistry
Supervisor	Mgr. David Kopečný, Ph.D.
The year of presentation	2019

Abstract

This bachelor thesis deals with the analysis of aldehyde dehydrogenase (ALDH) sequences in lower plants – algae and ferns. The theoretical part is focused on research of current knowledge about plant aldehyde dehydrogenase. The results include BLAST search and identification of ALDH genes in the *Klebsormidium nitens*, *Coccomyxa subellipsoidea* and *Chromochloris zofingiensis* genomes and *Azolla filiculoides* genome. The sequences of the encoded proteins were classified according to the approved classification into families and subfamilies and phylogenetic analysis of ALDH affinity of selected families with members from higher plants was performed using PhyML software. Most of the ALDH genes were found to be present in one copy in algal genomes and the absence of a strongly conserved ALDH7 gene was found in two algae. Furthermore, the presence of the ALDH gene was found, that was supposed to be present only in higher plants.

Keywords	Aldehyde dehydrogenase, aldehyde metabolism, Algae, BLAST, Fern, phylogenesis, PhyML
Number of pages	67
Number of appendices	0
Language	Czech

OBSAH

1 Úvod	1
2 Současný stav řešené problematiky	2
2.1 Mechy	2
2.1.1 Životní cyklus mechů	2
2.1.2 <i>Physcomytrella Patens</i>	3
2.2 Řasy	4
2.2.1 Životní cyklus řas	4
2.2.2 Charophyta a evoluce	5
2.2.2.1 <i>Klebsormidium</i>	7
2.2.2.2 <i>Chlorella</i>	7
2.2.2.3 <i>Coccomyxa</i>	8
2.3 Kapradiny	9
2.3.1 Životní cyklus kapradin	9
2.3.2 <i>Azolla filiculoides</i>	10
2.4 Aldehyddehydrogenasy	11
2.4.1 Nomenklatura Aldehyddehydrogenas	11
2.4.2 Rodina ALDH2	12
2.4.3 Rodina ALDH3	13
2.4.4 Rodina ALDH5	13
2.4.5 Rodina ALDH6	14
2.4.6 Rodina ALDH7	14
2.4.7 Rodina ALDH10	15
2.4.8 Rodina ALDH11	16
2.4.9 Rodina ALDH12	17
2.4.10 Rodina ALDH18	18
2.4.11 Rodina ALDH21	18
2.4.12 Rodina ALDH22	19
2.4.13 Rodina ALDH23 a ALDH24	19
3 Experimentální část	20
3.1 Materiál	20
3.2 Metody	21
3.2.1 BLAST	21
3.2.2.1 Algoritmus BLAST	22
3.2.2.2 Princip algoritmu	22
3.2.2.3 Algoritmus FASTA a dynamické programování	24
3.2.2.4 Program BLAST	26
3.2.2.5 Využití BLASTu	27
3.2.2 MUSCLE	27
3.2.3 GBlocks	31
3.2.4 PhyML	33
4 Výsledky	35
4.1 Přehled nalezených genů	35
4.2 Geny ALDH přítomné v řase <i>Klebsormidium nitens</i>	35
4.3 Geny ALDH přítomné v řase <i>Coccomyxa subellipsoidea</i>	38

4.4 Geny ALDH přítomné v řase <i>Chromochloris zoofingiensis</i>	40
4.5 Geny ALDH přítomné v kapradině <i>Azolla filiculoides</i>	42
4.6 Fylogenetická analýza	44
5 Diskuze	50
6 Závěr	53
7 Literatura	54

CÍLE PRÁCE

- Vypracování literární rešerše na téma nižší rostliny a nadrodiny ALDH v rostlinách na základě doporučené literatury.
- Hledání genů ALDH pomocí webového rozhraní GenBank a Phytozome (a jiných) s cílem zjištění počtu jednotlivých rodin a jednotlivých genů v řasách pro zástupce vývojové větve Charophyta a kapradinách.
- Rozřazení genů do rodin na základě shody se známými zástupci z vyšších rostlin.
- Provedení fylogenetické analýzy pomocí programu PhyML, pro vybrané rodiny genů *ALDH*, konkrétně ALDH2, ALDH5, ALDH7, ALDH 10, ALDH12, ALDH21, ALDH22.

Seznam užitých zkratek

3-PG	3-fosfoglycerát
ABAL	4-aminobutyraldehyd
ACAPAL	3-acetamidopropionaldehyd
ALDH	aldehyddehydrogenasa
AMADH	aminoaldehyddehydrogenasa
APAL	3-aminopropionaldehyd
ATP	Adenosintrifosfát
BADH	betainaldehyddehydrogenasa
BAL	β -alanin
BLAST	Basic Local Alignment Search Tool (algoritmus pro porovnávání sekvencí)
BLOSUM	BLOcks SUBstitution Matrix (typ substituční matice)
CLUSTAL	software pro mnohočetné porovnání sekvencí, či formát souboru sekvencí
CoA	koenzym A
DMSP	dimethylsulfoniopropionát
DMSPAL	3-dimethylsulfoniopropionaldehyd
FASTA	algoritmus pro porovnávání sekvencí nebo též formát souboru sekvencí
GABA	γ -aminobutyrát
GAP	glyceraldehyd-3-fosfát
GAPN	glyceraldehyd-3-fosfátdehydrogenasa
Gb (Mb)	giga páry bazí (mega páry bazí)
GB	glycin betain
GBAL	4-guanidinobutyraldehyd
GSAL	glutamát γ -semialdehyd
GSALDH	glutamát- γ -semialdehyddehydrogenasa
HSP	high-scoring segment pairs (párové segmenty sekvencí s vysokým skóre)
IUPAC	International Union of Pure and Applied Chemistry
MAFFT	multiple alignment using fast Fourier transform (software pro multiple alignment)

MDA	3,4-methylendioxyamfetamin
MMS	methylmalonylsemialdehyd
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NAD ⁺ /NADH	nikotinamidadenindinukleotid
NADP ⁺ /NADPH	nikotinadenindinukleotidfosfát
NCBI	National Center for Biotechnology Information
P5C	prolin-5-karboxylát
P5CS	Δ^1 -pyrrolin-5-karboxylátsynthasa
P6CDH	Δ^1 -piperidin-6-karboxylátdehydrogenasa
PAM	Point Accepted Mutation (typ substituční matice)
PhyML	Phylogenetic estimation using Maximum Likelihood (software pro fylogenetickou analýzu)
SSAL	sukcinát semialdehyd
SSALDH	sukcinátsemialdehyddehydrogenasa
T-coffee	Tree-based Consistency Objective Function for Alignment Evaluation (software pro multiple alignment)
TMABAL	N,N,N-trimethyl-4-aminobutyraldehyd

1 ÚVOD

Nadrodina aldehyddehydrogenas (ALDH) obsahuje mnoho různých enzymů katalyzujících metabolismus aldehydů za pomoci NAD⁺ nebo NADP⁺ jako kofaktorů. V posledních letech se výzkum na naší katedře v Olomouci zaměřuje na studium rostlinných ALDH, a to konkrétně na ALDH2, ALDH7, ALDH10, ALDH12 a ALDH21 (Tylichová et al., 2010; Kopečný et al., 2011; Kopečný et al., 2013; Končítíková et al., 2019).

S rostoucím množstvím sekvenovaných genomů nižších rostlin se nám nabízí možnost studovat *ALDH* geny v těchto rostlinách a společně s nimi i evoluci těchto genů spojenou s expanzí rostlin na souš a například i jejich vliv na přežití rostlin ve stresových podmínkách jako je zvýšená slanost nebo dehydratace. Tato práce se zaměřuje na přítomnost sekvencí ALDH v plně sekvenovaných genomech nižších rostlin jako jsou řasy *Klebsormidium nitens*, *Coccomyxa subellipsoidea* a *Chromochloris zofingiensis* nebo kapradina *Azolla filiculoides*. Výsledky jsou pak porovnávány s nalezenými ALDH rodinami z předešlých studií (Brocker et al., 2013).

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

2.1 Mechy

Mechy řadící se mezi Bryophyta jsou malé výtrusné rostliny, které se typicky vyskytují ve vlhkých a stinných lokacích kde rostou ve shlucích. Toto oddělení nemá ještě plně vyvinuté cévní svazky, a tedy obtížněji hospodaří s vodou, kterou přijímají převážně celým povrchem stélky. Také jsou známé svou značnou substrátovou specificitou, protože jejich přichytná vlákna neumožňují úplné využití živin jako u cévnatých rostlin.

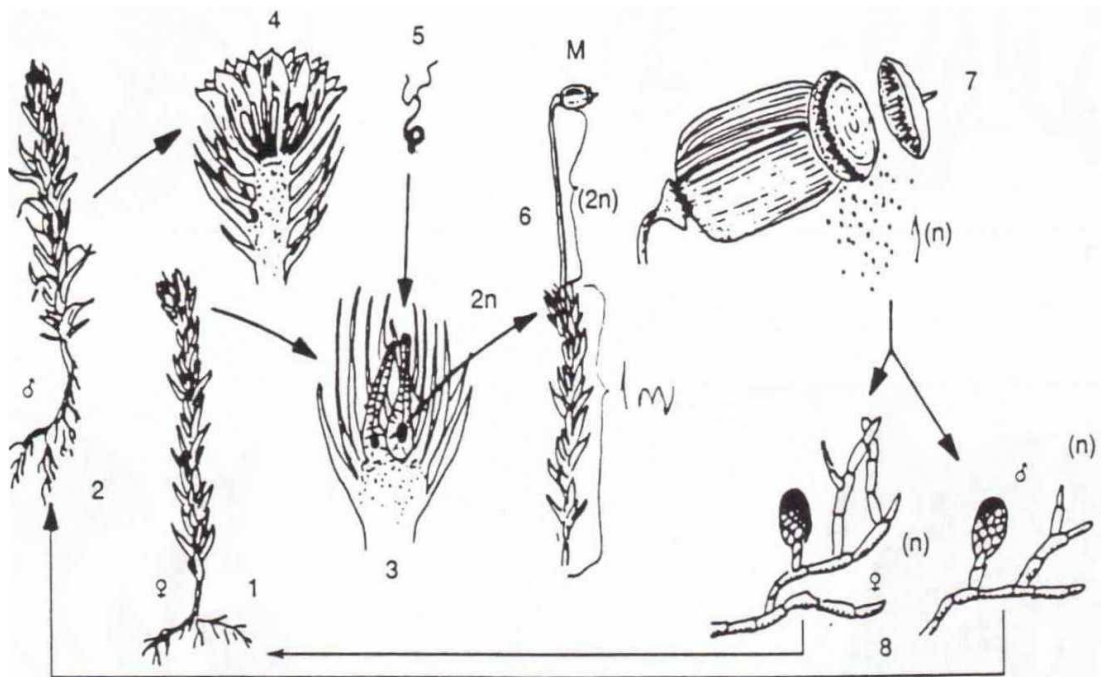
Mechy představují významný mezistupeň v evoluci rostlin, jejich přechodu od vodních rostlin k suchozemským druhům a vzniku pro tyto druhy specifických znaků, rozdělujících tyto dvě skupiny. Mechy mají více než 10 000 druhů a řadí se tedy mezi jedno z nejpočetnějších oddělení vyšších rostlin.

2.1.1 Životní cyklus mechů

Životní cyklus mechů se liší od semenných rostlin. Střídají se zde dvě fáze, kdy haploidní gametofyt, který je schopen fotosyntézy, je právě tou dominantní nad diploidním sporofytem, stejně jako u jätrovek nebo hlevíků. Gametofyt pozorujeme od vyklíčení haploidní spory, z níž vzniká rozvětvený vláknitý prvoklíček. Ten obsahuje vláknitá chloronemata, buňky, v nichž jsou obsaženy chloroplasty, tvořené vlákny s příčnými přehrádkami. Chloronema dále diferencují na rhizoidy. Prvoklíček obsahuje také kaulonemata, která obsahují značně méně chloroplastů. Z prvoklíčku poté vyrůstají pupeny, které jsou základem pro jednotlivé gametofory. Ty už se skládají z mnohobuněčných přichytných vláken (rhizoidy), kterými se liší například od jätrovek, dále také z lodyžky (kauloid) a lístků (fyloidy). Na vrcholu stonků gametoforu se poté vytvářejí pohlavní orgány, samičí zárodečníky nebo samčí pelatky. Mechy mohou být jednodomé, kdy se na jedné rostlince vyskytují oba typy pohlavních orgánů nebo dvoudomé, tedy obsahující pouze samčí nebo samičí orgány na rozdílných rostlinkách. Ve vodním prostředí se spermatozoidy z pelatek pomocí bičků dostávají k samičímu vajíčku v zárodečníku a dochází k oplodnění a následnému vzniku diploidního sporofytu. Sporofyt obsahuje štět s tobolek, kde produkují haploidní výtrusy, které jsou následně roznášeny větrem.

Některé mechy, jako třeba Čtyřzoubek průzračný (*Tetraphis pellucida*), mají možnost i nepohlavního rozmnožování, a to pomocí struktur nazývaných gemy. Ty se

mohou vytvářet po celé ploše mechové stélky a když se od rostliny oddělují, tvoří totožný klon.



Obr. 1 Metagenese mechů – heteromorfní rodokmen. Jednoznačně převažuje haploidní gametofyt. Znárodně je dvoudomý prokel (1, 2). Na něm vyrůstají zárodečníky (3) nebo pelatky (4) produkující spermatozoidy (5). Vzniká diploidní štět s tobolkou (6), z nějž se pak uvolňují spory tvořící prvoklíčky (8). Upraveno dle Jelínek a Zicháček (2011).

2.1.2 *Physcomyrella patens*

Physcomyrella Patens je zelený mech, jehož genom byl plně sekvenován roku 2008 (Rensing et al., 2008). Tato rostlina je často používána jako modelový organismus pro evoluční studie, společně se zelenými řasami *Chlamydomonas reinhardtii* nebo semennou rostlinou *Arabidopsis thaliana*, jelikož mechy tvoří mezistupeň při přechodu od jednobuněčných fotosyntetizujících organismů k cévnatým suchozemským rostlinám (Strobek et al., 2013). Porovnáváním genové funkce v jednotlivých zástupcích nám pak může posloužit k rozeznání konzervovaných genů v průběhu evoluce, ale také k odhalení nových genů vzniklých adaptací. *P. patens* je také významná pro reverzní genetiku, která může být užita ke zjištění funkce genů nebo molekulární evoluce rostlin, díky svým výhodným homologním rekombinacím (Kammerer and Cove, 1996).



Obr. 2 *Physcomitrella patens*, její gametofor (vlevo) a plně vyvinutý sporofyt (vpravo, vyznačen šipkou). Upraveno dle Strobek et al. (2013).

2.2 Řasy

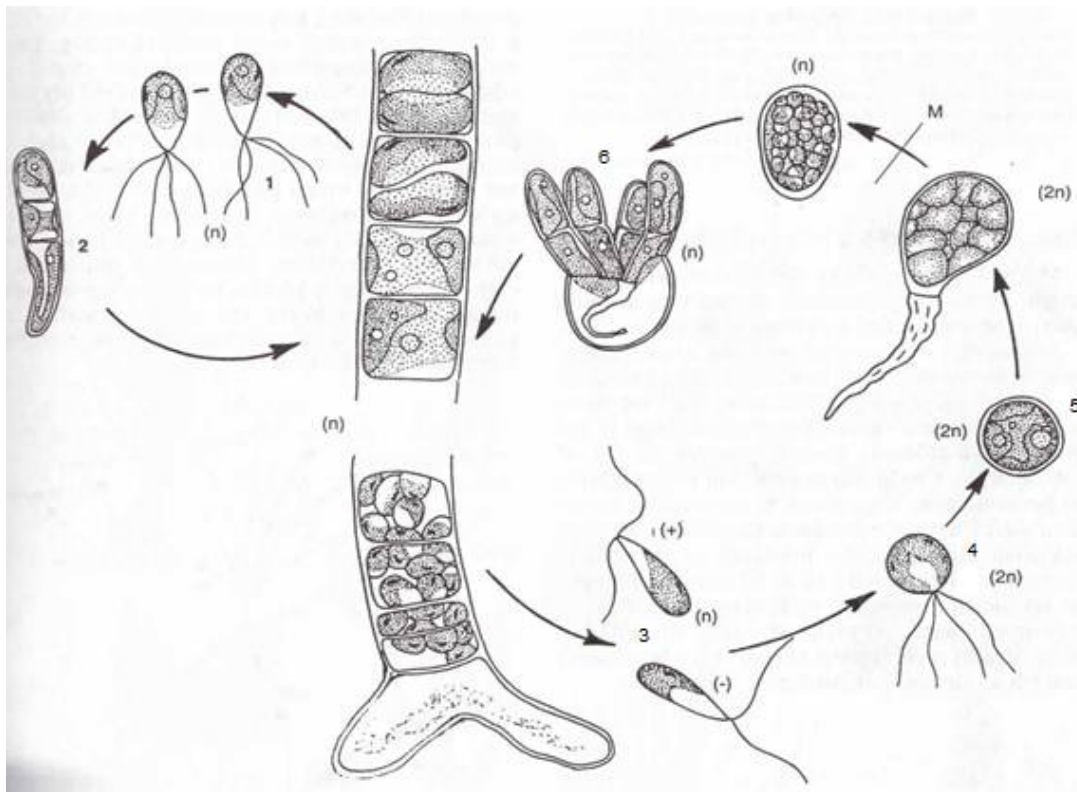
Jako řasy je označována skupina fotosyntetizujících organismů, které nemusí být navzájem příbuzné. Jejich zástupci zahrnují organismy od jednoduchých jednobuněčných řas, jako například *Chlorella*, až po mnohobuněčné řasy, jejichž tělo tvoří stélka. Jedná se převážně o vodní organismy, které neobsahují buněčná pletiva typické pro suchozemské rostliny jako je floém nebo xylém, stejně jako rostlinné orgány jako kořeny či listy. Neobsahují ani rhizoidy typické pro mechy. Většina řas je autotrofní, díky své schopnosti fotosyntézy, kterou získávají vznikem chloroplastů primární endosymbiózou se sinicí či sekundární symbiózou s jinou řasou. Některé zelené řasy, ale také mnohé zlativky a jiné řasy zvolily heterotrofní způsob výživy, nebo se dokonce jedná o parazity, kdy se spoléhají na fotosyntézu pro získání živin velmi omezeně nebo vůbec. (Figueroa-Martinez et al., 2015)

Skupina řas pojímá několik říší, rostliny *Archaeplastida*, *Excavata* a *Chromista*. Nejrozsáhlejší říší jsou *Archaeplastida*, které obsahují nejpočetnější oddělení *Chlorophyta* (zelené řasy) a *Rhodophyta* (červené řasy). Mezi nejkomplexnější sladkovodní řasy se řadí *Charophyta*, oddělení zelených řas, sesterská skupina *Chlorophyt*, která obsahuje například rody *Klebsormidium*, *Chlorella* nebo *Coccomyxa*.

2.2.1 Životní cyklus řas

Životní cyklus řas se značně liší. Obecně jej ale můžeme rozdělit na pohlavní a nepohlavní fázi rozmnožování. Během nepohlavní fáze se jedinec množí dělením

buněk, pomocí zoospor či aplanospor, z nichž následně vyrůstá nové vlákno řasy. Během pohlavní fáze se jedná o jedince haploidního, jež tvoří samčí nebo samičí gamety, které spolu splynou a vytvoří diploidní zygotu. Pro zaručení úspěšné pohlavní reprodukce je tvorba gamet regulována, v tomto procesu nejspíše hrají důležitou roli feromony (Frenkel et al., 2014). Pohlavní rozmnožování poskytuje vyšší variabilitu, na druhou stranu nepohlavní rozmnožování je výhodnější z hlediska efektivity. *Charophyta* nemusí obsahovat pohyblivé samčí pohlavní buňky, jako třeba spájivky (*Zygnematophyceae*), avšak u některých z nich se vyskytují, například u parožnatek.



Obr. 3 Metageneze řas. Nepohlavní rozmnožování řas znázorněné pomocí zoospor (1), z nichž pak vzniká nové vlákno řasy (2). Pohlavní rozmnožování zajištěné gametami (3), ty se spojí a vzniká planozygota (4) poté zygospora (5). Nakonec vzniká po meiotickém dělení sporofyt s haploidními vlákny (6). Upraveno dle Jelínek a Zicháček (2011).

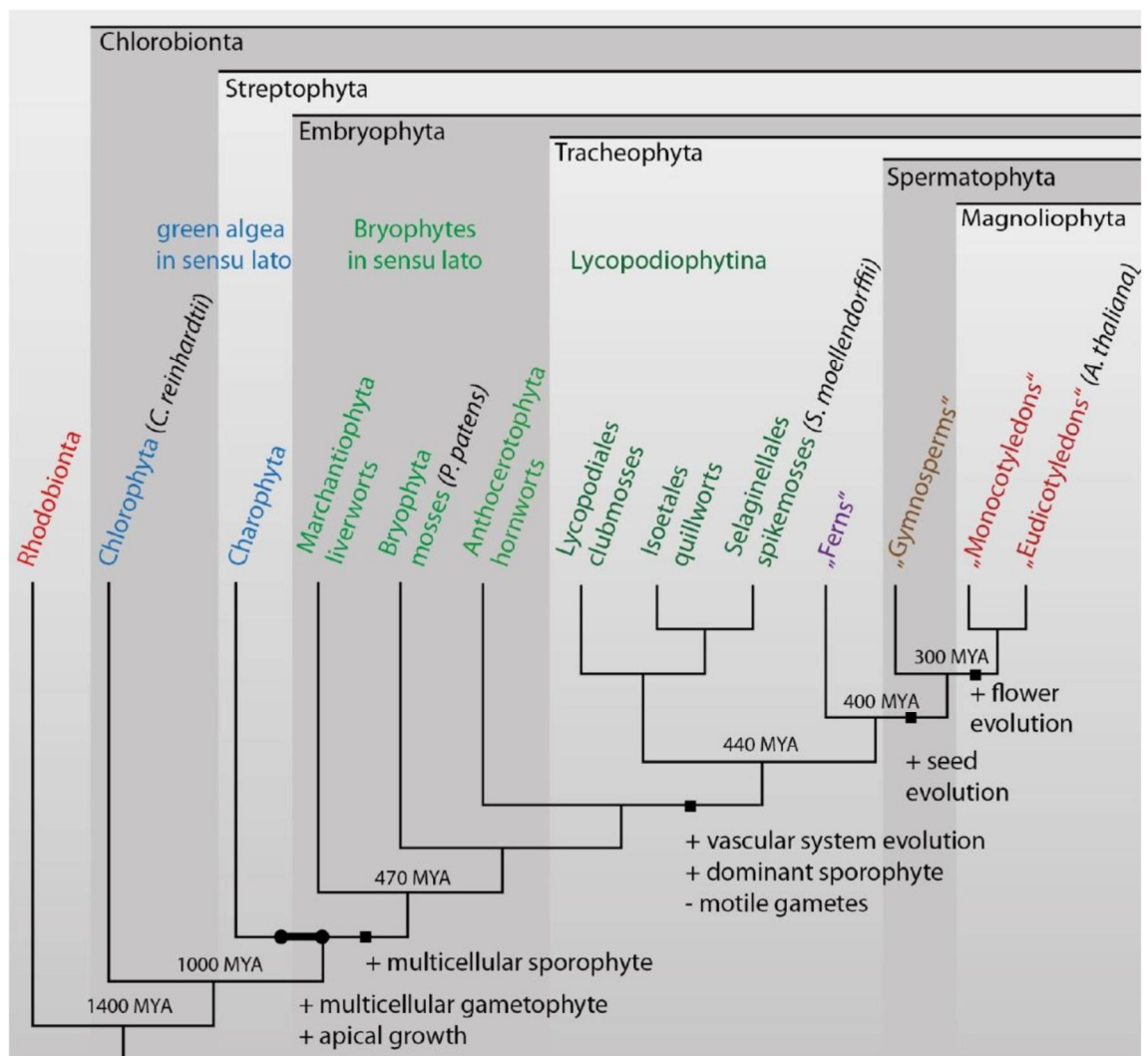
2.2.2 *Charophyta* a evoluce

Charophyta jsou skupinou sladkovodních zelených řas, o kterých se předpokládá, že suchozemské rostliny, přesněji *Embryophyta*, vzešly z této rodiny.

Před 500 miliony let se u *Charophyt* vyvinuly některé znaky vyšších rostlin, jako mnohobuněčný gametofyt, jehož buňky již byly ohraničeny buněčnou stěnou a propojeny plazmodezmaty. U *Embryophyt* se dále vyvinul diploidní sporofyt před 470

miliony lety. Před 400 miliony let se pak vyvinula první semena. Krytosemenné rostliny (*Magnoliophyta*) se vyvinuly o 100 milionů let později (Floyd and Bowman, 2007).

Charophyta jsou sesterskou skupinou s *Chlorophyta*. Společně s *Embryophyta* se řadí do skupiny *Viridiplantae* (zelené rostliny). Jejich společnými znaky jsou například nepřítomnost fykobilinů, dále přítomnost chlorofylu a, chlorofylu b, přítomnost celulózy v buněčných stěnách nebo také využití škrobu jako zásobní látky. S *Embryophyta* však sdílí také určité znaky, kterými se odlišují od zástupců *Chlorophyta* a to přítomností určitých enzymů, jako Cu/Zn superoxiddismutasa a glykolát oxidasa, nebo užití fragmoplastu při buněčném dělení (Leliaert et al., 2012). *Charophyta* a *Embryophyta* jsou tedy řazeny z evolučního hlediska do jedné skupiny (clade) a to *Streptophyta*.

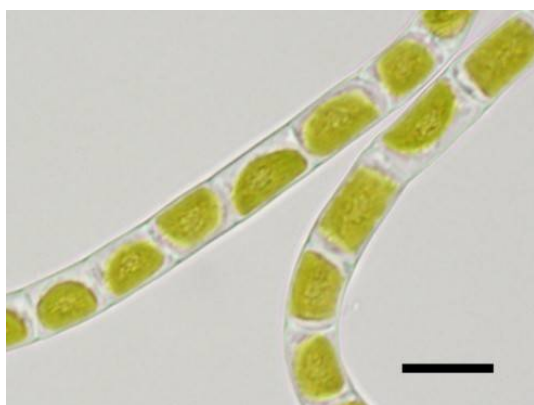


Obr. 4 Evoluční strom rostlin. Fylogenetický strom znázorňující evoluční vztah mezi jednotlivými odděleními od červených řas (*Rhodobionta*) až po krytosemenné rostliny

(*Magnobiontha*). Vyznačené jsou modelové organismy významné pro dané evoluční větve (Strobek et al., 2013).

2.2.2.1 Klebsormidium

Rod *Klebsormidium* obsahuje vláknité zelené řasy náležící do oddělení *Charophyta*. Existuje 20 známých zástupců tohoto rodu. Jejich pojmenování bylo zavedeno roku 1972. Ačkoliv se většina zástupců vyskytuje na souši, často ve vlhkých substrátech, existují také čistě vodní zástupci. Mnoho rostlin rodu *Klebsormidium* syntetizuje MMA aminokyseliny (mycosporine-like amino acids), které mají významnou funkci při ochraně organismu proti UV záření. Zástupci tohoto rodu se rozmnožují převážně fragmentací nebo pomocí aplanospor. Pohlavně se rozmnožují pomocí izogamie, což znamená, že splývající gamety jsou stejného tvaru a velikosti (izogamety) (Fritsch et al., 1935). Sekvenovaný genom zástupce tohoto rodu, *Klebsormidium nitens*, byl publikován roku 2014.



Obr. 5 *Klebsormidium nitens* – obrázek získán mikroskopem při 1000 x zvětšení. Černá linka představuje měřítko (10 μ m). Převzato ze stránky:

http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/

2.2.2.2 Chlorella

Chlorella je rod řadící se mezi sladkovodní jednobuněčné řasy, popsány roku 1890 mikrobiologem M. W. Beijerinck. Vzhledem k ostatním rostlinám obsahují zástupci tohoto rodu velké množství chlorofylu a také proteinů. *Chlorella* je tedy také vhodná pro redukci oxidu uhličitého a produkci kyslíku v uzavřeném prostoru jako například vesmírné stanice. V ruském experimentu roku 1965 byla pěstována v barelech pod umělým světlem (<https://www.permanent.com/russian-celss.html>). Druhy chlorelly můžeme od sebe rozlišit jednoduše sekvencováním DNA. Mezi hlavní druhy patří *Chlorella vulgaris*, *Chlorella pyrenoidosa*, *Chlorella lobophora*, *Parachlorella kessleri*,

Heterochlorella luteoviridis, *Chloroidium ellipsoideum*, *Mychonastes homosphaera* nebo *Chromochloris zofingiensis* (*Chlorella zofingiensis*).

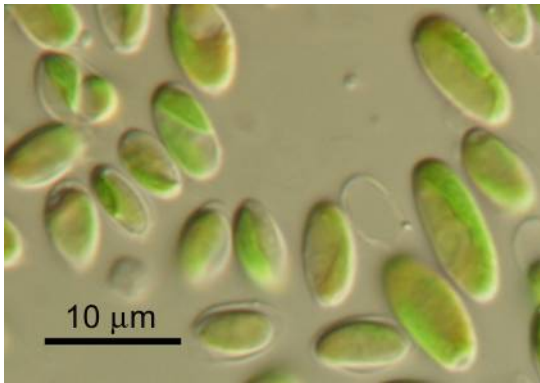


Obr. 6 *Chromochloris zofingiensis*. Převzato z webu: <https://botany.natur.cuni.cz/algo/database/node/1207>

2.2.2.3 *Coccomyxa*

Coccomyxa je rod zelených řas řadící se do rodiny *Coccomyxaceae*. Jsou typické svým elipsovým tvarem těla a tím, že obsahují jednoduchý chloroplast, ten však neobsahuje pyrenoidy. Jsou to bílkovinná tělíska obsahující RUBISCO vážící oxid uhličitý v temnostní fázi fotosyntézy. Na povrchu pyrenoidu je většinou škrob.

Zástupci tohoto rodu mohou mít rozličný styl života, ať už parazitický, kdy parazitují na měkkýších jako například slávky, symbiotický, a to například se stopkovýtrusnými či vřeckovýtrusnými houbami anebo volně žijící (Darienko et al., 2015). *Coccomyxa* jsou po většinu svého života haploidní a rozmnožují se převážně nepohlavně (Grube et al., 2017). *Coccomyxa* je ideální jako modelový organismus a také je ideální pro experimenty, jelikož byl publikován celý její genom v podobě *Coccomyxa subelipsoidea* (Blanc et al., 2012, Darienko et al., 2015).



Obr. 7 *Coccomyxa subellipsoidea* (C-169)

(The Microbial Culture Collection at NIES, Tsukuba, Japan, via NCBI)

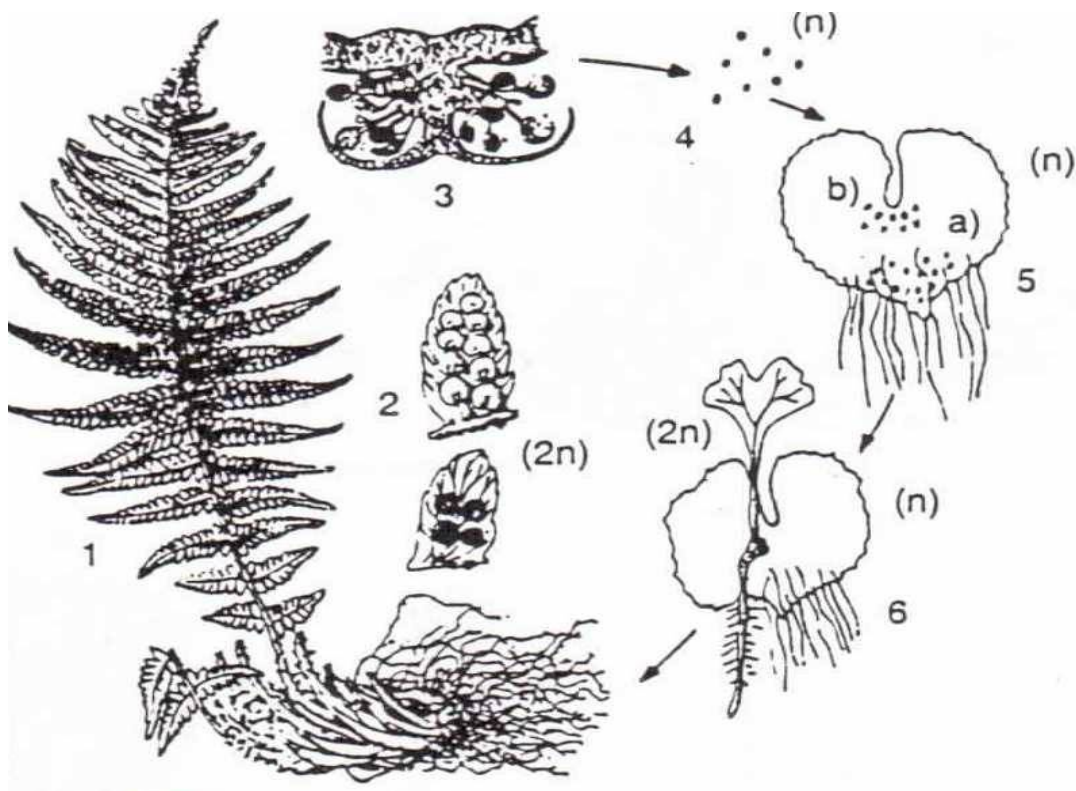
2.3 Kapradiny

Kapradiny jsou podtřídou kaprad'orostů patřící mezi výtrusné rostliny. Jedná se o cévnaté rostliny, to značí, že obsahují cévní svazky, tkáně uzpůsobené pro rozvádění živin po těle rostliny. Kapradiny mají na rozdíl od mechů složité listy zvané megafyly a nemají schopnost druhotného tloustnutí čili tloustnutí stonku nebo kořenu zapříčiněné činností felogénu nebo laterálního meristému kambia. Nedávno byl identifikován gen, specifický pro kapradiny, který zaručuje jejich odolnost vůči hmyzu, tento gen byl nejspíše získán horizontálním přenosem genové informace od bakterií (Li et al., 2018). Oddělení kapradin obsahuje jednu hlavní třídu *Polypodiopsida*, která obsahuje přes 10 000 známých zástupců (Christenhusz et al., 2016). Většina současných druhů jsou byliny avšak, když se kapradiny objevily poprvé ve spodním devonu, asi před 360 miliony lety, jednalo se o stromovité druhy. Současné druhy se vyvinuly až před 145 miliony lety, během období kříd, společně s kvetoucími rostlinami. Evolučním předkem kapradin jsou nejspíše *Rhyniophyta*.

2.3.1 Životní cyklus kapradin

V životním cyklu kapradin, na rozdíl od mechů, převládá sporofyt, podobně jako u semenných rostlin. Z haploidní spory se vyvine fotosyntetizující prokel. Ten představuje pohlavní generaci neboli gametofyt, který obsahuje rozmnožovací útvary (gametangia) a produkuje haploidní pohlavní buňky (gamety). Archegonia tvoří vajíčka a Antheridia tvoří spermatozoidy, nejčastěji se vyskytují na jednom organismu samčí i samičí gametangia. Samčí pohyblivé buňky následně oplodní vajíčko, které je připevněno k prokelu. Vzniká diploidní zygota, ze které následně vyklíčí diploidní

sporofyt, ten je většinou tvořen velikými listy vyrůstajícími z podzemního stonku (oddenku). Na spodní straně listů sporofytu se nacházejí výtrusnice (sporangia). Často se nacházejí ve shlucích chráněných ostěrou. Ve sporangiích vznikají meiózou haploidní spory. Za sucha se uvolňuje voda, ostěra se roztrhne a uvolní spory, proces se opakuje.



Obr. 8 Metagenese Kapradin. Z oddenku v zemi vyrůstají velké složené listy (1), nesoucí na rubu kupky výtrusnic (2, 3). Z uvolněných spor (4) vyklíčí v tomto případě jednodomý prokel (5). Spojení gamet pak vytvoří znovu sporofyt (6). Převzato z Jelínek a Zicháček, 2011).

2.3.2 *Azolla filiculoides*

Rod *Azolla* jsou malé vodní kapradiny z čeledi nepukalkovitých. Rostliny tohoto rodu často rostou ve shlucích na vodní hladině a připomínají mechové rostlinky, na spodní straně mají splývavé kořeny. *Azolla* žije v symbióze se sinicí *Anabaena azollae*, která je ukrytá v listech kapradiny (Lhotský, 2015). Sinice je schopná fixovat dusík, který získává ze vzduchu a vytváří amonné ionty pomocí specializovaných buněk heterocyst. Tyto ionty už může využít samotná kapradina. Symbióza je tedy výhodná pro oba organismy, kdy sinice získává vhodné místo pro svůj vývoj a kapradina dostává pravidelný přísun dusíku. *Azolla* je jednou z nejrychleji rostoucích kapradin a je tedy významná díky své schopnosti pohlcovat znečišťující látky z ovzduší. Genom Azolly, o velikosti pouze 0,75 Giga párů bazí (Gb) (Obermayer et al., 2002), je relativně velmi

malý v porovnání s ostatními kapradinami, známými pro svůj mohutný genom, jako například až 148Gb (Hidalgo et al., 2017). Nejmenší zaznamenaný genom se vyskytuje u sesterského rodu *Azolla*, u zástupce *Salvinia cucullata*, který má velikost pouze 0,26Gb (Li et al., 2018).



Obr. 9 *Azolla filiculoides*. Převzato z Li et al. (2018.)

2.4 Aldehyddehydrogenasy

Nadrodina aldehyddehydrogenas (ALDH) je rozmanitá skupina prokaryotických i eukaryotických enzymů, které se podílejí na metabolismu aldehydů za pomoci NAD^+ nebo NADP^+ jako kofaktorů. V rostlinách se vyskytuje právě 14 různých rodin těchto enzymů, a to: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH19, ALDH21, ALDH22, ALDH23 a ALDH24. Zastoupení rodin ALDH se liší u každé rostliny, přičemž převažuje trend, že u komplexnějších rostlin se vyskytuje větší množství genů. Některé rodiny, specificky ALDH10, ALDH12, ALDH19, ALDH21, ALDH22, ALDH23 a ALDH24, jsou typické pouze pro rostliny. Ostatní rodiny pak mají známé živočišné orthology. Předpokládá se, že rodiny ALDH21, ALDH23 a ALDH24 jsou specifické pro nižší rostliny, naopak rodina ALDH22 se převážně vyskytuje až u vyšších rostlinách (Brocker et al., 2013).

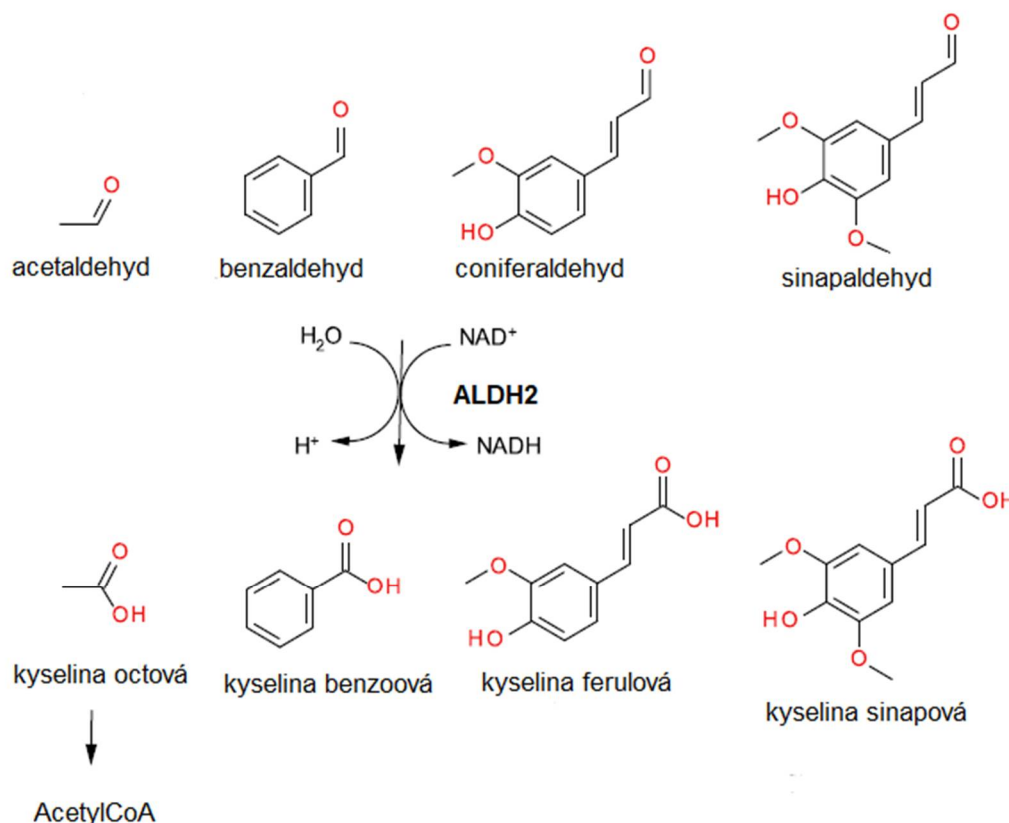
2.4.1 Nomenklatura Aldehyddehydrogenas

Nomenklatura neboli systém zařazení a pojmenování, nadrodiny ALDH vznikla v roce 1999 (Vasiliou et al. 1999) a řídí se pravidly Human Gene Nomenclature. Všechny sekvence jsou pojmenovány stejným základem „ALDH“, následovaným číslicí označující příslušnost do jednotlivých rodin. Toto označení je pak následováno písmenem abecedy, značícím podrodinu a případně dalším číslem označujícím jednotlivé geny v dané podrodině. Rozřazení ALDH se řídí pomocí párového porovnání

sekvencí aminokyselin. Sekvence, které mají vzájemnou podobnost 40 % a více, řadíme do stejné rodiny. Pokud je podobnost sekvencí vyšší než 60 % řadí se také do stejné podrodiny (Brocker et al., 2013).

2.4.2 Rodina ALDH2

Jedná se o skupinu mitochondriálních a cytosolických enzymů s poměrně vysokou substrátovou specificitou. Tyto enzymy jsou aktivní jako homotetramery (Marchitti et al. 2008). ALDH2 v rostlinách metabolizují acetaldehyd vytvářený při fermentaci ethanolu. Následně se touto reakcí vytváří acetát, který se využívá při syntéze koenzymu A pomocí acetyl-CoAsynthasy, tento proces také nazýváme „pyruvátdehydrogenasovou drahou“ (op den Camp and Kuhlemeier 1997; Wei et al. 2009). V *Arabidopsis* k této reakci přispívá hlavně ALDH2B4 (Wei et al., 2009). Další studie také naznačují, že gen ALDH2C4 má vliv na biosyntézu kyseliny ferulové a sinapové, které zvyšují pevnost buněčných stěn (Grabber et al. 2000; Nair et al. 2004).



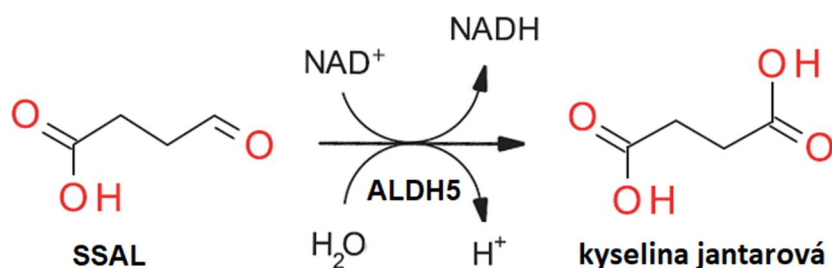
Obr. 10 Reakce katalyzované pomocí ALDH2. Acetaldehyd se přeměňuje na kyselinu octovou, ta se pak využívá k přípravě AcetylCoA, benzaldehyd na benzoovou kyselinu, koniferaldehyd na kyselinu ferulovou, sinapaldehyd na kyselinu sinapovou. Upraveno dle Koncitikova et al. (2015).

2.4.3 Rodina ALDH3

Tato rodina enzymů se může v rostlinách vyskytovat v rozličných organelách, například v cytosolu, chloroplastu nebo v mitochondriích (Kirch et al. 2004; Jimenez-Lopez et al. 2010; Stiti et al. 2011). Exprese většiny genů z této rodiny je regulována kyselinou abscisovou, reagující na stresové podmínky. Tuto rodinu můžeme dále rozdělit na šest podrodin ALDH3E, 3F, 3H, 3I, 3J a 3K. V *Arabidopsis thaliana* se vyskytují tři podrodiny (3I, 3H a 3F), kdy můžeme vidět, že ALDH3I1 například se vyskytuje převážně v listech a svou expresí odpovídá na stresové podmínky jako dehydratace, pesticidy, těžké kovy, nebo zvýšená salinita, stejně jako ALDH3H1 (Kirch et al. 2001; Stiti et al. 2011; Sunkar et al. 2003). Avšak ALDH3F1 nereaguje na stresové podmínky vůbec jako zbylé dvě podrodiny. Na základě těchto zjištění v *Arabidopsis thaliana* bylo vydedukováno, že se tato rodina vyvinula následkem specializace určitých tkání a organel (Kirch et al. 2004).

2.4.4 Rodina ALDH5

Rodina ALDH5 představuje sukcinátsemialdehyddehydrogenasy (SSALDH), které katalyzují přeměnu sukcinát semialdehydu (SSAL) na sukcinát během katabolismu γ -aminobutyátu (GABA). U savců hraje GABA důležitou roli jako neurotransmitter. V rostlinách pomáhá při opylování, odpuzování býložravců nebo při oxidativním stresu (Fait et al. 2008; Palanivelu et al. 2003). Pro tuto rodinu se předpokládá, že se jedná převážně o mitochondriální enzymy (Bouche et al. 2003; Gao and Han 2009). Mutace ALDH5 pak mohou způsobit nadprodukcii meziproductů reaktivního kyslíku a buněčnou smrt následkem světelného či tepelného stresu (Bouche et al. 2003).



Obr. 11 Reakce sukcinátsemialdehyddehydrogenasy (ALDH5), využívající NAD^+ jako koenzym, při níž vzniká kyselina jantarová. Upraveno dle Kopečná et al. (2015).

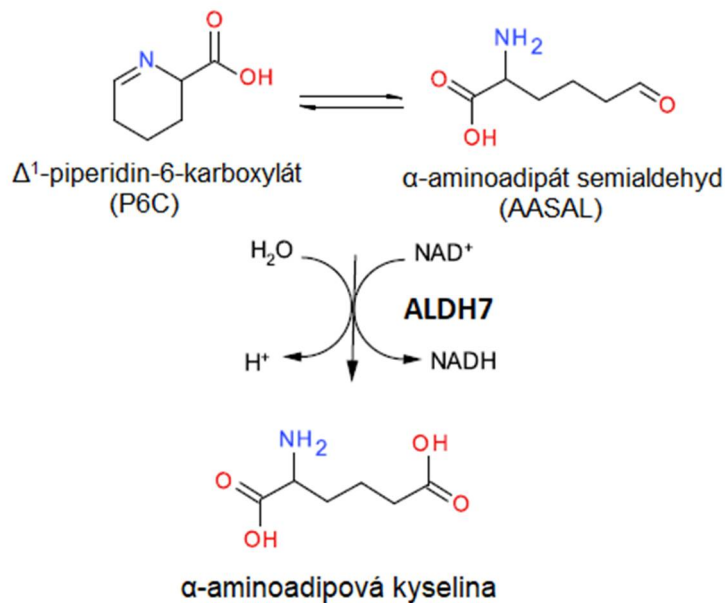
2.4.5 Rodina ALDH6

Rodina ALDH6 je skupina enzymů, také zvaných methylmalonylsemialdehyddehydrogenasy. Usnadňují reakce probíhající během katabolismu valinu a pyrimidinu. Je to jediná rodina ALDH, která při svých reakcích užívá koenzym A jako kofaktor, jelikož katalyzují koenzym A-dependentní reakci přeměny MMS na propionyl CoA. Mimo to jsou enzymy ALDH6 schopné také metabolizovat malonátsemialdehyd na acetyl-CoA (Marchitti et al. 2008). Je však zapotřebí více studií, jelikož přesná funkce této rodiny stále ještě není známa.

2.4.6 Rodina ALDH7

Enzymy rodiny ALDH7 nazýváme Δ^1 -piperidin-6-karboxylátdehydrogenasy (P6CDH), α -aminoadipátsemialdehyddehydrogenasy nebo antiquitiny. Tato ALDH rodina je značně konzervovaná. Jelikož je míra podobnosti i mezi ALDH7 z evolučně velmi vzdálenými druhy velmi vysoká, je značně pravděpodobné, že i jejich funkce je významná pro funkci buňky. Například lidská ALDH7A sdílí asi 60% identitu s rostlinnými zástupci z podrodiny ALDH7B. Také bylo zjištěno, že ALDH7 z kukuřice (*Zea mays*) preferuje substrát α -aminoadipát semialdehyd, stejně jako lidská ALDH7A (Končítíková et al., 2015).

Genová exprese těchto rostlinných enzymů odpovídá nejspíše na stresové podmínky, stejně jako je tomu například u ALDH3. Například zvýšená exprese ALDH7B je důsledkem stresových podmínek jako je dehydratace, ultrafialové záření, salinita nebo nízká či příliš vysoká teplota (Kotchoni et al. 2006; Rodrigues et al. 2006). Specifická funkce ALDH7B je například v rýži, kde důležitá pro dozrávání semen. Rekombinantní ALDH7B byla také schopna metabolizovat MDA, acetaldehyd nebo glycerinaldehyd (Shin et al., 2009). Rozdílná je lokalizace těchto enzymů, zatímco u rostlin se vyskytují pouze v cytosolu, u savců byly nalezeny taktéž v mitochondriích (Wong et al., 2010).

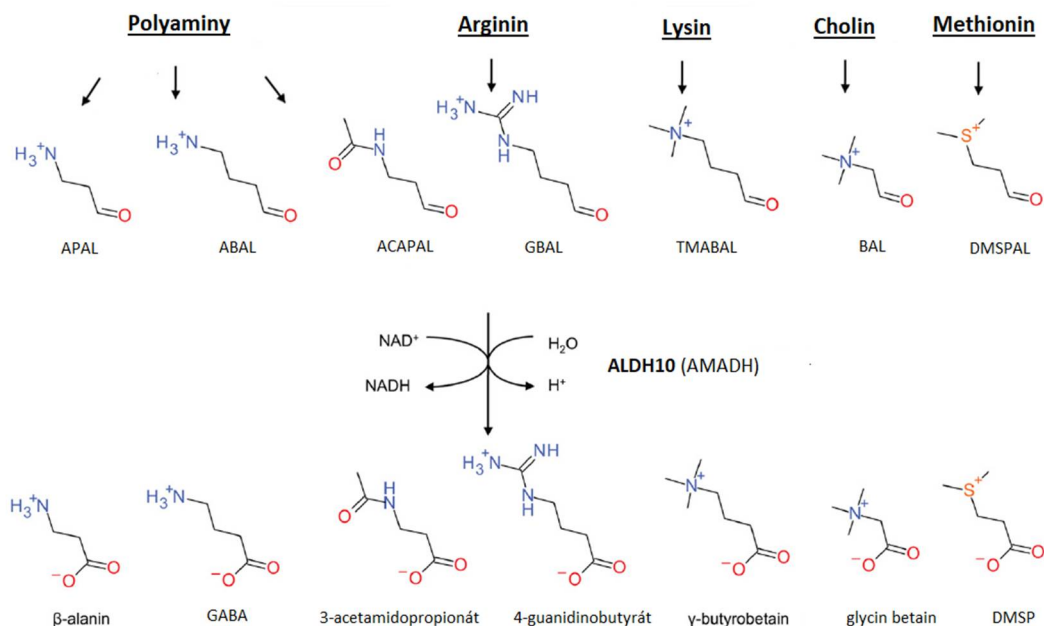


Obr. 12 Oxidace α -aminoadipát semialdehydu na α -aminoadipovou kyselinu pomocí ALDH7 s NAD^+ koenzymem. Upraveno dle Končítíková *et al.* (2015).

2.4.7 Rodina ALDH10

Rodina ALDH10 obsahuje enzymy známé jako aminoaldehyddehydrogenasy (AMADH), 4-aminobutyraldehyddehydrogenasy, 4-guanidinobutyraldehyddehydrogenasy nebo také betainaldehyddehydrogenasy (BADH). Tyto enzymy hrají důležitou roli při katabolismu polyaminů, kdy deaktivují reaktivní ω -aminoaldehydy jako 3-aminopropionaldehyd (APAL) nebo 4-aminobutyraldehyd (ABAL), které vznikají při oxidaci polyaminů rostlinnými aminooxidasami (Šebela *et al.* 2000). ALDH10 přeměňují tyto aminoaldehydy na netoxické metabolity jako β -alanin a γ -aminobutyrát (GABA). Dále ale také produkují osmoprotektivní glycin betain (GB) jako odpověď na suché podmínky. GB funguje jako významný buněčný osmolyt, který vyrovnává osmotickou rovnováhu a také chaperon, kdy pomáhá stabilizovat strukturu a funkci proteinů (Allakhverdiev *et al.* 2008).

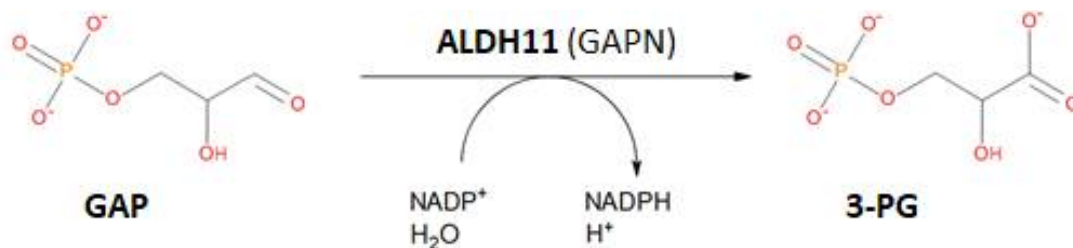
U savců se ALDH10 nevyskytují, jejich funkci převážně nahrazují ALDH9, které jsou považovány za hlavní enzymy syntetizující GB nebo γ -butyrobetain (Brocker *et al.* 2010; Vaz *et al.* 2000; Končítíková *et al.* 2019).



Obr. 13 Přehled reakcí oxidovaných ALDH10 (AMADH). APAL se přemění na β-alanin, ABAL na GABA, ACAPAL na 3-acetamidopropionát, GBAL na 4-guanidinobutyrát, TMABAL na γ-butyrobetain, BAL na glycin betain a DMSPAL na DMSP. Upraveno dle Kopečný *et al.* (2013).

2.4.8 Rodina ALDH11

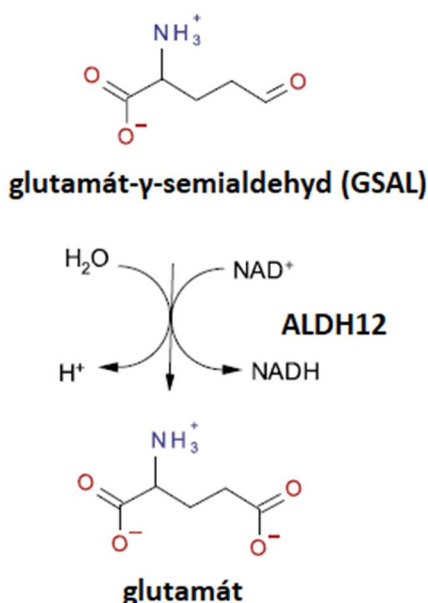
Genová rodina ALDH11 je reprezentována převážně cytosolickými nefosforylujícími glyceralddehyd-3-fosfátdehydrogenasami (GAPN). Tyto katalyzují jednosměrnou na NADP⁺ dependentní oxidaci GAP na 3-fosfoglycerát (3-PG) a NADPH (Valverde *et al.* 1999). Tato reakce je potřebná pro tzv. glykolytický bypass, což je reakce specifická pro fotosyntetizující eukaryotické organismy, která dovoluje přeskočit první fosforylační krok glykolýzy na úrovni substrátu (Plaxton 1996; Valverde *et al.* 1999). Tato reakce je mimo jiné také hlavním zdrojem NADPH rostlin, ta se pak využívá při biosyntéze mannitolu (Gao a Loescher, 2000).



Obr. 14 Přeměna glyceraldehyd-3-fosfátu na 3-fosfoglycerát. Nefosforylující reakce za pomoci ALDH11.

2.4.9 Rodina ALDH12

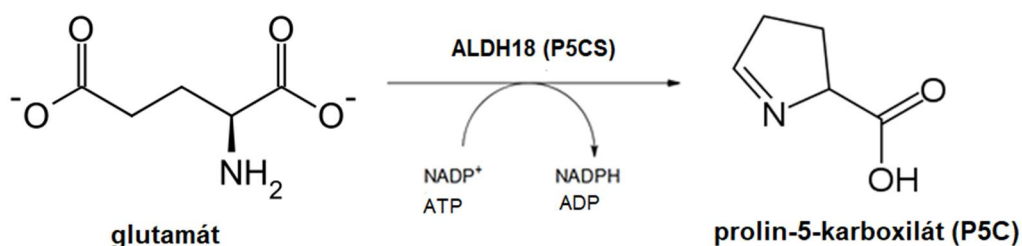
Geny ALDH12 představují Δ -1-pyrrolin-5-karboxylátdehydrogenasy nebo-li NAD^+ -dependentní glutamát- γ -semialdehyddehydrogenasy (GSALDH). Vyskytují se hojně u *Heterokont*, zelených řas *Charophyta* i *Chlorophyta*, taktéž u všech *Embryophyt*. Tyto enzymy hrají důležitou roli v procesu degradace prolinu nebo argininu na glutamát, kdy nezvratně přetvářejí glutamát γ -semialdehyd (GSAL) na glutamát. Exprese těchto genů je snížena jako odpověď na sucho či zvýšenou slanost, nejspíše z důvodu zajištění správné hladiny prolinu. Tato rodina má velmi nízkou podobnost k živočišným či bakteriálním GSALDH, které se řadí do rodiny ALDH4. Přesto jsou tyto dvě rodiny stále nejpodobnější v ALDH nadrodině, jelikož i jejich krystalická struktura je víceméně stejná, což značí funkční podobnost. Fylogenetická analýza naznačuje, že z ALDH4 vznikly ALDH12 ještě před vznikem zelených řas nebo přechodem rostlin na souš (Korasick et al. 2019).



Obr. 15 Oxidace GSAL na glutamát za pomoci NAD^+ jako kofaktoru. Tato reakce je součástí reakce degradace prolinu či argininu v rostlinách. Tato reakce může být zprostředkována také ALDH4 v jiných druzích. Upraveno dle Korasick et al. (2019).

2.4.10 Rodina ALDH18

Rodina ALDH18 obsahuje enzymy Δ -1-pyrrolin-5-karboxylátsynthasy (P5CS). Tyto enzymy jsou bifunkční proteiny, které obsahují N-koncovou kinásovou doménu a C-koncovou aldehyddehydrogenasovou doménu. U savců ALDH18A1 usnadňuje na ATP a NADPH dependentní přeměnu glutamátu na prolin-5-karboxylát (P5C), ten je pak přeměněn na ornitin a užít pro syntézu prolinu a argininu (Marchitti et al. 2008). U většiny rostlin se zvyšuje množství ALDH18 při nedostatku vody, jelikož prolin stejně jako betain funguje jako osmolyt a dokáže vyrovnávat osmotickou rovnováhu v organismu (Yoshida et al. 1997).

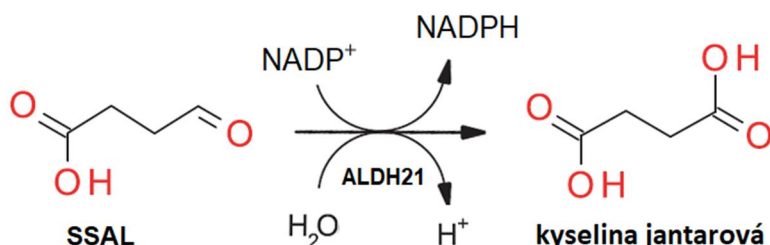


Obr. 16 Reakce katalyzovaná ALDH18 u savců. Přeměna glutamátu na prolin-5-karboxylát (P5C).

2.4.11 Rodina ALDH21

Tato rodina ALDH obsahuje geny specifické pro nižší rostliny, to značí zelené řasy, Bryophyta či kapradiny, nebyly však identifikovány v kvetoucích rostlinách. Exprese genů je zvýšena při dehydrataci. Zástupcem nalezeným v mechu *Physcomitrella patens* je tetramerní NADP^+ -dependentní sukcinátsemialdehyddehydrogenasa (SSALDH), která oxiduje v cytosolu sukcinát semialdehyd na sukcinát. Tato reakce je významná pro dráhu γ -aminomáselné kyseliny. Rodina ALDH21 silně preferuje substrát SSAL společně s NADP^+ jako koenzymem, jelikož NAD^+ je pro ni velmi špatný koenzym s aktivitou pouze asi 3% na rozdíl od ALDH5.

Nejpříbuznější rodinou ADLH z funkčního hlediska je tedy rodina ALDH5. Bylo zjištěno, že mechy, ale nejspíše i zelené řasy či kapradiny, které obsahují obě tyto rodiny, jsou schopné oxidovat SSAL na sukcinát nejen v mitochondriích ale i v cytosolu, na rozdíl od vyšších rostlin (Kopečná et al., 2017).



Obr. 17 Reakce sukcinátsemiialdehyddehydrogenasy (ALDH21), využívající NADP⁺ jako koenzym, při níž vzniká kyselina jantarová. Upraveno dle Kopečná et al. (2015).

2.4.12 Rodina ALDH22

Rodina ALDH22 byla identifikována teprve nedávno a není o ní příliš známo. Bylo zjištěno, že ALDH22A1 přítomná v kukuřici *Zea mays*, se vyskytuje převážně na plastidech a reaguje zvýšením exprese, na mnoho stresových podmínek jako dehydratace nebo zvýšená salinita (Huang et al. 2008). Toto zvýšené množství ALDH22 pak zvyšuje odolnost vůči těmto stresovým podmínkám.

2.4.13 Rodina ALDH23 a ALDH24

Rodiny ALDH23 a ALDH24 byly objeveny teprve nedávno, a tedy jejich význam a funkce v rostlinách je doposud neznámá. Bylo však zaznamenáno, že tyto rodiny jsou fylogeneticky příbuzné, jelikož například sekvence ALDH24 nalezena v *Chlamydomonas reinhardtii* sdílí přibližně 30 % identitu se sekvencemi některých genů z rodiny ALDH23. Sekvence rodiny ALDH23 byly identifikovány doposud pouze v nižších rostlinách jako například mech *Physcomitrella patens* nebo plavuně *Selaginella moellendorffii* (Brockner et al., 2013).

3 EXPERIMENTÁLNÍ ČÁST

3.1 Materiál

Jako sekvence pro hledání podobnosti jsme zvolili záznamy ALDH sekvencí publikované v Brocker et al. (2013), zaznamenané ve databázi Phytozome, především ALDH rodiny z kukuřice (*Zea mays*):

ALDH2B2 - GRMZM2G058675, ALDH2B5 - GRMZM2G125268, ALDH2C1 - GRMZM2G071021, ALDH2C2 - GRMZM2G097706, ALDH3E1 - GRMZM2G169458, ALDH3E2 - GRMZM2G155502, ALDH3H1 - GRMZM2G103546, ALDH3H2 - GRMZM2G060800, ALDH3H3 - GRMZM2G118800, ALDH5F2 - GRMZM2G119482, ALDH6B1 - GRMZM2G001898, ALDH7B6 - GRMZM2G130440, ALDH10A5 - GRMZM2G135470, ALDH10A8 - GRMZM2G013214, ALDH10A9 - GRMZM2G016189, ALDH12A1 - GRMZM2G090087, ALDH22A1 - GRMZM2G135341.

Dále sekvence z mechu (*Physcomitrella patens*): ALDH2B1 - Pp1s93_81V6, ALDH2B2 - Pp1s496_1V6, ALDH3H1 - Pp1s71_128V6, ALDH3H2 - Pp1s272_3V6, ALDH3K1 - Pp1s26_253V6, ALDH3K2 - Pp1s90_226V6, ALDH3K3 - Pp1s124_90V6, ALDH5F1 - Pp1s6_180V6, ALDH5F2 - Pp1s279_72V6, ALDH6B1 - Pp1s27_198V6, ALDH7B4 - Pp1s237_29V6, ALDH10A1 - Pp1s23_38V6, ALDH11A1 - Pp1s73_32V6, ALDH11A2 - Pp1s16_156V6, ALDH11A3 - Pp1s10_228V6, ALDH11A4 - Pp1s44_64V6, ALDH12A1 - Pp1s41_177V6, ALDH18B1 - Pp1s22_40V6, ALDH21A1 - Pp1s111_161V6, ALDH23A1 - Pp1s18_148V6.

Dále sekvence z huseníčku (*Arabidopsis thaliana*): ALDH2B4 - AT3G48000, ALDH2B7 - AT1G23800, ALDH2C4 - AT3G24503, ALDH3F1 - AT4G36250, ALDH3H1 - AT1G44170, ALDH3I1 - AT4G34240, ALDH5F1 - AT1G79440, ALDH6B2 - AT2G14170, ALDH7B4 - AT1G54100, ALDH10A8 - AT1G74920, ALDH10A9 - AT3G48170, ALDH11A3 - AT2G24270, ALDH12A1 - AT5G62530, ALDH22A1 - AT3G66658.

Ty jsem poté využil pro hledání podobných sekvencí ve zkoumaných genomech *Klebsormidium nitens*, *Coccomyxa subellipsoidea*, *Azolla filiculoides* a *Chromochloris zofingerensis*.

Oficiální stránka projektu *Klebsormidium nitens*:

http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/ (release: 171026)

Coccomyxa subellipsoidea phytozome databáze:

https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_CsubellipsoideaC_169

Azolla filiculoides, stránka projektu FernBase, software Blast:

https://www.fernbase.org/tools/blast?db_id=34

Chromochloris zofingerensis Phytozome databáze:

https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Czofingiensis_er

3.2 Metody

3.2.1 BLAST

Pro hledání genů ALDH ve zkoumaných rostlinách byl použit software BLAST, převážně jeho verze *blastp* (protein-protein blast) a *tblastn* (protein-translated nucleotide blast). Software jsem užíval skrze webové stránky NCBI, Phytozome nebo samotné stránky jednotlivých projektů sekvenovaných genomů jako Plantmorphogenesis nebo Fernbase. Program BLAST přijímá sekvence ve formátu FASTA, kdy je možné zadávat jednotlivé sekvence ručně, nebo rovnou celý soubor sekvencí. Následně se zvolí databáze, skrze kterou chceme vyhledávat a software se spustí.

The screenshot shows the NCBI BLAST search interface. At the top, there is a text input field for the query, labeled '1)'. To its right are 'Clear' and 'Query subrange' options. Below the query field are 'Or, upload file' and 'Job Title' sections. The 'Or, upload file' section has a 'Vybrat soubor' button and a 'Soubor nevybrán' label, labeled '2)'. The 'Job Title' section has a text input field and a label 'Enter a descriptive title for your BLAST search'. Below this is a checkbox for 'Align two or more sequences'. The 'Choose Search Set' section has a 'Database' dropdown menu set to 'Non-redundant protein sequences (nr)', labeled '3)'. Below it is an 'Organism' text input field with an 'exclude' checkbox, labeled '4)'. There are also checkboxes for 'Exclude' options: 'Models (XM/XP)', 'Non-redundant RefSeq proteins (WP)', and 'Uncultured/environmental sample sequences'. The 'Entrez Query' section has a text input field and a 'Create custom database' link. The 'Program Selection' section has radio buttons for different algorithms: 'Quick BLASTP (Accelerated protein-protein BLAST)', 'blastp (protein-protein BLAST)' (which is selected), 'PSI-BLAST (Position-Specific Iterated BLAST)', 'PHI-BLAST (Pattern Hit Initiated BLAST)', and 'DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)', labeled '5)'. At the bottom, there is a 'BLAST' button and a 'Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)' label, with a checkbox for 'Show results in a new window'.

Obr. 18 Interface programu BLAST (specificky verze blastp) na webu NCBI. Vzhled se může lišit na jednotlivých webových stránkách, princip je však vždy stejný. Program přijímá jednotlivé sekvence ve formátu FASTA (1) nebo lze nahrát i celý soubor sekvencí (2). Následně si můžeme zvolit typ databáze (3) (v tomto případě různé typy proteinových databází). Zvolíme organismus, jehož sekvence chceme prohledávat (4). Nakonec můžeme ještě upřesnit verzi programu BLAST kterou chceme využít (5).

3.2.1.1 Algoritmus BLAST

Basic Local Alignment Search Tool neboli BLAST je algoritmus používaný pro srovnávání neboli alignment sekvencí nukleotidů nebo aminokyselin z proteinů. Umožňuje tedy srovnání zadávané sekvence vůči určité databázi známých sekvencí a zjištění podobnosti těchto sekvencí. Program BLAST byl navržen v USA a spolupracovali na něm Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, a David J. Lipman z NIH (National Institutes of Health). Publikovali jej v Journal of Molecular Biology roku 1990 (Altschul et al., 1990).

3.2.1.2 Princip algoritmu

Pro průběh BLASTu je tedy potřeba hledaná sekvence ve formátu FASTA a cílová sekvence nebo databáze ve které hledáme. Program pak hledá subsekvence v databázi, které jsou podobné částem dotazované sekvence. BLAST hledá high-scoring

segment páry (HSP), pomocí heuristických algoritmů podobných Smith-Watermanovu algoritmu, který je sice přesnější, ale z důvodu své rychlosti nevhodný pro prohledávání větších databází (Mount, D. W., 2004).

Samotný algoritmus pak probíhá v několika krocích. Nejprve se ze zadané sekvence odstraní tzv. „low-complexity regions“ a repetice, které by mohli zpomalovat proces algoritmu a ovlivnit jeho přesnost při vyhledávání v databázi. Následně se zadaná sekvence rozdělí na k-prvkové podsekvence, obvykle se jedná o 3-prvkové sekvence pro proteiny nebo 11-prvkové nukleotidové sekvence, které jsou hledány v databázi. Podobnost se sekvencemi v databázi se určuje pomocí substitučních matic, jejichž typy mohou být IUPAC (identity matrix), která se využívá pro nukleové kyseliny, shody mají kladnou konstantní hodnotu, neshody zápornou či nulovou. Pro proteiny se pak využívají PAM nebo BLOSUM. Nejlépe hodnocené úseky sekvence pak nazýváme high-scoring páry. Alignment sekvencí, se svou substituční maticí, je dále rozšiřován, dokud se zvyšuje celkový součet skóre high-scoring párů. Jakmile se zvyšovat přestane, algoritmus končí. (Xiong, J., 2006)

Matice PAM („Point Accepted Mutation“)

Tato matice je historicky nejstarší maticí pro proteiny vytvořena týmem Margaret Dayhoff v roce 1978, zohledňuje velmi malé množství evolučních předpokladů, byla vypočítána u příbuzných proteinů. PAM1 byla například vytvořena pro proteiny s identitou větší než 85% a je základem pro tvorbu dalších matic. V současnosti se nejvíce využívají matice PAM30, PAM70 nebo PAM250 (Dayhoff et al., 1978).

Matice BLOSUM („BLOcks SUBstitution Matrix“)

Empirická matice vytvořena na základě lokálního přiřazení konzervativních domén v méně příbuzných sekvencích. Na rozdíl od matice PAM pracuje lépe s evolučně se lišícími sekvencemi. Čím vyšší je číslo matice, tím vyšší byl podíl konzervativních domén v porovnávaných sekvencích. Algoritmus BLAST využívá matici BLOSUM62. (Henikoff a Henikoff, 1992)

```

# Matrix made by matblas from blosum62.iiij
# * column uses minimum score
# BLOSUM Clustered Scoring Matrix in 1/2 Bit Units
# Blocks Database = /data/blocks_5.0/blocks.dat
# Cluster Percentage: >= 62
# Entropy = 0.6979, Expected = -0.5209
  A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  B  Z  X  *
A  4 -1 -2 -2  0 -1 -1  0 -2 -1 -1 -1 -1 -2 -1  1  0 -3 -2  0 -2 -1  0 -4
R -1  5  0 -2 -3  1  0 -2  0 -3 -2  2 -1 -3 -2 -1 -1 -3 -2 -3 -1  0 -1 -4
N -2  0  6  1 -3  0  0  0  1 -3 -3  0 -2 -3 -2  1  0 -4 -2 -3  3  0 -1 -4
D -2 -2  1  6 -3  0  2 -1 -1 -3 -4 -1 -3 -3 -1  0 -1 -4 -3 -3  4  1 -1 -4
C  0 -3 -3 -3  9 -3 -4 -3 -3 -1 -1 -3 -1 -2 -3 -1 -1 -2 -2 -1 -3 -3 -2 -4
Q -1  1  0  0 -3  5  2 -2  0 -3 -2  1  0 -3 -1  0 -1 -2 -1 -2  0  3 -1 -4
E -1  0  0  2 -4  2  5 -2  0 -3 -3  1 -2 -3 -1  0 -1 -3 -2 -2  1  4 -1 -4
G  0 -2  0 -1 -3 -2 -2  6 -2 -4 -4 -2 -3 -3 -2  0 -2 -2 -3 -3 -1 -2 -1 -4
H -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3  0  0 -1 -4
I -1 -3 -3 -3 -1 -3 -3 -4 -3  4  2 -3  1  0 -3 -2 -1 -3 -1  3  3 -3 -1 -4
L -1 -2 -3 -4 -1 -2 -3 -4 -3  2  4 -2  2  0 -3 -2 -1 -2 -1  1 -4 -3 -1 -4
K -1  2  0 -1 -3  1  1 -2 -1 -3 -2  5 -1 -3 -1  0 -1 -3 -2 -2  0  1 -1 -4
M -1 -1 -2 -3 -1  0 -2 -3 -2  1  2 -1  5  0 -2 -1 -1 -1 -1  1 -3 -1 -1 -4
F -2 -3 -3 -3 -2 -3 -3 -3 -1  0  0 -3  0  6 -4 -2 -2  1  3 -1 -3 -3 -1 -4
P -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4  7 -1 -1 -4 -3 -2 -2 -1 -2 -4
S  1 -1  1  0 -1  0  0  0 -1 -2 -2  0 -1 -2 -1  4  1 -3 -2 -2  0  0  0 -4
T  0 -1  0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1  1  5 -2 -2  0 -1 -1  0 -4
W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1  1 -4 -3 -2 11  2 -3 -4 -3 -2 -4
Y -2 -2 -2 -3 -2 -1 -2 -3  2 -1 -1 -2 -1  3 -3 -2 -2  2  7 -1 -3 -2 -1 -4
V  0 -3 -3 -3 -1 -2 -2 -3 -3  3  1 -2  1 -1 -2 -2  0 -3 -1  4 -3 -2 -1 -4
B -2 -1  3  4 -3  0  1 -1  0 -3 -4  0 -3 -3 -2  0 -1 -4 -3 -3  4  1 -1 -4
Z -1  0  0  1 -3  3  4 -2  0 -3 -3  1 -1 -3 -1  0 -1 -3 -2 -2  1  4 -1 -4
X  0 -1 -1 -1 -2 -1 -1 -1 -1 -1 -1 -1 -1 -2  0  0 -2 -1 -1 -1 -1 -1 -1 -4
* -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4

```

Obr. 19 Příklad skórovací matice BLOSUM62. Jednotlivé báze jsou ohodnoceny vzhledem k podobnosti a hodnota za mezeru (*) je -4.

3.2.1.3 Algoritmus FASTA a dynamické programování

FASTA je podobný algoritmus jako BLAST, oba algoritmy užívají heuristické přístupy pro rychlý alignment sekvencí, kdy hledají krátké totožné nebo velmi podobné úseky dvou sekvencí. BLAST je ovšem časově efektivnější díky hledání pouze high-scoring segmentů, přičemž přesnost obou algoritmů je srovnatelná. U těchto metod není zaručeno, že nalezneme všechny homology sekvencí, avšak jsou mnohem rychlejší než algoritmy Needleman-Wunch nebo Smith-Waterman, užívající dynamické programování (Xiong, J., 2006).

Algoritmus BLAST byl odvozen od algoritmů Needleman-Wunsch a Smith-Waterman, které využívají dynamické programování. Porovnává sekvence na lokální úrovni, stejně jako Smith-Waterman algoritmus.

Needleman-Wunsch algoritmus

Algoritmus Needleman-Wunsch porovnává sekvence na globální úrovni, tedy výsledný alignment sekvencí je stejně dlouhý jako nejdelší ze zadaných sekvencí. Jeho princip spočívá ve vytvoření tabulky, kdy na prvním horním řádku je zapsaná první sekvence o délce m , v podobě jednotlivých bazí a v prvním levém sloupci je zapsaná

druhá sekvence o délce n. Vytvoří se tedy tabulka o velikosti ((m+1) x (n+1)). Pro zvolení ideálního alignmentu sekvencí je užito skórovacího systému, který je nejčastěji založen na substituční matici. Algoritmus také zaznamenává směr procházení tabulky. Tabulka se vyplňuje podle vzorce:

$$\max ((H(i-1; j)+d), (H(i, j-1)+d), (H(i-1, j-1) +M(i, j)))$$

To znamená, že maximem je buď hodnota skóre pole vlevo + hodnota vložení mezery (inzerce), nebo skóre pole nahoře + hodnota vložení mezery (delece), nebo skóre pole vlevo nahoře + hodnota shody/neshody znaků dané pozice sekvencí (příslušné danému poli v matici). Cesta průchodu maticí se značí šipkami.

-	-	A	T	C	G	A	C
-	0	-4	-8	-12	-16	-20	-24
C	-4	-3	-7	-3	-7	-11	-15
A	-8	1	-3	-7	-6	-2	-6
T	-12	-3	6	2	-2	-6	-5
A	-16	-7	2	3	-1	3	-1
C	-20	-11	-2	-1	0	-1	8

Obř. 20 Tabulka ohodnocení Needleman-Wunsch algoritmu pro řetězce ATCGAC a CATAC. Červenými šipkami je značené porovnání s nejvyšším skóre.

Smith-Waterman algoritmus

Algoritmus Smith-Waterman je modifikací algoritmu Needleman-Wunsch, kdy porovnává sekvence na lokální úrovni, tedy pracuje s celou sekvencí, ale výsledkem může být alignment podsekvencí dané sekvence o různých délkách. Vstupem jsou dvě sekvence, substituční matice a sankční hodnota mezery. Liší se od Needleman-Wunsch algoritmu převážně tím, že hodnota skóre nikdy nebude záporná. Nejdřív se sestaví matice o velikosti ((m+1) x (n+1)). Ta se následně vyplňuje podle vzorce:

$$\max ((H(i-1; j)+d), (H(i, j-1)+d), (H(i-1, j-1) +M(i, j)), 0)$$

Tento vzorec funguje stejně jako u Needleman-Wunsch algoritmu, pouze nezachází do záporných hodnot.

-	-	A	T	C	G	A	A
-	0	0	0	0	0	0	0
C	0	0	0	5	1	0	0
A	0	5	1	1	2	5	5
T	0	1	10	6	2	1	2
A	0	5	6	7	3	7	6
C	0	1	2	11	7	3	4

Obr. 21 Tabulka ohodnocení Smith-Watermanova algoritmu pro řetězce ATCGAA a CATAC. Červenými šipkami je značené porovnání s nejvyšším skóre.

3.2.1.4 Program BLAST

BLAST v podobě programu lze stáhnout pod názvem „blastall“, ten následně funguje lokálně pomocí příkazového řádku nebo lze BLAST užívat na webovém serveru NCBI (The National Center for Biotechnology Information), kde může kdokoliv prohledávat databázi se sekvencemi většiny osekvenovaných organismů. Jelikož je BLAST open-source, bylo vyvinuto mnoho modifikací původního programu. Tyto verze se navzájem liší ať už v typu porovnávaných dat nebo způsobem srovnávání.

Nukleotid-nukleotid BLAST (blastn) – Verze programu, která bere jako vstup sekvenci DNA a prohledává databázi nukleotidových sekvencí DNA.

Protein-protein BLAST (blastp) – Verze programu, kde vstupem je proteinová sekvence a hledá se nejpodobnější v dané proteinové databázi.

Position-Specific Iterative BLAST (PSI-BLAST, blastpgp) – Verze programu využívaná k hledání vzdáleně příbuzných proteinů. Nejdříve nalezne blízké příbuzné proteiny, z nich se následně vytvoří profilová sekvence, která obsahuje význačné znaky těchto sekvencí. Za pomoci profilové sekvence se prohledává databáze a nalezne se opět velké množství sekvencí proteinů. Z nalezených sekvencí se vytvoří opět jedna profilová sekvence a proces se opakuje. PSI-BLAST je přesnější než ostatní verze BLAST, jelikož identifikuje i značně vzdálené evoluční vztahy jednotlivých příbuzných proteinů.

Nucleotide 6-frame translation-protein (blastx) – Verze BLAST, která porovnává abstraktní translační produkty šesti čtecích rámců zadané nukleotidové sekvence oproti proteinové databázi.

Nucleotide 6-frame translation-nucleotide 6-frame translation (tblastx) – Verze programu, která porovnává šest čtecích rámců dotazované nukleotidové sekvence vůči šesti čtecím rámcům sekvencí v zadané nukleotidové databázi. Takto je algoritmus schopen nalézt vzdálené vztahy mezi sekvencemi za cenu rychlosti prohledávání.

Protein-nucleotide 6-frame translation (tblastn) – Verze programu která porovnává proteinovou sekvenci vůči nukleotidové databázi v šesti čtecích rámcích sekvencí této databáze.

3.2.1.5 Využití BLASTu

BLAST je významný pro mnoho procesů, využívá se například při identifikaci druhů, kdy pomocí BLASTu můžeme zjistit pravděpodobnost identity druhů nebo nalézt druhy homologní, pokud pracujeme s DNA neznámého druhu. Dále se využívá při lokalizaci domén v sekvenci proteinu, stanovení fylogeneze a vytvoření fylogenetických stromů, ačkoli fylogeneze založena pouze na BLAST nemusí být vždy přesná, nebo DNA mapování, kdy v databázi známého druhu hledáme pozici určitého genu na chromozomu.

3.2.2 MUSCLE

Pro nalezené sekvence programem BLAST jsem pak provedl multiple alignment pomocí softwaru MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>). Program opět přijímá sekvence ve formátu FASTA, zadané sekvence pak porovnává mezi sebou. MULTiple Sequence Comparison by Log-Expectation (MUSCLE) je software sloužící k mnohočetnému porovnání sekvencí proteinů nebo nukleotidů. Tento algoritmus publikoval poprvé Robert C. Edgar v roce 2004 v Nucleic Acids Research (Edgar R. C., 2004).

The screenshot shows the MUSCLE web interface with three steps:

- STEP 1 - Enter your input sequences:** A text area for entering or pasting sequences. A red '1)' is shown in the text area. Below the text area are buttons for 'Vybrat soubor' (selected) and 'Soubor nevybrán' (2). There are also links for 'Use a example sequence', 'Clear sequence', and 'See more example inputs'.
- STEP 2 - Set your Parameters:** A dropdown menu for 'OUTPUT FORMAT:' with 'ClustalW' selected (3). Below it is a note: 'The default settings will fulfill the needs of most users.' and a button for 'More options...' with a link to view or change default settings.
- STEP 3 - Submit your job:** A checkbox for 'Be notified by email' (with a note to tick if notified) and a 'Submit' button.

Obr. 19 Interface programu MUSCLE. Je možné zadat sekvence jednotlivě ve formátu FASTA (1), nebo jako celý soubor (2). Zvolí se výstupní formát (3) a spustí se algoritmus.

Samotný algoritmus funguje ve třech etapách. Nejdříve se vytvoří konceptuální multiple alignment, který není velmi přesný, to je však kompenzováno rychlostí jeho vytvoření. Následně se využije Kimurova dvouparametrová vzdálenost, která opravuje porovnání, kdy bere v potaz transnitní a transverzní substituci purinu. Zpřesní se odhad konceptuálního alignmentu a vznikne nový multiple alignment. Poslední fáze algoritmu ještě více zpřesní alignment vzniklý v druhé etapě za cenu zvýšené časové složitosti. MUSCLE se často využívá místo známějšího softwaru Clustal, jelikož tvoří většinou přesnější alignmenty a jeho proces je také rychlejší (Edgar R. C., 2004). Jedná se o globální přiřazení, sekvence by tedy měly být přibližně stejně dlouhé a evolučně příbuzné.

Podobným softwarem využívaným pro multiple alignment je T-coffee, jehož hlavní výhodou je například možnost porovnávat společně proteinové i RNA sekvence a je také rychlejší a přesnější než MUSCLE, pokud se nejedná o porovnávání příliš evolučně vzdálených sekvencí (Notredame et al., 2000). MAFFT je další multiple alignment software, který využívá při svém algoritmu rychlou Furierovu transformaci pro seřazení sekvencí a je nejvýhodnější při využití pro velmi velký počet porovnávaných sekvencí (Kato et al., 2002).

Výstupem programu MUSCLE je automaticky formát CLUSTAL, podle stejnojmenného programu, je též označován jako „interleaved“. Sekvence jsou zapsány jako přiřazení pod sebou a jejich názvy mají přesně danou délku svých názvů. Jako

výstup můžeme zvolit i jiné formáty, jako jsou například FASTA nebo PHYLIP 3.2, na rozdíl od CLUSTAL jsou označovány jako „sequential“. Sekvence jsou zapsány odděleně i s nově vzniklými mezerami z přiřazení. Pro další práci se sekvencemi jsem užíval formát FASTA.

```

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

azALDH5-2      -----MIRDRAARGTSAQGRGRSRRGSPLERKTATSRSLTFPCLVSWSGLTRPSRVSPRE
azALDH5-3      -----
klALDH5F       ---MNVKQARIILGFRQLFCHRSAGATSIASQTSGARYRSCVVPNSIQDVHSTLHSRCG
VvALDH5F3      -----MAKFLFDYYFYQYFNQNRP-----THLASWHYNSVMTYGSQLLFIVSEGFSRFG
azALDH5-1      MMMVFYSRRGWDFGLLVQKESACVPHPDLIHSSSFFLTAAAAAVPALNPSSTPLSSSST
PpALDH5A       -----
PpALDH5B       -----
SmALDH5F1      -----
VvALDH5F2      -----
OsALDH5        -----MAMAMAMRRAAALGARHILAA
SbALDH5F1      -----MAMAMMTLRRRAVALGARHIPAV
ZmALDH5        -----MAMAMMAMRRAVALGARHIPAA
AtALDH5F1      -----MVIGAAARVAIGGCCKLIS
VvALDH5F1      -----MGLLRSSCSALCH
PtALDH5F1      -----MTLGRIALARIPATRCKIHSLFS
coALDH5G       -----MAGLHRLSLTLSLHRQMMMLTSSIS----
chALDH5F       -----MLAAAGAGHLLAAGTLVGSQHLLAH

azALDH5-2      -----RRGHKERPRIAPRPFEDRLRDHSDAQAEILDGVVIEAVFRAFAALAGLLDAT
azALDH5-3      -----QDFLIAVTRSKVFNNGIPMDIVKDASLFRE----AFYSDGA
klALDH5F       SGGGGAGKRDLIGRFTQRTFHLTPTCQMALDSGSSNSVPKLKDPSLLKD----LGLIGGK
VvALDH5F3      -----FHNLILLICHFYRQNFLLIQISMDPQSAIAQVKRAGLFRT----QGLIGGK
azALDH5-1      VMQASLSKLRLLHHASKAITTGFAQLRYLSTPSDEAVIHKIKEAGLLQS----KGLIGGK
PpALDH5A       -----MGFAQVGAMVQLKNMGLFKT----QGLINGE
PpALDH5B       -----MVQLNEAGLFKS----QGLIGDK
SmALDH5F1      -----MATIASAGASQAVLDKINNAGLLRA----KGLISGE
VvALDH5F2      -----MEAQSVITRFQNSGVFRT----QGLIGGK
OsALDH5        -----SSTSSSGVLLRRHMSVDAGAAMEKVRAAGLLRT----QGLIGGK
SbALDH5F1      -----AAASSRIVPLRHMSAGAGAAVEKIRAAGLLRT----QGLIAGK
ZmALDH5        -----AASSFRVVSRLRHMSADAGAAMEKIRAAGLLRT----QGLIAGQ
AtALDH5F1      -----SHTSLLLVSQCRCQMSMDAQSVSEKLRSSGLLRT----QGLIGGK
VvALDH5F1      -----GPRTASVLRPSAVLTRQISMDTQNLVARLNSSGLLRS----QCLIGGK
PtALDH5F1      -----APHSSSSTNRTSPPLSRHMSMKSENLVSKLTSSGLLKT----QGLIDGK
coALDH5G       -----RLQQFVGTQAFHVTA-AEELLNKVHDKDLVKT----LGFINGE
chALDH5F       -----ALRWGMRAASSAQPEPEMARI-AEDLTSRVTDKDLLHT----CGFVGGK

.           :..

azALDH5-2      ERGDFGGDQAGVDSDHAIFFELGNPEDFRDVARVE-IGRQAE LGVVGHGNDLFLGVELEQ
azALDH5-3      WHKG-GADVIVT-----NPATGKVIQTVPKATTEQTRAAIEAAHAAPDWAARP

```

Obr. 20 Příklad výstupu formátu Clustal programu MUSCLE pro zde zkoumané ALDH5. Jednotlivé sekvence mají předpony podle rostlin, ze kterých byly získány. Použité zkratky: az – *Azolla filiculoides*, kl – *Klebsormidium nitens*, Vv – *Vitis vinifera* (vinná réva), Pp – *Physcomitrella patens*, sm – *Selaginella moellendorffii*, os – *Oriza sativa* (rýže), sb – *Sorghum bicolor* (čirok), At – *Arabidopsis thaliana* (huseníček rolní), Pt – *Populus trichocarpa* (topol), Zm – *Zea mays* (kukuřice), co – *coccomyxa subellipsoidea*, ch – *Chromochloris zofingiensis*.

```

>azALDH5-2
-----MIRDRAARGTSAQGRGRSRRGSPLERKTATSRSLTFPCLVSWSGLTRPSRVSPRE
-----RRGHKERPRIAPRPFEDRLRDHSDAQAEILD LGVVIEAVFRAFAALAGLLDAT
ERGDFGGDQAGVSDHAIFFELGNPEDFRDVARVE-IGRQAE LGVVGHGNDLFLGVELEQ
RRDGAEGFLAGDLHLSGDAG-----ED
GRLEEQAALRIRLAADRHS GALGHGIGDVF LDLVDGGAGDQRTGIDAGRKAVADLHRRHL
GGQLFGEGIIDAGLDIEAVGA---DAGLAVVTVFGDHRAFDGGIKIGVVKDDERRIAAQ
FQAE LHDLV GALAHQDLADTGRAGKQQLADLVILGELLADRRRILRSDDVENAGGQAGAF
GQQRQRQRQRCLGGGLGDHGAAGGHGRRELAGQHGVGEVPRGDRGDDADRLD DDDDALV
RLVAGNGVAIDALGF LAEPLDIGGAI EHLALGLGHRLAH-----FHGQDG
-GEVIGIGDHQVVELAQHGGAFLAGL-----LGPDFLRGVGG-----
-----IDGAGGLCAVHVGQRGDDSTIGRIGDIKGLAAVALGN
PFARDQRLVSSQRRVLELGQQSFGNRHGSGLRSAFAQCVFENIHRFGKHCILDGDRRQEA
DDIAIDAAGQEDQPLGQRLFAHLVGEIGIRGLGARNRLQALLEAVAERHGPFVEALFLDD
VEHGIAGSDADRVA AIGAAEAAGAGRIHDVGAAGDGS D-----
-----
-----REAAAEALGHGGDVGDNAVMFHRKQFSRAGKA-----
-----
>azALDH5-3
-----
-----QDFLIAVTRSKVFNRGIPMDIVKDASLFRE----AFYSDGA
WHKG-GADVIQVT-----NPATGKVI GTVPKATTEQTRAAIEAAHAAPDWAARP
AKERSAILRKW-----FE
LIVENADDLGRLLTAEQ----GKPF AEARGEI-LYGAFFEFWFAEDAKRISGET----I
PAPTRDKRMLVIRQPIGVCSAITPWNFP MAMIP----RKAGPAIAAGCT-----MV-
-----VVTGAAGVVGEELVTNPLV-----
-----RKITFTG-----STEVGRELMAKAAKTIKKVSLELGGNAPLIVFDD----
-----ASLETAVQ GALASKFRNMG--QTCVCANRIYVQ-----DG
IYDAFVERFAEVVSAMKVG DGFEEGVL-----QGP--LIDEAA-----
-----VAKVEDHVRDAL-EGGARIVTG--GKRHAL-----GG
TFYEPTVLADATQSMKIAREETFGP-----VAPIIRF-----TD
EKEVIRLANDTEFGLSAYFFTRDL-----ARTWRVAEAL----
-EAGQMGIN TGIT----AFEGAPFGGVKQSGIGREGSHHGIDEYPSLESDIDEIGAEDSG
DLGPAGR DYVNALARGLEVIKVFTRFTPKMTLSEVSQATGMTRATVRRFLLTLAHEGYVE
ISAKHFSLRPKILELGFSALSSMDLWDVVQPVINDLSEKLGESCFSAVL DGD SVIYVARA
TPNRRVNVGISIGSRVPAHCVSTGRVLLAALPEEQ LHQY
>k1ALDH5F
---MNVKQARIILGFRQLFCHRSAGATSIASQTS GARYRSSCVVPSNIQDVHSTLHSRCG
SGGGGAGKRDLIGRFTQRTFHLTPTCQMALDSGSSNSV PKLKDPSLLKD----LGLIGGK
WVGAHDGKTMPVH-----NPATGEVVAE VPLMGAQETA EAIAAADEAFQTWSKTT
AKHRSQ L LKKW-----YD
LMLANKEDLAVLMTVEQ----GKPLAEARGEI-EYGASFVEWFAEEAKRMYGDV----I
PAARPDNRILVLKQPVGVIAAITPWNFPVAMAT----RKLAPALAAGCT-----AVL

```

Obr. 21 Výstup programu MUSCLE ve formátu FASTA. Na rozdíl od formátu Clustal jsou vždy zapsané celé sekvence postupně za sebou.

3.2.3 Gblocks

Následně byly v programu Gblocks odstraněny větší mezery a také úseky s nízkou podobností porovnávaných sekvencí, díky tomu mohla být provedena přesnější fylogenetická analýza. Tento program přijímá sekvence ve formátu FASTA. Webový interface je k dispozici na adrese:

http://molevol.cmima.csic.es/castresana/Gblocks_server.html

Gblocks Server

Paste an alignment in NBRF/PIR or FASTA format:

1)

Or upload an alignment file: 2)

Vybrat soubor | Soubor nevybrán

Type of sequence: 3)

DNA || Protein || Codons

Options for a less stringent selection: 4)

Allow smaller final blocks

Allow gap positions within the final blocks

Allow less strict flanking positions

Options for a more stringent selection:

Do not allow many contiguous nonconserved positions

Get Blocks Clear

The software and its accompanying documentation are provided as is, without guarantee of support or maintenance.

*Institut de Biologia Evolutiva (CSIC-UPF)
Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain*

Obr. 22 Interface programu Gblocks. Je možné zadat sekvence jednotlivě ve formátu FASTA (1), nebo jako celý soubor (2). Dále zvolíme, o jaké sekvence se jedná (3). Nakonec se můžeme ještě vybrat, zda chceme ponechat určité mezery mezi sekvencemi nebo třeba méně přesné porovnání, a tedy eliminovat pouze naprosto nepodobné části sekvencí (4).

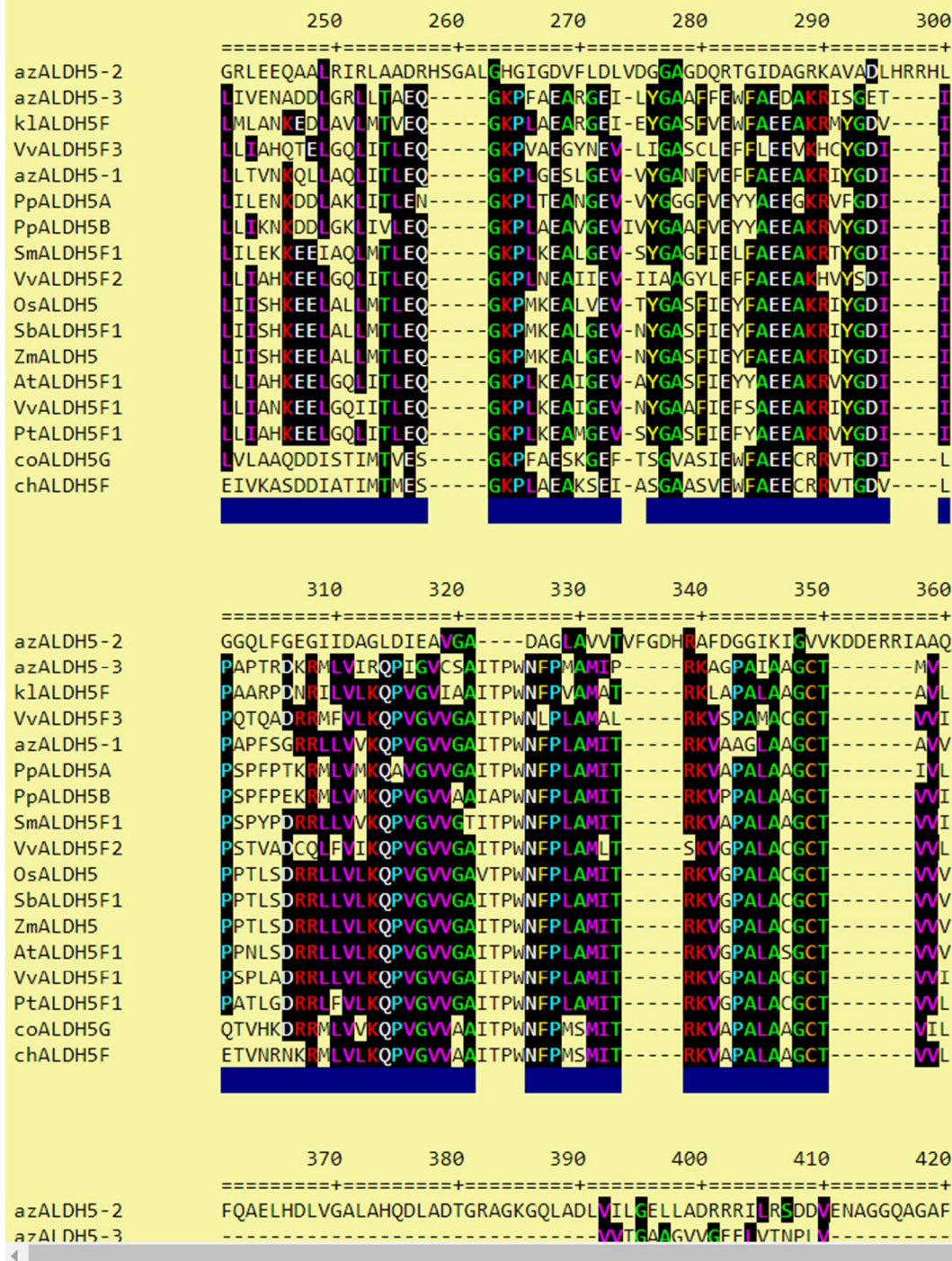
Gblocks 0.91b Results

Processed file: **ALDH5-MUSCLE.fas**

Number of sequences: **17**

Alignment assumed to be: **Protein**

New number of positions: **367** (selected positions are underlined in blue)



Obr. 23 Výstup programu Gblocks. Modrými čarami jsou označeny části sekvencí s největší podobností, a tedy ty části, které zůstávají ve výsledném souboru. Výsledný soubor programu Gblocks je opět FASTA, obsahující však pouze části sekvencí s vyšší podobností.

3.2.4 PhyML

Na zkrácených a upravených sekvencích nakonec byla provedena fylogenetická analýza pomocí softwaru PhyML. Phylogenetic estimation using Maximum Likelihood (PhyML) je software, který určuje maximum likelihood fylogeneze na základě alignmentu sekvencí nukleotidů nebo aminokyselin. Jako vstup přijímá program pouze sekvence ve formátu PHYLIP. Hlavní výhodou tohoto softwaru je jeho velký výběr substitučních modelů, od velmi rychlých metod až po pomalejší, ale mnohonásobně přesnější metody. Nejnovějším substitučním modelem je model LG, který je momentálně taktéž nejvyužívanější. Oproti starším modelům jako je WAG nebo JTT byla zjištěna zvýšená přesnost v 90 % porovnávaných sekvencí. (<http://www.atgc-montpellier.fr/models/index.php?model=lg>)

Pro svou analýzu jsem využíval AIC (Akaike Information Criterion), který automaticky vybere vhodný model pro zadané sekvence, na základě určité penalizační hodnoty, pro počet parametrů, aby ztráta dat při fylogenezi byla minimální. Druhou možností automatického výběru substitučních modelů je BIC (Bayesian Information Criterion), který funguje na značně podobném principu jako AIC, pouze využívá vyšší penalizační hodnoty. PhyML mimo jiné také využívá dvou metod pro ohodnocení jednotlivých větví fylogenetických stromů v daném statistickém rámci, a to takzvaný bootstrapping (obr. 25) nebo test přibližného poměru pravděpodobností (the approximate likelihood ratio test).

PhyML online execution

Input Data

Sequences (PHYLIP format) Soubor nevybrán **1)** File Example (AA file) (from Phylogenetic Handbook) DNA Amino-Acids **2)**

Substitution Model

Automatic model selection by SMS

Selection criterion AIC (Akaike Information Criterion) BIC (Bayesian Information Criterion) **3)**

Set by user

Substitution model LG

Equilibrium frequencies model empirical estimated

Transition / transversion ratio (DNA models) fixed estimated

Proportion of invariable sites fixed estimated

Number of substitution rate categories fixed estimated

Gamma shape parameter fixed estimated

Tree Searching

Starting tree(s) Soubor nevybrán File BIONJ

Type of tree improvement NNI

Number of random starting tree yes no

Branch Support

Fast likelihood-based method **4)** Use aLRT or aBayes to save computing time yes no

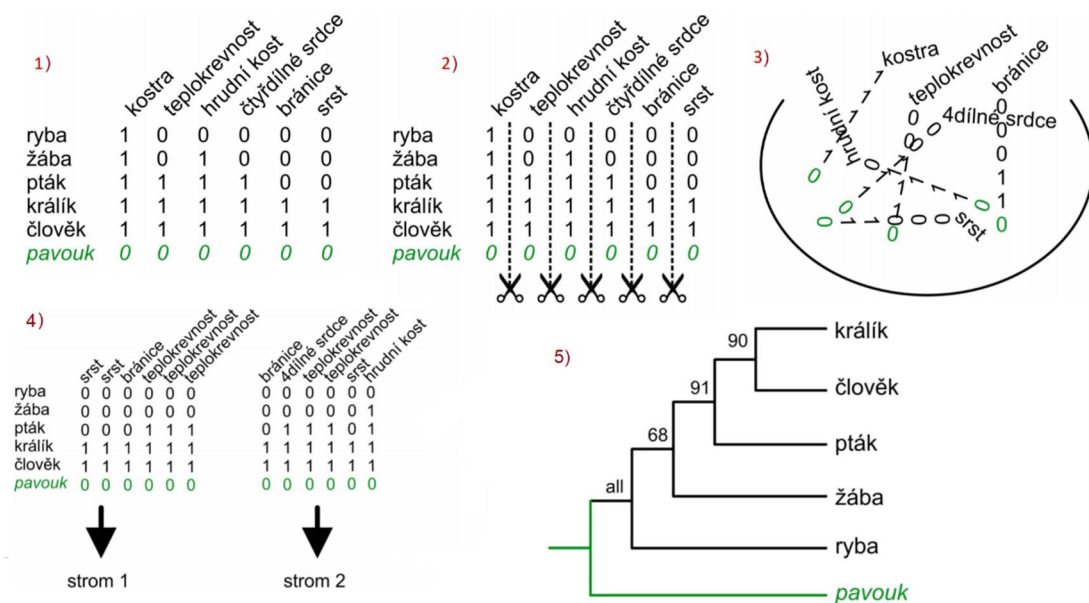
Perform bootstrap 100 yes no

Name of your analysis

Your email

Please confirm your email

Obr. 24 Interface programu PhyML. Přijímá soubory pouze ve formátu PHYLIP (1). Dále zvolíme, o jaké sekvenze se jedná (DNA/Aminokyseliny) (2). Můžeme si vybrat různé substituční modely a nastavit si je podle našich preferencí. Pro své sekvenze jsem však využíval pouze automaticky vybraný model dle AIC (3). Nakonec můžeme zvolit ohodnocení jednotlivých větví fylogenetických stromů. Využíval jsem bootstrapping o počtu 100 bootstrapů. (4)



Obr. 25 Princip bootstrappingu. Jednotlivé sekvenze z rostlin si můžeme představit jako živočichy s určitými znaky (1). Jednotlivé znaky se pak rozdělí (2) a následně náhodně vytahují (3) a skládají dohromady pro vytvoření „pseudomatic“ (tyto mají stejný počet druhů a znaků, některé znaky se však mohou opakovat) (4). Těchto matic, a tedy i stromů se vytvoří zadaný počet (v našem případě 100) a ve výsledném stromě je pak zaznamenáno kolika procentuálně byly jednotlivé znaky zastoupeny dohromady ve všech pseudomaticích (5).

4 VÝSLEDKY

4.1 Přehled nalezených genů

Zkoumal jsem genomy řas *Klebsormidium nitens*, *Coccomyxa subellipsoidea*, *Chromochloris zofingiensis* a kapradiny *Azolla filiculoides*. Výsledkem je nalezení 13 *ALDH* genů u řasy *Klebsormidium nitens*, 10 *ALDH* genů u řasy *Coccomyxa subellipsoidea* a pouze 9 *ALDH* genů u *Chromochloris zofingiensis*, naopak u kapradiny bylo nalezeno už 26 *ALDH* genů. Potvrzuje se tedy předpokládaný trend, že *ALDH* nadrodina expandovala s přechodem rostlin z moře na souš (viz Tab. 1).

Tab 1. Souhrnný počet *ALDH* genů v *Arabidopsis thaliana* (huseníček rolní), *Zea mays* (kukuřice setá), *Physcomitrella patens* (čepenka odstálá), *Selaginella moellendorffii* podle (Brocker et al. 2013) a počet *ALDH* v zástupcích *Azolla filiculoides* (Azola americká), *Klebsormidium nitens*, *Coccomyxa subellipsoidea*, *Chromochloris zofingiensis* zkoumaných v této bakalářské práci. Velikost genomu je udávána v mega párech bazí (Mb).

Druh	Oddělení	Velikost genomu (Mb)	Počet chromozómů	Počet genů <i>ALDH</i>
<i>Arabidopsis thaliana</i>	Krytosemenné rostliny	135	5	16
<i>Zea mays</i>	Krytosemenné rostliny	3000	10	22
<i>Physcomitrella patens</i>	Mechy	480	27	21
<i>Selaginella moellendorffii</i>	Plavuně	100	27	24
<i>Azolla filiculoides</i>	Kapradiny	750	22	26
<i>Klebsormidium nitens</i>	Řasy (<i>Charophyta</i>)	104	6	13
<i>Coccomyxa subellipsoidea</i>	Řasy (<i>Charophyta</i>)	49	20	10
<i>Chromochloris zofingiensis</i>	Řasy (<i>Charophyta</i>)	58	19	9

4.2 Geny *ALDH* přítomné v řase *Klebsormidium nitens*

V řase *Klebsormidium nitens* byly identifikovány všechny rostlinné rodiny *ALDH*, přičemž každá obsahovala právě jednoho zástupce. Výjimkou byly rodiny *ALDH2*, která obsahovala dva geny a rodina *ALDH24*, která v této řase identifikována nebyla.

Tab. 2 Geny *ALDH* přítomné v řase *Klebsormidium nitens*. Genom organismu byl prohledáván skrze webovou databázi NCBI a web projektu Plantmorphogenesis. V tabulce zaznačeny jsou ID jednotlivých proteinů z těchto databází kódovaných příslušnými geny.

Rodina ALDH	Název genu	NCBI protein ID	Plantmorphogenesis ID	Počet aminokyselin
<i>Rodina 2</i>	<i>ALDH2B1</i>	GAQ84628.1	kfl00198_0100_v1.1	495
	<i>ALDH2F1</i>	GAQ83917.1	kfl00168_0120_v1.1	573
<i>Rodina 3</i>	<i>ALDH3L1</i>	GAQ82469.1	kfl00112_0130_v1.1	478
<i>Rodina 5</i>	<i>ALDH5F1</i>	GAQ91062.1	kfl00722_0010_v1.1	580
<i>Rodina 6</i>	<i>ALDH6B1</i>	GAQ92677.1	kfl01091_0010_v1.1	564
<i>Rodina 7</i>	<i>ALDH7B1</i>	GAQ80312.1	kfl00051_0130_v1.1	521
<i>Rodina 10</i>	<i>ALDH10A1</i>	GAQ87015.1	kfl00325_0190_v1.1	544
<i>Rodina 11</i>	<i>ALDH11A1</i>	GAQ90687.1	kfl00674_0010_v1.1	497
<i>Rodina 12</i>	<i>ALDH12A1</i>	GAQ87509.1	kfl00358_0030_v1.1	535
<i>Rodina 18</i>	<i>ALDH18B1</i>	GAQ89957.1	kfl00582_0010_v1.1	746
<i>Rodina 21</i>	<i>ALDH21A1</i>	GAQ88729.1	kfl00454_0140_v1.1	575
<i>Rodina 22</i>	<i>ALDH22A1</i>	GAQ84711.1	kfl00202_0060_v1.1	621
<i>Rodina 23</i>	<i>ALDH23A1</i>	GAQ77842.1	kfl00004_0420_v1.1	496

Tab. 3 Tabulka podobnosti ALDH sekvencí z *Klebsormidium nitens* s mechem *Physcomyrella patens* (čepenka odstálá), kukuřicí *Zea mays* (kukuřice setá) a *Arabidopsis thaliana* (huseníček rolní). PpALDH značí proteinové sekvence z mechu *Physcomyrella Patens*, ZmALDH značí proteiny z jednoděložné (monokot) kukuřice, AtALDH značí sekvence z dvouděložné (dikot) *Arabidopsis*. Značení „-“ znamená chybějící sekvence.

Rodina ALDH	Název proteinu	Nejpodobnější ALDH v mechu	Podobnost k mechu (%)	Nejpodobnější ALDH v kukuřici	Nejpodobnější ALDH v Arabidopsis	Podobnost k monokotům/dikotům (%)
<i>Rodina 2</i>	ALDH2B1 ALDH2F1	PpALDH2B PpALDH2B	67 47	ZmALDH2-1 ZmALDH2-6	AtALDH2B7 AtALDH2B4	66 / 66 48 / 48
<i>Rodina 3</i>	ALDH3L1	PpALDH3E	49	ZmALDH3-5	AtALDH3I1	46 / 47
<i>Rodina 5</i>	ALDH5F1	PpALDH5B	65	ZmALDH5	AtALDH5F1	63 / 61
<i>Rodina 6</i>	ALDH6B1	PpALDH6B	72	ZmALDH6	AtALDH6B2	69 / 67
<i>Rodina 7</i>	ALDH7B1	PpALDH7B	73	ZmALDH7	AtALDH7B4	73 / 71
<i>Rodina 10</i>	ALDH10A1	PpALDH10A	66	ZmALDH10-1B	AtALDH10A9	63 / 65
<i>Rodina 11</i>	ALDH11A1	PpALDH11D	75	ZmALDH11	AtALDH11A3	79 / 78
<i>Rodina 12</i>	ALDH12A1	PpALDH12	63	ZmALDH12	AtALDH12A1	61 / 62
<i>Rodina 18</i>	ALDH18B1	PpALDH18B1	66	ZmALDH18B1	AtALDH18B2	64 / 63
<i>Rodina 21</i>	ALDH21A1	PpALDH21	63	-	-	0 / 0
<i>Rodina 22</i>	ALDH22A1	-	0	ZmALDH22	AtALDH22	56 / 58
<i>Rodina 23</i>	ALDH23A1	PpALDH23	59	-	-	0 / 0

4.3 Geny ALDH přítomné v řase *Coccomyxa subellipsoidea*

V řase *Coccomyxa subellipsoidea* byly identifikovány všechny rostlinné rodiny ALDH kromě rodin ALDH3, ALDH7 a rodin ALDH23 a ALDH24. Všechny nalezené rodiny byly zastoupeny právě jedním genem, výjimkou byla rodina ALDH2, která obsahovala právě dva zástupce.

Tab. 4 ALDH geny přítomné v řase *Coccomyxa subellipsoidea*, jejíž genom byl prohledán skrze webovou databázi NCBI a web Phytozome, kde je možné zjistit i počet exonů. V tabulce označeny jsou ID jednotlivých proteinů z těchto databází kódovaných příslušnými geny. Značení „-“ znamená chybějící sekvence.

Rodina ALDH	Název genu	NCBI protein ID	Phytozome ID	Počet exonů	Počet aminokyselin
Rodina 2	<i>ALDH2B1</i>	XP_005642981.1	31641	12	486
	<i>ALDH2B2</i>	XP_005642984.1	54894	14	565
Rodina 5	<i>ALDH5F1</i>	XP_005651730.1	34715	12	537
Rodina 6	<i>ALDH6B1</i>	XP_005647038.1	29754	13	544
Rodina 10	<i>ALDH10A1</i>	XP_005645499.1	37718	8	498
Rodina 11	<i>ALDH11A1</i>	XP_005650540.1	52649	15	496
Rodina 12	<i>ALDH12A1</i>	XP_005643005.1	54900	11	542
Rodina 18	<i>ALDH18E1</i>	XP_005645012.1	-	-	778
Rodina 21	<i>ALDH21A1</i>	XP_005644523.1	19137	10	500
Rodina 22	<i>ALDH22B1</i>	XP_005650418.1	12562	9	517

Tab. 5 Tabulka podobnosti ALDH sekvencí z *Coccomyxa subellipsoidea* s mechem *Physcomyrella patens* (čepenka odstálá), kukuřičí *Zea mays* (kukuřice setá) a *Arabidopsis thaliana* (huseníček rolní). PpALDH značí proteinové sekvence z mechu *Physcomyrella patens*, ZmALDH značí proteiny z jednoděložné (monokot) kukuřice, AtALDH značí sekvence z dvouděložné (dikot) *Arabidopsis*. Značení „-“ znamená chybějící sekvence.

Rodina ALDH	Název genu	Nejpodobnější ALDH v mechu	Podobnost k mechu (%)	Nejpodobnější ALDH v kukuřici	Nejpodobnější ALDH v Arabidopsis	Podobnost k monokotům /dikotům (%)
Rodina 2	<i>ALDH2B1</i>	PpALDH2B	61	ZmALDH2-2	AtALDH2B4	58 / 64
	<i>ALDH2B2</i>	PpALDH2B	62	ZmALDH2-2	AtALDH2B4	61 / 60
Rodina 5	<i>ALDH5F1</i>	PpALDH5F2	58	ZmALDH5	AtALDH5F1	57 / 55
Rodina 6	<i>ALDH6B1</i>	PpALDH6A	68	ZmALDH6	AtALDH6B2	69 / 67
Rodina 10	<i>ALDH10A1</i>	PpALDH10A	62	ZmALDH10-1B	AtALDH10A8	61 / 62
Rodina 11	<i>ALDH11A1</i>	PpALDH11D	68	ZmALDH11	AtALDH11A3	70 / 71
Rodina 12	<i>ALDH12A1</i>	PpALDH12A1	68	ZmALDH12	AtALDH12A1	65 / 63
Rodina 18	<i>ALDH18E11</i>	PpALDH18B1	54	ZmALDH18B1	AtALDH18B1	53 / 53
Rodina 21	<i>ALDH21A1</i>	PpALDH21	59	-	-	0 / 0
Rodina 22	<i>ALDH22B1</i>	-	0	ZmALDH22	AtALDH22A1	52 / 52

4.4 Geny ALDH přítomné v kapradině *Azolla filiculoides*

V kapradině *Azolla filiculoides* byly identifikovány všechny rostlinné rodiny ALDH, kromě rodin ALDH23 a ALDH24.

Tab. 6 Geny ALDH přítomné v kapradině *Azolla filiculoides*, jejíž genom byl prohledán skrze webové rozhraní projektu Ferbase. V tabulce jsou zaznačeny počty nukleotidů i aminokyselin po přepisu sekvence, jelikož jsem využíval tblastn i blastp.

Rodina ALDH	Název genu	FernBase protein ID	Počet nukleotidů	Počet aminokyselin
<i>Rodina 2</i>	<i>ALDH2B1</i>	Azfi_s0021.g015740	1533	510
	<i>ALDH2B2</i>	Azfi_s0083.g038897	1494	497
	<i>ALDH2C1</i>	Azfi_s0078.g038198	1398	465
	<i>ALDH2C2</i>	Azfi_s0049.g030778	1407	468
	<i>ALDH2F1</i>	Azfi_s0002.g004343	1440	480
<i>Rodina 3</i>	<i>ALDH3L1</i>	Azfi_s0011.g012726	1557	518
	<i>ALDH3L2</i>	Azfi_s0109.g045319	1374	457
	<i>ALDH3L3</i>	Azfi_s0587.g078563	1551	516
	<i>ALDH3L4</i>	Azfi_s0036.g025798	1278	425
	<i>ALDH3L5</i>	Azfi_s0636.g080483	1722	573
<i>Rodina 5</i>	<i>ALDH5H1</i>	Azfi_s0033.g025123	2019	672
	<i>ALDH5H2</i>	Azfi_s0002.g003146	2544	733
	<i>ALDH5H3</i>	Azfi_s0002.g003388	3168	636
<i>Rodina 6</i>	<i>ALDH6B1</i>	Azfi_s0012.g012829	1704	567
	<i>ALDH6C1</i>	Azfi_s0002.g003326	1728	576
<i>Rodina 7</i>	<i>ALDH7B1</i>	Azfi_s0068.g036288	1494	497
<i>Rodina 10</i>	<i>ALDH10A1</i>	Azfi_s0092.g043158	1395	464
<i>Rodina 11</i>	<i>ALDH11A1</i>	Azfi_s0160.g054025	1641	546
	<i>ALDH11A2</i>	Azfi_s0003.g007703	1377	458
	<i>ALDH11A3</i>	Azfi_s0006.g010360	2835	644
<i>Rodina 12</i>	<i>ALDH12A1</i>	Azfi_s0027.g023626	1716	571
<i>Rodina 18</i>	<i>ALDH18B1</i>	Azfi_s1111.g097729	2217	738
	<i>ALDH18B2</i>	Azfi_s0157.g053821	2265	754
	<i>ALDH18B3</i>	Azfi_s0022.g015914	2229	742
<i>Rodina 21</i>	<i>ALDH21A1</i>	Azfi_s0412.g068547	3423	1140
<i>Rodina 22</i>	<i>ALDH22A1</i>	Azfi_s0022.g016122	1101	366

Mnoho rodin obsahovalo velké množství *ALDH* genů. Rodiny ALDH2 a ALDH3 obsahovaly 5 zástupců, ALDH5, ALDH11 a ALDH18 obsahovaly 3 zástupce a rodina

ALDH6 byla zastoupena dvěma geny. Zbylé identifikované rodiny měly právě jednoho zástupce.

Tab. 7 Tabulka podobnosti ALDH sekvencí z *Azolla filiculoides* s mechem *Physcomyrella patens* (čepenka odstálá), kukuřicí *Zea mays* (kukuřice setá) a *Arabidopsis thaliana* (huseníček rolní). PpALDH značí proteinové sekvence z mechu *Physcomyrella patens*, ZmALDH značí proteiny z jednoděložné (monokot) kukuřice, AtALDH značí sekvence z dvouděložné (dikot) *Arabidopsis*. Značení „-“ znamená chybějící sekvence.

Rodina ALDH	Název genu	Nejpodobnější ALDH z mechu	Podobnost k mechu (%)	Nejpodobnější ALDH v kukuřici	Nejpodobnější ALDH v Arabidopsis	Podobnost k monokotům /dikotům (%)
Rodina 2	<i>ALDH2B1</i>	PpALDH2B	60	ZmALDH2-4	AtALDH2B4	62 /62
	<i>ALDH2B2</i>	PpALDH2B	57	ZmALDH2-4	AtALDH2C4	58 /56
	<i>ALDH2C1</i>	PpALDH2B	59	ZmALDH2-4	AtALDH2C4	63 /61
	<i>ALDH2C2</i>	PpALDH2B	57	ZmALDH2-3	AtALDH2C4	61 /57
	<i>ALDH2F1</i>	PpALDH2A	42	ZmALDH2-1	AtALDH2B4	43 /42
Rodina 3	<i>ALDH3L1</i>	PpALDH3B	51	ZmALDH3-1	AtALDH3H1	51 /50
	<i>ALDH3L2</i>	PpALDH3A	51	ZmALDH3-4	AtALDH3H1	49 /49
	<i>ALDH3L3</i>	PpALDH3B	49	ZmALDH3-1	AtALDH3H1	48 /48
	<i>ALDH3L4</i>	PpALDH3D	53	ZmALDH3-5	AtALDH3H1	52 /53
	<i>ALDH3L5</i>	PpALDH3B	41	ZmALDH3-1	AtALDH3H1	51 /49
Rodina 5	<i>ALDH5H1</i>	PpALDH5B	54	ZmALDH5	AtALDH5F1	54 /53
	<i>ALDH5H2</i>	PpALDH5A	54	ZmALDH5	AtALDH5F1	53 /53
	<i>ALDH5H3</i>	PpALDH5A	52	ZmALDH5	AtALDH5F1	51 /54
Rodina 6	<i>ALDH6B1</i>	PpALDH6B	78	ZmALDH6	AtALDH6B2	71 /71
	<i>ALDH6C1</i>	PpALDH6B	48	ZmALDH6	AtALDH6B2	48 /48
Rodina 7	<i>ALDH7B1</i>	PpALDH7B	74	ZmALDH7	AtALDH7B4	72 /73
Rodina 10	<i>ALDH10A1</i>	PpALDH10A	73	ZmAL10-1B	AtALDH10A9	73 /75
Rodina 11	<i>ALDH11A1</i>	PpALDH11B	70	ZmALDH11	AtALDH11A3	74 /75
	<i>ALDH11A2</i>	PpALDH11C	73	ZmALDH11	AtALDH11A3	78 /77
	<i>ALDH11A3</i>	PpALDH11D	72	ZmALDH11	AtALDH11A3	75 /75
Rodina 12	<i>ALDH12A1</i>	PpALDH12A1	70	ZmALDH12	AtALDH12A1	69 /67
Rodina 18	<i>ALDH18B1</i>	PpALDH18B1	69	ZmALDH18B1	AtALDH18B1	67 /66
	<i>ALDH18B2</i>	PpALDH18B1	71	ZmALDH18B1	AtALDH18B1	67 /67
	<i>ALDH18B3</i>	PpALDH18B1	70	ZmALDH18B1	AtALDH18B1	69 /68
Rodina 21	<i>ALDH21A1</i>	PpALDH21	57	-	-	0/0
Rodina 22	<i>ALDH22A1</i>	-	0	ZmALDH22	AtALDH22A1	65 /65

4.5 Geny ALDH přítomné v řase *Chromochloris zofingiensis*

V řase *Chromochloris zofingiensis* jsem identifikoval nejméně rodin rostlinných ALDH. Chyběly rodiny ALDH3, ALDH7, ALDH21, ALDH23 a ALDH24. Nalezené rodiny byly převážně zastoupeny pouze jedním genem, kromě rodiny ALDH2, která obsahovala dva zástupce.

Tab. 8 Geny ALDH přítomné v řase *Chromochloris zofingiensis*, jejíž genom byl prohledán skrze webovou databázi Phytozome. V tabulce jsou zaznačeny kromě ID těchto proteinů taktéž exony zaznamenané v této databázi.

Rodina ALDH	Název genu	Phytozome ID	Počet exonů	Počet aminokyselin
<i>Rodina 2</i>	<i>ALDH2B1</i>	Cz03g20090.t1	8	479
	<i>ALDH2F1</i>	Cz11g04290.t1	13	490
<i>Rodina 5</i>	<i>ALDH5H1</i>	Cz05g02090.t1	11	559
<i>Rodina 6</i>	<i>ALDH6B1</i>	Cz12g20220.t1	9	524
<i>Rodina 10</i>	<i>ALDH10A1</i>	Cz03g10210.t1	10	498
<i>Rodina 11</i>	<i>ALDH11A1</i>	Cz06g05120.t1	6	497
<i>Rodina 12</i>	<i>ALDH12A1</i>	Cz05g18170.t1	9	501
<i>Rodina 18</i>	<i>ALDH18E1</i>	Cz04g39260.t1	6	438
<i>Rodina 22</i>	<i>ALDH22B1</i>	Cz09g10170.t1	12	613

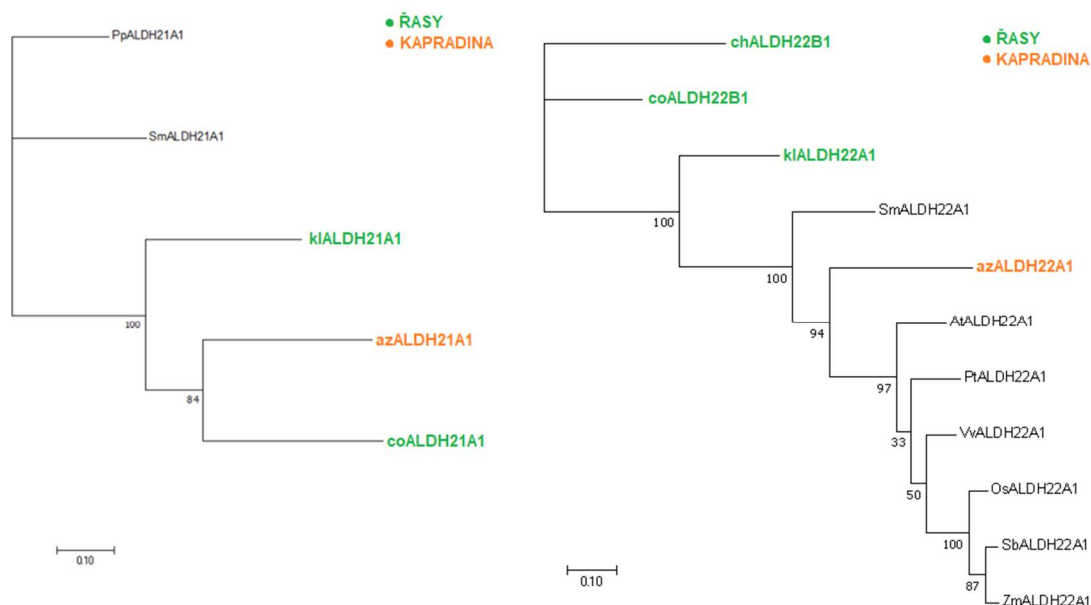
Tab. 9 Tabulka podobnosti sekvencí ALDH z *Chromochloris zofingiensis* s mechem *Physcomyrella patens* (čepenka odstálá), kukuřicí *Zea mays* (kukuřice setá) a *Arabidopsis thaliana* (huseníček rolní). PpALDH značí proteinové sekvence z mechu *Physcomyrella patens*, ZmALDH značí proteiny z jednoděložné (monokot) kukuřice, AtALDH značí sekvence z dvouděložné (dikot) *Arabidopsis*. Značení „-“ znamená chybějící sekvence.

Rodina ALDH	Název genu	Nejpodobnější ALDH v mechu	Podobnost k mechu (%)	Nejpodobnější ALDH v kukuřici	Nejpodobnější ALDH v Arabidopsis	Podobnost k monokotům /dikotům (%)
<i>Rodina 2</i>	<i>ALDH2B1</i> <i>ALDH2F1</i>	PpALDH2B1 PpALDH2B1	67 45	ZmALDH2-1 ZmALDH2-1	AtALDH2B4 AtALDH2B4	62 /65 45 /47
<i>Rodina 5</i>	<i>ALDH5H1</i>	PpALDH5B	53	ZmALDH5	AtALDH5F1	54 /55
<i>Rodina 6</i>	<i>ALDH6B1</i>	PpALDH6B	66	ZmALDH6	AtALDH6B2	61 /65
<i>Rodina 10</i>	<i>ALDH10A1</i>	PpALDH10A	63	ZmAL10-1A	AtALDH10A8	61 /60
<i>Rodina 11</i>	<i>ALDH11A1</i>	PpALDH11B	70	ZmALDH11	AtALDH11A3	72 /68
<i>Rodina 12</i>	<i>ALDH12A1</i>	PpALDH12A1	71	ZmALDH12	AtALDH12A1	69 /70
<i>Rodina 18</i>	<i>ALDH18E1</i>	PpALDH18B1	47	ZmALDH18B1	AtALDH18B1	44 /45
<i>Rodina 22</i>	<i>ALDH22B1</i>	-	0	ZmALDH22	AtALDH22A1	53 /52

4.6 Fylogenetická analýza

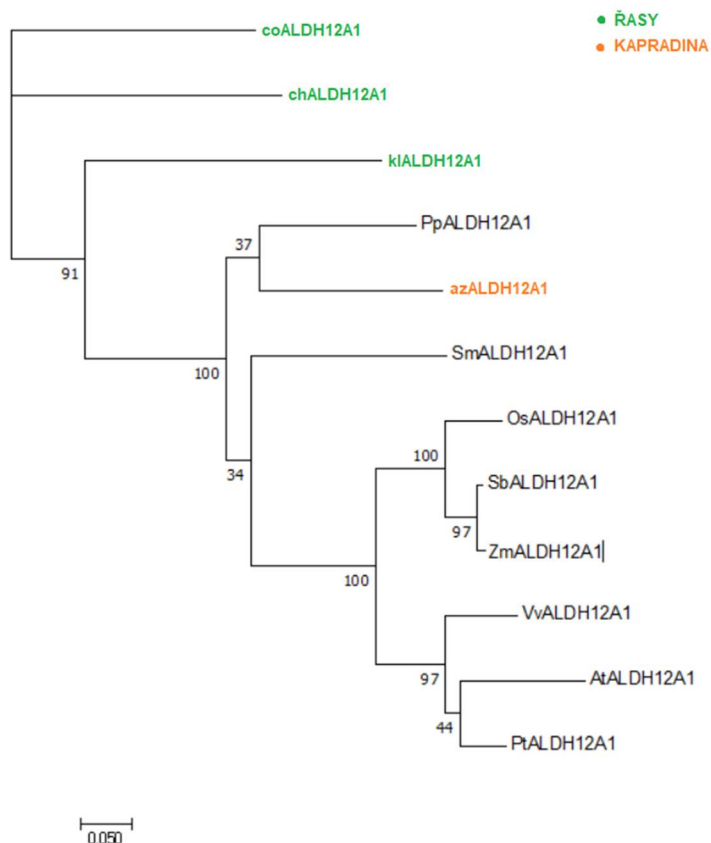
Fylogenetická analýza byla provedena pouze pro vybrané rodiny *ALDH* genů, a to konkrétně rodiny *ALDH2*, *ALDH5*, *ALDH7*, *ALDH10*, *ALDH12*, *ALDH21* a *ALDH22*. U každého stromu je uvedeno vzdálenostní měřítko, které značí míru odlišnosti jednotlivých sekvencí.

V rodině *ALDH21* se všichni nalezení zástupci řadili do jedné podrodiny na základě podobnosti, graf fylogenetického stromu tedy neznázorňuje přílišné odchylky. V rodině *ALDH22* se u řasy *Coccomyxa* a *Chromochloris* vyskytla nová podrodina *ALDH22B*, což můžeme pozorovat na fylogenetickém stromě této rodiny, jejich jednoznačným vychýlením vzhledem k ostatním zástupcům (Obr. 26).



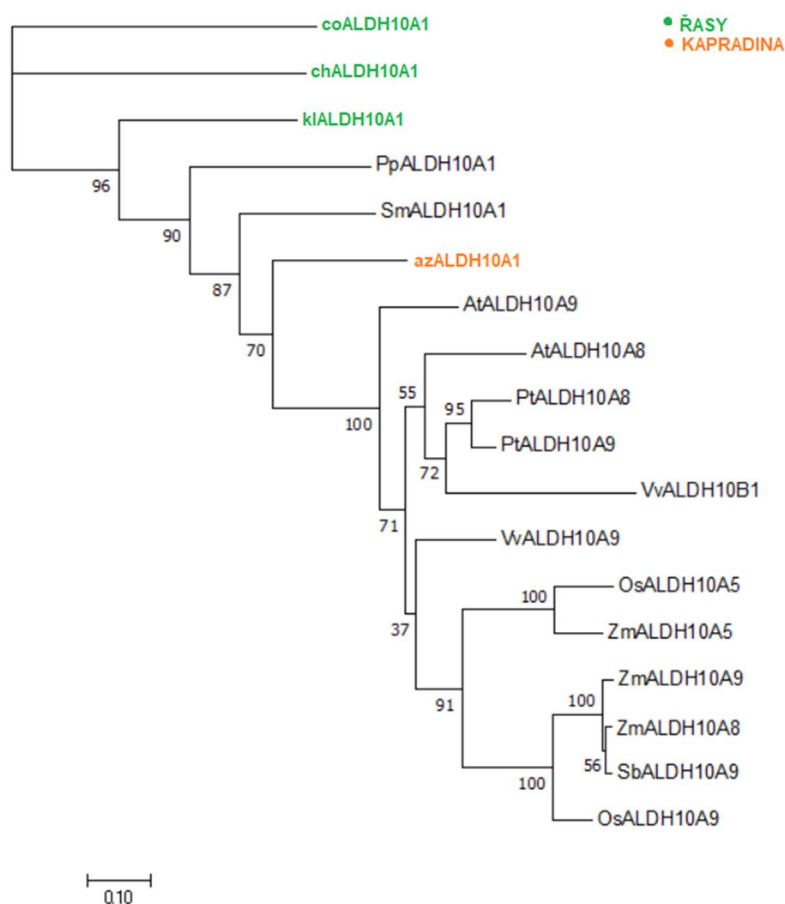
Obr. 26 Fylogenetické stromy pro rodiny *ALDH21* a *ALDH22* s ohodnocením větví pomocí bootstrappingu. Nalezené sekvence jsem porovnával se sekvencemi anotovanými v Brocker et al., 2013. Měřítko odpovídá odlišnosti sekvencí (měřítko 0,1 značí míru odlišnosti sekvencí 10%). Použité zkratky: Ch – *Chromochloris zoefingiensis*, Az – *Azolla filiculoides*, Co – *Coccomyxa subellipsoidea*, Kl – *Klebsormidium nitens*, At – *Arabidopsis thaliana* (Huseníček rolní), Zm – *Zea mays* (Kukuřice setá), Pp – *Physcomitrella patens*, Os – *Oriza sativa* (Rýže setá), Sm – *Selaginella moellendorffii*, Sb – *Sorghum bicolor* (Čirok dvoubarevný), Pt – *Populus trichocarpa* (Topol chlupatoplodý), Vv – *Vitis vinifera* (Vinná réva).

Pro rodinu ALDH12 jsou všechny sekvence značně konzervované, a tedy žádný *ALDH* gen příliš nevystupuje z fylogenetického stromu. Lze vidět například že monokoty (rýže, čirok, kukuřice) a dikoty (réva, huseníček, topol) se párují k sobě uvnitř stromu (Obr. 27).



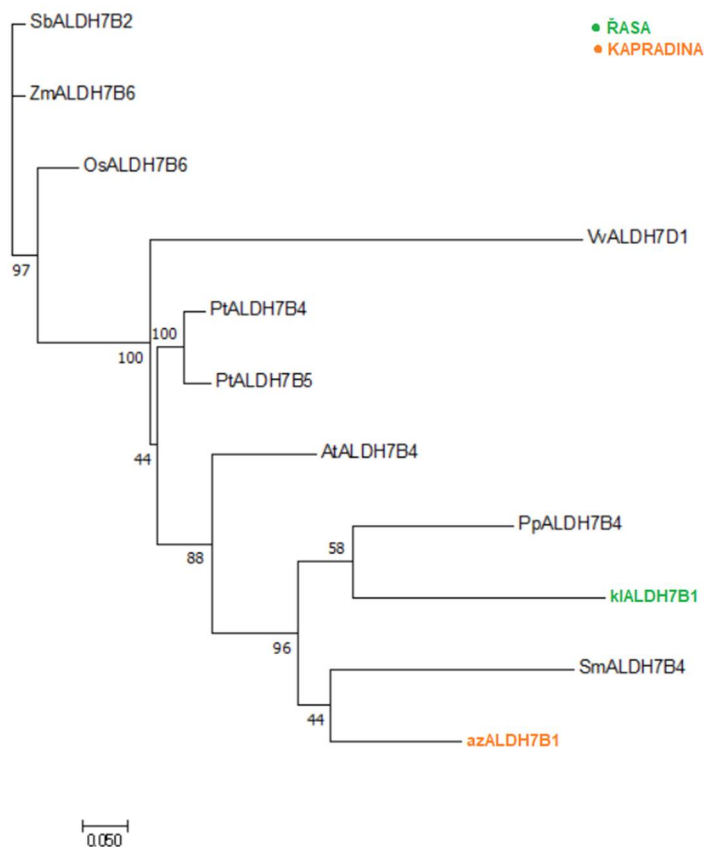
Obr. 27 Fylogenetický strom pro rodinu ALDH12 s ohodnocením větví pomocí bootstrappingu. Nalezené sekvence jsem porovnával se sekvencemi anotovanými v Brocker et al., 2013. Měřítko odpovídá odlišnosti sekvencí (měřítko 0,05 značí míru odlišnosti sekvencí 5%). Použité zkratky: Ch – *Chromochloris zofingiensis*, Az – *Azolla filiculoides*, Co – *Coccomyxa subellipsoidea*, Kl – *Klebsormidium nitens*, At – *Arabidopsis thaliana* (Huseníček rolní), Zm – *Zea mays* (Kukuřice setá), Pp – *Physcomitrella patens*, Os – *Oriza sativa* (Rýže setá), Sm – *Selaginella moellendorffii*, Sb – *Sorghum bicolor* (Čirok dvoubarevný), Pt – *Populus trichocarpa* (Topol chlupatoplodý), Vv – *Vitis vinifera* (Vinná réva).

Sekvence rodiny ALDH10 byly také značně konzervované a řadili se většina do podrodiny ALDH10A, a tedy ve fylogenetickém stromě žádný gen *ALDH* příliš nevystupuje, kromě zástupce z vinné révy řadícího se do podrodiny ALDH10B (Obr. 28).



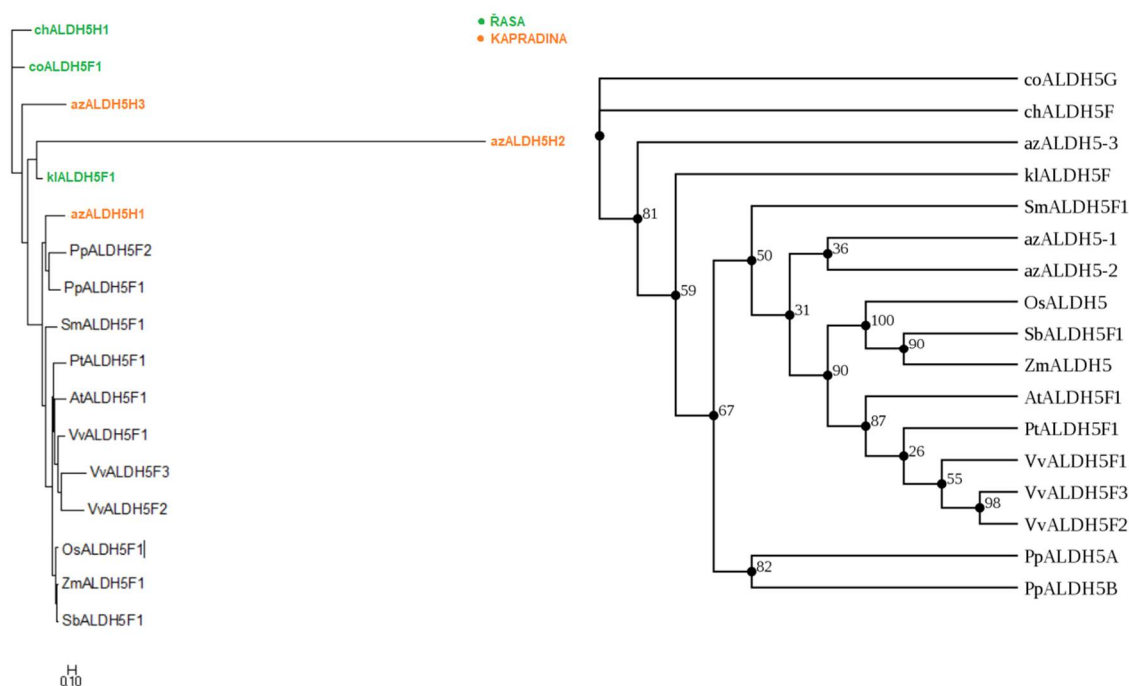
Obr. 28 Fylogenetický strom pro rodinu ALDH10 s ohodnocením větví pomocí bootstrappingu. Nalezené sekvence jsem porovnával se sekvencemi anotovanými v Brocker et al., 2013. Měřítko odpovídá odlišnosti sekvencí (měřítko 0,1 značí míru odlišnosti sekvencí 10%). Použité zkratky: Ch – *Chromochloris zoŕingiensis*, Az – *Azolla filiculoides*, Co – *Coccomyxa subellipsoidea*, Kl – *Klebsormidium nitens*, At – *Arabidopsis thaliana* (Huseníček rolní), Zm – *Zea mays* (Kukuřice setá), Pp – *Physcomitrella patens*, Os – *Oriza sativa* (Rýže setá), Sm – *Selaginella moellendorffii*, Sb – *Sorghum bicolor* (Čirok dvoubarevný), Pt – *Populus trichocarpa* (Topol chlupatoplodý), Vv – *Vitis vinifera* (Vinná réva).

Rodina ALDH7 je velmi konzervovaná, a tedy se většina sekvencí řadí do jediné podrodiny ALDH7B. Na fylogenetickém stromě můžeme pozorovat jednoho zástupce podrodiny ALDH7D z vinné révy, který se jednoznačně liší od ostatních (Obr. 29).



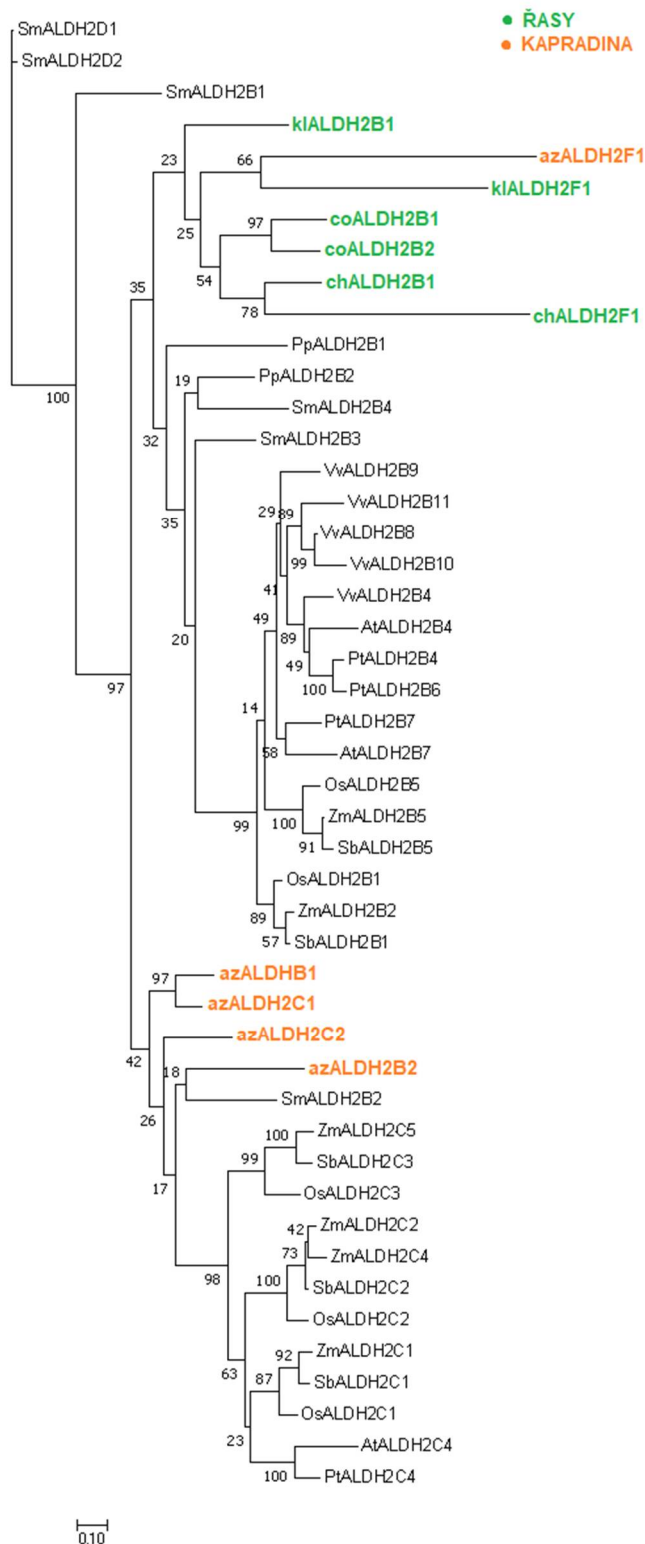
Obr. 29 Fylogenetický strom pro rodinu ALDH7 s ohodnocením větví pomocí bootstrappingu. Nalezené sekvence jsem porovnával se sekvencemi anotovanými v Brocker et al., 2013. Měřítko odpovídá odlišnosti sekvencí (měřítko 0,05 značí míru odlišnosti sekvencí 5%). Použité zkratky: Az – *Azolla filiculoides*, Kl – *Klebsormidium nitens*, At – *Arabidopsis thaliana* (Huseníček rolní), Zm – *Zea mays* (Kukuřice setá), Pp – *Physcomitrella patens*, Os – *Oriza sativa* (Rýže setá), Sm - *Selaginella moellendorffii*, Sb – *Sorghum bicolor* (Čirok dvoubarevný), Pt - *Populus trichocarpa* (Topol chlupatoplodý), Vv – *Vitis vinifera* (Vinná réva).

Fylogenetický strom pro rodinu ALDH5 znázorňuje vysokou podobnost většiny zástupců této rodiny, jeden zástupce z kapradiny *Azolla* však vykazuje značnou odlišnost oproti ostatním sekvencím, a tedy vystupuje z fylogenetického stromu (Obr. 30). Jednoznačně tedy můžeme říct, že se jedná o zástupce odlišné podrodiny, než u ostatních sekvencí.



Obr. 30 Fylogenetický strom pro rodinu ALDH5 (vpravo s ohodnocením větví bootstrapingem). Nalezené sekvence jsem porovnával se sekvencemi anotovanými v Brocker et al. (2013). Měřítko odpovídá odlišnosti sekvencí (měřítko 0,1 značí míru odlišnosti sekvencí 10%). Použité zkratky: Ch – *Chromochloris zofingiensis*, Az – *Azolla filiculoides*, Co - *Coccomyxa subellipsoidea*, Kl - *Klebsormidium nitens*, At – *Arabidopsis thaliana* (Huseniček rolní), Zm – *Zea mays* (Kukuřice setá), Pp – *Physcomitrella patens*, Os – *Oriza sativa* (Rýže setá), Sm - *Selaginella moellendorffii*, Sb – *Sorghum bicolor* (Čirok dvoubarevný), Pt - *Populus trichocarpa* (Topol chlupatoplodý), Vv – *Vitis vinifera* (Vinná réva).

Nakonec rostlinná rodina ALDH2, jejíž zástupci se vyskytují v hojném počtu u suchozemských rostlin se dělí na hlavní podrodiny ALDH2B, ALDH2C a ALDH2D. Některé z nalezených sekvencí však nespádaly do žádné z těchto rodin, kvůli své nízké identitě. Řadily se tedy do nové podrodiny ALDH2F, která je jasně vidět na fylogenetickém stromě (Obr. 31).



Obr. 31 **Fylogenetický strom pro rodinu ALDH2 s ohodnocením větví pomocí bootstrappingu.** Nalezené sekvence jsem porovnával se sekvencemi anotovanými v Brocker et al. (2013). Měřítka odpovídá odlišnosti sekvencí (měřítka 0,1 značí míru odlišnosti sekvencí 10%). Použité zkratky: Ch – *Chromochloris zofingiensis*, Az – *Azolla filiculoides*, Co – *Coccomyxa subellipsoidea*, Kl – *Klebsormidium nitens*, At – *Arabidopsis thaliana* (Huseniček rolní), Zm – *Zea mays* (Kukuřice setá), Pp – *Physcomitrella patens*, Os – *Oriza sativa* (Rýže setá), Sm – *Selaginella moellendorffii*, Sb – *Sorghum bicolor* (Čirok dvoubarevný), Pt – *Populus trichocarpa* (Topol chlupatoplodý), Vv – *Vitis vinifera* (Vinná réva).

5 DISKUZE

Ve zkoumaných řasách jsem identifikoval vždy dva zástupce rodiny ALDH2. U *Coccomyxa* se jednalo o mitochondriální ALDH2B, metabolizující acetaldehydy, avšak u zbylých dvou řas byla pouze jedna z nalezených sekvencí podobná z více jak 60% referenční ALDH2B, a tedy jsou zde ALDH2F, které se svou odlišností tvoří novou podrodinu, nejspíše tedy došlo k diversifikaci sekvencí ALDH. To mohlo být způsobeno adaptací na specifický substrát a tato funkce pak nejspíše již nebyla potřeba u většiny suchozemských rostlin, a tedy tato rodina vymizela. U kapradiny bylo nalezeno více, a to 5 sekvencí ALDH2, což je typické pro suchozemské rostliny. Byla nalezena mitochondriální ALDH2B i cytosolická ALDH2C podrodina, ale také jedna nepříliš podobná sekvence ALDH řadící se tedy do nové podrodiny ALDH2F.

Podle počtu exonů, který jsem zjistil pouze u řas *Chromochloris* a *Coccomyxa*, jelikož jsem jejich genom zkoumal skrze server Phytozome, můžeme vidět, že *Coccomyxa* obsahuje více exonů v sekvenci rodiny ALDH2B než vyšší rostliny jako *Arabidopsis*, *Physcomitrella* nebo kukuřice, které mají zástupce rodiny ALDH2B maximálně o 12 exonech. *Chromochloris* však obsahuje pouze 8 exonů v této sekvenci, což je poměrně běžné.

Rodina ALDH3 je velmi evolučně zachovalá rodina, velmi rozšířená a diverzifikovaná u suchozemských rostlin. Bylo identifikováno 5 genů u kapradiny *Azolla*, neměly však vysokou identitu s žádnou doposud identifikovanou rodinou ALDH3, jsou tedy řazeny do nové rodiny ALDH3L. Co se týče řas, byl identifikován gen *ALDH3* pouze u řasy *Klebsormidium*, u zbylých dvou zástupců geny této rodiny chyběly. Toto zjištění naznačuje, že teorie, že tato rodina expandovala a diverzifikovala, až během adaptace vodních rostlin na souš (Brocker et al., 2013), bude nejspíše pravdivá.

V každém zkoumaném zástupci se pak vyskytovala alespoň jedna sekvence *ALDH5*, což není překvapením, vzhledem k tomu, že hrají důležitou roli v metabolismu GABA, tzv. „GABA shunt“. V řase *Klebsormidium* a *Coccomyxa* měli zástupci této rodiny vysokou identitu k ostatním známým, a tedy byli zařazeni do rodiny ALDH5F. U řasy *Chromochloris* a kapradiny *Azolla* nebyla identita se známými zástupci příliš vysoká a řadím je tedy do nové rodiny ALDH5H.

Na druhou stranu počet exonů *Chromochloris* a *Coccomyxy* byl téměř stejný, ačkoli se liší v podrodinách. V obou případech pak byl značně nižší, než je obvyklé u

ortologů známých z vyšších rostlin, které téměř pro každou sekvenci z rodiny ALDH5 obsahovali kolem 20 exonů. Můžeme tedy říct, že s největší pravděpodobností mají řasy delší exonové sekvence ALDH5 než vyšší rostliny.

Všechny zkoumané rostliny obsahovaly alespoň jednoho zástupce rodiny ALDH6. Tito zástupci měli vysokou identitu s referenčními sekvencemi z vyšších rostlin (větší než 60 %) a spadají tedy do stejné podrodiny ALDH6B. Genom kapradiny obsahoval 2 sekvence ALDH6, avšak druhá sekvence měla identitu pouze asi 50 % a neřadí se tedy do stejné podrodiny, řadím tuto rodinu tedy do nové podrodiny ALDH6C.

Počet exonů vyšších rostlin rodiny ALDH6 se značně liší u jednotlivých zástupců, od *Arabidopsis* s 19 exony až po *Physcomytrinu* se svými 4 exony. Počet exonů ALDH6 řas se nejvíce přiblížil exonům v kukuřici, u které byla nalezena sekvence o 13 exonech.

Rodina ALDH7 je velmi konzervovaná stejně jako mnoho dalších rodin ALDH v rostlinách, proto bylo překvapením, že v řase *Coccomyxa* a *Chromochloris* nebyla identifikována žádná sekvence spadající do této rodiny. Podobně nebyly tyto ALDH nalezeny ani například v *Chlamydomonas reinhardtii* zkoumané v předešlé publikované studii, kdy nebylo jasné, zda tyto geny zmizely příčinou evolučního tlaku, či delece genů v těchto zástupcích (Brocker et al., 2013). Nalezené ALDH7 v kapradině a *Klebsormidiu* však měly obě velmi vysokou identitu, a to více jak 70% a tedy se jednoznačně řadí do známé rodiny ALDH7B.

Většina identifikovaných sekvencí spadajících do rodiny ALDH10 má vysokou míru homologie sekvencí, a tedy je řadíme do stejné podrodiny, ALDH10A. V genomu každé ze zkoumaných rostlin byla nalezena právě jedna sekvence ALDH10. U vyšších rostlin je počet exonů poměrně uniformní, většina obsahuje sekvence o 15 exonech. V mnou zkoumaných řasách bylo identifikováno 8 a 10 exonů což je nejspíše typické pro řasy, jelikož můžeme pozorovat například 11 exonů u řasy *Chlamydomonas reinhardtii* z předchozí studie (Brocker et al., 2013).

Zástupci rodiny ALDH11 byli identifikováni ve všech zkoumaných rostlinách, každý s vysokou homologií sekvence (podobnost >70 %) a všichni se řadí do stejné podrodiny, ALDH11A, podobně jako u téměř všech předešle zkoumaných rostlin. Počet exonů řasy *Coccomyxa* je značně vysoký vzhledem ke všem ostatním známým zástupcům této rodiny, které doposud dosahovaly maximálně 11 exonů, může se tedy

jednat o to, že exony této řasy jsou kratší než u ostatních rostlin, nebo je sekvence řasy *Coccomyxa* něčím specifická.

Rodina ALDH12 je silně konzervovaná, v každé rostlině se vyskytoval právě jeden gen z této rodiny s podobností >60 %, všechny tedy řadíme do podrodiny ALDH12A. Vzhledem k této míře podobnosti sekvencí, existuje nejspíše silný evoluční tlak pro to, aby sekvence této rodiny zůstala stále stejná. Počet exonů řas víceméně odpovídá známému zástupci *Chlamydomonas reinhartii*, vyšší rostliny obsahují většinou více, a to kolem 15 exonů v sekvenci ALDH12.

ALDH18 jsou silně konzervované, a tedy přítomné taktéž v každé ze zkoumaných rostlin. Sekvence z kapradiny *Azolla* a řasy *Klebsormidium* sdílí podobnost větší než 60% s referenčními sekvencemi. U zbylých řas tomu tak však není a jejich podobnost je přibližně 50%. Proto se rozdělují na dvě podrodiny, známou ALDH18B a novou rodinu ALDH18E. Většina vyšších rostlin obsahuje v sekvencích ALDH18 kolem 20 exonů a ačkoli řasa *Chlamydomonas reinhartii* obsahuje méně a to 16, tyto počty se neshodují s výsledky z *Chromochloris*, která obsahovala pouze 6 exonů v sekvenci ALDH18.

Rodina ALDH21 byla nalezena téměř u každé zkoumané rostliny, což není překvapivé, jelikož je známo, že se tato rodina vyskytuje převážně u nižších rostlin, a tedy výsledky tento trend nevyvracejí. Podobně se hovoří o rodině ALDH23, ta však byla identifikována pouze u zástupce *Klebsormidium nitens* (Tab. 2). Počet exonů zástupce rodiny ALDH21 nalezený v *Coccomyxa* byl 10 exonů, což je podobné jako u řasy *Chlamydomonas reinhartii* nebo plavuně *Selaginella*, kde byly identifikováno 8 exonů.

Překvapivý je nález rodiny ALDH22 ve všech zástupcích, jelikož se u této rodiny doposud předpokládalo, že se vyskytuje pouze u vyšších rostlin, a tedy by se v řasách neměla vyskytovat. Výsledky této práce však svědčí o opaku. O této rodině není příliš známo, ale co se týče počtu exonů, většina zástupců této rodiny u vyšších rostlin obsahovala 14 exonů. V této práci zkoumané řasy obsahovali odlišné podrodiny ALDH22 a to ALDH22B, které obsahovaly 12 exonů u *Coccomyxy* a 9 exonů u *Chromochloris*.

6 ZÁVĚR

Identifikoval jsem 13 genů ALDH v genomu řasy *Klebsormidium nitens*, dále 10 genů ALDH v genomu řasy *Coccomyxa subellipsoidea* a pouze 9 genů ALDH v řase *Chromochloris zofingiensis*. Naproti tomu v genomu kapradiny *Azolla filiculoides* bylo identifikováno 26 genů ALDH. Můžeme tedy pozorovat značný nárůst počtu ALDH genů u suchozemských rostlin oproti vodním rostlinám jako jsou například řasy. Speciálně rodiny ALDH2 a ALDH3 expandovaly značně s přechodem rostlin na souš, což můžeme pozorovat jak u *Azolla filiculoides* tak i u v minulosti zkoumaného mechu *Physcomytrella Patens* (Brocker et al. 2013).

Mimo identifikace jednotlivých zástupců rodin ALDH, byly tyto zástupci také rozděleni do rodin a podrodin, přičemž jsem pomocí fylogenetické analýzy byl schopen určit nové podrodiny, které neměly dostačující identitu, aby náležely do již známých podrodin. Tyto rodiny byly pojmenovány ALDH2F, ALDH3L, ALDH5H, ALDH6C a ALDH22B.

Jedná se pouze o část celkové analýzy. Pokud bych měl zjistit další detaily, musel bych ověřit subcelulární lokalizaci, nalézt krystalovou strukturu každé rodiny a pro ty co doposud nemají tuto strukturu zjištěnou, bych musel vytvořit modely těchto struktur. Dále bych mohl pomocí softwaru weblogo zjistit konzervovanost aktivního místa, tyto výsledky bych následně koreloval s výsledky z nových podrodin.

7 Literatura

Allakhverdiev SI, Kreslavski VD, Klimov VV, Los DA, Carpentier R, Mohanty P (2008) Heat stress: an overview of molecular responses in photosynthesis. *Photosynth Res* **98**:541–550

Altschul, S; Gish, W; Miller, W; Myers, E; Lipman, D. Basic local alignment search tool. *Journal of Molecular Biology*. 1990, s. 403–410.

Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, et al. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol*. 2012; 13: R39 10.1186/gb-2012-13-5-r39

Bouche N, Fait A, Bouchez D, Moller SG, Fromm H (2003) Mitochondrial succinic-semialdehyde dehydrogenase of the gamma-aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proc Nat Acad Sci USA* **100**:6843–6848

Brocker C, Lassen N, Estey T, Pappa A, Cantore M, Orlova VV, Chavakis T, Kavanagh KL, Oppermann U, Vasiliou V (2010) Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *J Biol Chem* **285**:18452–18463

Brocker C., Vasiliou M., Carpenter S., Carpenter C., Zhang Y., Wang X., Kotchoni S.O., Wood A.J., Kirch H.H., Kopečný D. Nebert D.W. Vasiliou V. (2013) Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics. *Planta* **237**, 189–210.

Darienko, Tatyana; Gustavs, Lydia; Eggert, Anja; Wolf, Wiebke; Pröschold, Thomas (2015-06-16). "Evaluating the Species Boundaries of Green Microalgae (Coccomyxa, Trebouxiophyceae, Chlorophyta) Using Integrative Taxonomy and DNA Barcoding with Further Implications for the Species Identification in Environmental Samples". *PLOS ONE*. **10** (6): e0127838. doi:10.1371/journal.pone.0127838

Dayhoff, M.O., Schwartz, R. and Orcutt, B.C. (1978). "A model of Evolutionary Change in Proteins". *Atlas of protein sequence and structure (volume 5, supplement 3 ed.)*. Nat. Biomed. Res. Found. pp. 345–358. ISBN 978-0-912466-07-1.

Edgar RC (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Research*. **32** (5): 1792–97. doi:10.1093/nar/gkh340

Edgar RC (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity". *BMC Bioinformatics*. **5** (1): 113. doi:10.1186/1471-2105-5-113

Fait A, Fromm H, Walter D, Galili G, Fernie AR (2008) Highway or byway: the metabolic role of the GABA shunt in plants. *Trends Plant Sci* **13**:14–19

Figueroa-Martinez, F.; Nedelcu, A. M.; Smith, D. R.; Reyes-Prieto, A. (2015). "When the lights go out: the evolutionary fate of free-living colorless green algae". *New Phytologist*. **206** (3): 972–982. doi:10.1111/nph.13279

Floyd S.K., Bowman J.L. (2007) The ancestral developmental tool kit of land plants. *International Journal of Plant Sciences* **168**, 1-35.

Frenkel J, Vyverman W, Pohnert G (2014). "Pheromone signaling during sexual reproduction in algae". *Plant J*. **79** (4): 632–644. doi:10.1111/tpj.12496

Fritsch, F.E. (1935). *The Structure and Reproduction of the Algae, vol I*. Cambridge University Press. pp. 205–206.

Gao C, Han B (2009) Evolutionary and expression study of the aldehyde dehydrogenase (ALDH) gene superfamily in rice (*Oryza sativa*). *Gene* **431**:86–94

Gao Z, Loescher WH (2000) NADPH supply and mannitol biosynthesis. Characterization, cloning, and regulation of the nonreversible glyceraldehyde-3-phosphate dehydrogenase in celery leaves. *Plant Physiol* **124**:321–330

Grabber JH, Ralph J, Hatfield RD (2000) Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. *J Agric Food Chem* **48**:6106–6113

Grube, Martin; Seckbach, Joseph; Muggia, Lucia (2017). *Algal and cyanobacteria symbioses*. Grube, Martin, Seckbach, J. (Joseph), Muggia, Lucia. New Jersey. ISBN 978-1786340573

Henikoff, S.; Henikoff, J.G. (1992). Amino Acid Substitution Matrices from Protein Blocks. *PNAS*. **89** (22): 10915–10919. doi:10.1073/pnas.89.22.10915

Hidalgo, O. et al. Is there an upper limit to genome size? *Trends Plant Sci.* **22**, 567–573 (2017).

Huang W, Ma X, Wang Q, Gao Y, Xue Y, Niu X, Yu G, Liu Y (2008) Significant improvement of stress tolerance in tobacco plants by overexpressing a stress-responsive aldehyde dehydrogenase gene from maize (*Zea mays*). *Plant Mol Biol* **68**:451–463

Christenhusz, M. J. M.; Byng, J. W. (2016). "The number of known plants species in the world and its annual increase". *Phytotaxa*. Magnolia Press. **261** (3): 201–217. doi:10.11646/phytotaxa.261.3.1

Jelínek, Jan; Zicháček, Vladimír. *Biologie pro gymnázia*. Nakladatelství Olomouc, s.r.o., 2011. ISBN 978-80-7182-213-4.

Jimenez-Lopez JC, Gachomo EW, Seufferheld MJ, Kotchoni SO (2010) The maize ALDH protein superfamily: linking structural features to functional specificities. *BMC Struct Biol* **10**:43

Katoh, Kazutaka; Misawa, Kazuharu; Kuma, Kei-ichi; Miyata, Takashi (2002). "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform". *Nucleic Acids Research*. **30** (14): 3059–66. doi:10.1093/nar/gkf436

Kirch HH, Nair A, Bartels D (2001) Novel ABA- and dehydrationinducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*. *Plant J Cell Mol Biol* **28**:555–567

Kirch HH, Bartels D, Wei Y, Schnable PS, Wood AJ (2004) The ALDH gene superfamily of *Arabidopsis*. *Trends Plant Sci* **9**:371–377

Končítíková R., Vigoroux A., Kopečná M., Andree T., Bartoš J., Šebela M., Moréra S., Kopečný D., 2015. Role and structural characterization of plant aldehyde dehydrogenases from family 2 and family 7. *Biochemical Journal* **468**, 109–123.

Kopečná M. et al. (2017) The ALDH21 gene found in lower plants and some vascular plants codes for a NADP⁺-dependent succinic semialdehyde dehydrogenase, doi: 10.1111/tpj.13648.

D. A. Korasick, R. Končítíková, M. Kopečná, et al., (2019) Structural and Biochemical Characterization of Aldehyde Dehydrogenase 12, the Last Enzyme of Proline Catabolism in Plants, *Journal of Molecular Biology*, <https://doi.org/10.1016/j.jmb.2018.12.010>

Kotchoni SO, Kuhns C, Ditzer A, Kirch HH, Bartels D (2006) Overexpression of different aldehyde dehydrogenase genes in *Arabidopsis thaliana* confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environ* **29**:1033–1048

Leliaert, Frederik; Smith, David R.; Moreau, Hervé; Herron, Matthew D.; Verbruggen, Heroen; Delwiche, Charles F.; De Clerck, Olivier (2012). "Phylogeny and molecular evolution of the green algae" (PDF). *Critical Reviews in Plant Sciences*. **31**: 1–46. doi:10.1080/07352689.2011.615705

Lhotský, Josef. *Úvod do studia symbiotických interakcí. Nový pohled na viry a bakterie*. Praha: Academia, 2015. 208 s. ISBN 978-80-200-2480-0. S. 121.

Li, Fay-Wei; Brouwer, Paul; Carretero-Paulet, Lorenzo; Cheng, Shifeng; de Vries, Jan; Delaux, Pierre-Marc; Eily, Ariana; Koppers, Nils; Kuo, Li-Yaung; Li, Zheng; Simenc, Mathew; Small, Ian; Wafula, Eric; Angarita, Stephany; Barker, Michael S.; Bräutigam, Andrea; dePamphilis, Claude; Gould, Sven; Hosmani, Prashant S.; Huang, Yao-Moan; Huettel, Bruno; Kato, Yoichiro; Liu, Xin; Maere, Steven; McDowell, Rose; Mueller, Lukas A.; Nierop, Klaas G. J.; Rensing, Stefan A.; Robison, Tanner; Rothfels, Carl J.; Sigel, Erin M.; Song, Yue; Timilsena, Prakash R.; Van de Peer, Yves; Wang, Hongli; Wilhelmsson, Per K. I.; Wolf, Paul G.; Xu, Xun; Der, Joshua P.; Schluepmann, Henriette; Wong, Gane K.-S.; Pryer, Kathleen M. (2018/07/01) Fern genomes elucidate land plant evolution and cyanobacterial symbioses, *Nature Plants* **4**, 460-472. <https://doi.org/10.1038/s41477-018-0188-8>

Marchitti SA, Brocker C, Stagos D, Vasiliou V (2008) Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol* **4**:697–720

Mount, D. W. *Bioinformatics: Sequence and Genome Analysis*. 2nd. vyd. [s.l.]: Cold Spring Harbor Press, 2004. ISBN 978-0-8796-9712-9.

Nair RB, Bastress KL, Ruegger MO, Denault JW, Chapple C (2004) The *Arabidopsis thaliana* REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* **16**:544–554

Notredame C, Higgins DG, Heringa J (2000-09-08). "T-Coffee: A novel method for fast and accurate multiple sequence alignment". *J Mol Biol*. **302** (1): 205–217. doi:10.1006/jmbi.2000.4042

Obermayer, R., Leitch, I. J., Hanson, L. & Bennett, M. D. Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Ann. Bot.* **90**, 209–217 (2002).

op den Camp RG, Kuhlemeier C (1997) Aldehyde dehydrogenase in tobacco pollen. *Plant Mol Biol* **35**:355–365

Palanivelu R, Brass L, Edlund AF, Preuss D (2003) Pollen tube growth and guidance is regulated by POP2, an *Arabidopsis* gene that controls GABA levels. *Cell* **114**:47–59

PhyML models: <http://www.atgc-montpellier.fr/models/index.php?model=lg> (9.5.2019)

Plaxton WC (1996) The organization and regulation of plant glycolysis. *Annu Rev Plant Physiol Plant Mol Biol* **47**:185–214

Program Selection Tables of the Blast NCBI website:

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=ProgSelectionGuide (9.5.2019)

Rodrigues SM, Andrade MO, Gomes AP, Damatta FM, Baracat- Pereira MC, Fontes EP (2006) Arabidopsis and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. *J Exp Bot* **57**:1909–1918

Russian CELSS Studies: <https://www.permanent.com/russian-celss.html> (9.5.2019)

Sebela M, Brauner F, Radova A, Jacobsen S, Havlis J, Galuszka P, Pec P (2000) Characterisation of a homogeneous plant aminoaldehyde dehydrogenase. *Biochim Biophys Acta* **1480**:329–341

Shin JH, Kim SR, AnG(2009) Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol* **149**:905–915

Stiti N, Missihoun TD, Kotchoni SO, Kirch HH, Bartels D (2011) Aldehyde dehydrogenases in Arabidopsis thaliana: biochemical requirements, metabolic pathways, and functional analysis. *Frontiers Plant Sci* **2**:65

Stropek C., Krinninger S., Frank W. (2013) The moss *Physcomitrella patens*: methods and tools from cultivation to targeted analysis of gene function. *International Journal of Developmental Biology* **57**, 533-564.

Sunkar R, Bartels D, Kirch HH (2003) Overexpression of a stressinducible aldehyde dehydrogenase gene from Arabidopsis thaliana in transgenic plants improves stress tolerance. *Plant J Cell Molecular Biol* **35**:452–464

Valverde F, Losada M, Serrano A (1999) Engineering a central metabolic pathway: glycolysis with no net phosphorylation in an Escherichia coli gap mutant complemented with a plant GapN gene. *FEBS Lett* **449**:153–158

Vaz FM, Fouchier SW, Ofman R, Sommer M, Wanders RJ (2000) Molecular and biochemical characterization of rat gammatrimethylaminobutyraldehyde dehydrogenase and evidence for the involvement of human aldehyde dehydrogenase 9 in carnitine biosynthesis. *J Biol Chem* **275**:7390–7394

Wei Y, Lin M, Oliver DJ, Schnable PS (2009) The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of Arabidopsis. *BMC Biochem* **10**:7

Wong J.W.Y., Chan C.L., Tang W.K., Cheng C.H.K., Fong W.P., 2010. Is Antiquitin a Mitochondrial Enzyme? *Journal of Cellular Biochemistry* **109**, 74–81.

Xiong, J. Essential Bioinformatics. [s.l.]: Cambridge University Press, 2006. ISBN 978-0-521-60082-8.

Yoshiba Y, Kiyosue T, Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (1997) Regulation of levels of proline as an osmolyte in plants under water stress. *Plant Cell Physiol* **38**:1095–1102