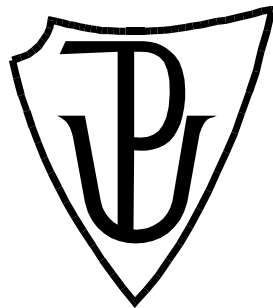


# UNIVERZITA PALACKÉHO V OLMOUCI

Přírodovědecká fakulta

Katedra biochemie



## **Studium nadrodiny ALDH v dostupných rostlinných genomech a tvorba webového rozhraní**

### **DIPLOMOVÁ PRÁCE**

|                   |                                  |
|-------------------|----------------------------------|
| Autor:            | <b>Šimon Pavlů</b>               |
| Studijní program: | B1406 Biochemie                  |
| Studijní obor:    | Bioinformatika                   |
| Forma studia:     | Prezenční                        |
| Vedoucí práce:    | <b>Mgr. David Kopečný, Ph.D.</b> |
| Konzultant:       | <b>Mgr. Martin Raus, Ph.D.</b>   |
| Rok:              | 2021                             |

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním diplomové práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl jsem seznámen 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 diplomanta*

Poděkování:

Chtěl bych poděkovat vedoucímu své diplomové práce Mgr. Davidu Kopečnému PhD., za odborné vedení, čas, ochotu a jeho cenné rady během psaní této diplomové práce. Dále bych chtěl také poděkovat Mgr. Martinu Rausovi PhD., za jeho všestrannou pomoc při realizaci praktické i teoretické části diplomové práce.

## Bibliografická identifikace

|                         |   |
|-------------------------|---|
| Jméno a příjmení autora | Šimon Pavlů   |
| Název práce             | Studium nadrodiny ALDH v dostupných rostlinných genomech a tvorba webového rozhraní |
| Typ práce               | Diplomová   |
| Pracoviště              | Katedra biochemie   |
| Vedoucí práce           | Mgr. David Kopečný, Ph.D.   |
| Rok obhajoby práce      | 2021  |

### Abstrakt

Tato diplomová práce se zabývá analýzou sekvencí aldehyddehydrogenas (ALDH) v dostupných rostlinných genomech od nižších po vyšší rostliny a tvorbou webového rozhraní s funkcí identifikace neznámých zástupců ALDH nadrodiny. Teoretická část je zaměřena na rešerši současných poznatků o nadrodině rostlinných a živočišných aldehyddehydrogenas. Výsledková část zahrnuje vyhledávání a identifikaci *ALDH* genů v genomech řas, mechů a kapradin, ale i jednoděložných či dvouděložných rostlin, pomocí softwaru BLAST, které pak tvoří databázi nové aplikace pro identifikaci a zařazení neznámých rostlinných ALDH sekvencí. Ty byly rozřazeny dle schválené klasifikace do rodin a podrodin a byla provedena fylogenetická analýza příbuznosti sekvencí, pro ověření přesnosti daného rozřazení, pomocí softwaru PhyML. Jsou popsány také možnosti užití a funkčnost vytvořeného webového rozhraní. Bylo zjištěno průměrné zastoupení jednotlivých ALDH rodin v rozličných odděleních a třídách a také existence dodatečných podrodin, které nebyly doposud uvedeny.

|               |  |
|---------------|--|
| Klíčová slova | aldehyddehydrogenasa, webová aplikace, BLAST, PhyML, metabolismus aldehydů, fylogeneze |
| Počet stran   | 95   |
| Počet příloh  | 00   |
| Jazyk         | Český  |

## **Bibliographical identification**

|                                |  |
|--------------------------------|--|
| Autor's first name and surname | Šimon Pavlů  |
| Title                          | Study of the ALDH superfamily in available plant genomes and creation of a web interface |
| Type of thesis                 | Diploma  |
| Department                     | Department of biochemistry   |
| Supervisor                     | Mgr. David Kopečný, Ph.D.  |
| The year of presentation       | 2021   |

### **Abstract**

This diploma thesis deals with the analysis of aldehyde dehydrogenase (ALDH) sequences in available plant genomes from lower to higher plants and the creation of a web interface with the function of identifying unknown representatives of the ALDH superfamily. The theoretical part is focused on the research of current knowledge about the superfamily of plant and animal aldehyde dehydrogenases. The results include BLAST search and identification of ALDH genes in the genomes of algae, mosses, and ferns, as well as monocotyledonous or dicotyledonous plants, which then form a database of the new application for identification and sorting of unknown plant ALDH sequences. These sequences were classified according to the approved nomenclature into families and subfamilies, and phylogenetic analysis of sequence relatedness was performed to verify the accuracy of the classification, using PhyML software. The possibilities of the usage and functionality of the created web interface are also described. The average occurrence of individual ALDH families in various divisions and classes was found as well as the existence of additional subfamilies that have not yet been listed.

|                      |   |
|----------------------|---|
| Keywords             | aldehyde dehydrogenase, web application, BLAST, PhyML, aldehyde metabolism, phylogeny |
| Number of pages      | 95  |
| Number of appendices | 00  |
| Language             | Czech   |

# OBSAH

|  |    |
|--|----|
| <b>1 Úvod</b>                                  | 1  |
| <b>2 Současný stav řešené problematiky</b>     | 2  |
| 2.1 Aldehyddehydrogenasy                       | 2  |
| 2.2 Rostlinné ALDH                             | 2  |
| 2.2.1 Rodina ALDH2                             | 3  |
| 2.2.2 Rodina ALDH3                             | 4  |
| 2.2.3 Rodina ALDH5                             | 4  |
| 2.2.4 Rodina ALDH6                             | 5  |
| 2.2.5 Rodina ALDH7                             | 6  |
| 2.2.6 Rodina ALDH10                            | 7  |
| 2.2.7 Rodina ALDH11                            | 8  |
| 2.2.8 Rodina ALDH12                            | 9  |
| 2.2.9 Rodina ALDH18                            | 10 |
| 2.2.10 Rodina ALDH19                           | 10 |
| 2.2.11 Rodina ALDH21                           | 11 |
| 2.2.12 Rodiny ALDH22, ALDH23 a ALDH24          | 12 |
| 2.3 Živočišné a jiné ALDH                      | 12 |
| 2.3.1 Rodina ALDH1                             | 12 |
| 2.3.2 Rodina ALDH4                             | 13 |
| 2.3.3 Rodina ALDH8                             | 14 |
| 2.3.4 Rodina ALDH9                             | 15 |
| 2.3.5 Rodiny ALDH13, ALDH14 a ALDH15           | 16 |
| 2.3.6 Rodina ALDH16                            | 16 |
| 2.3.7 Rodiny ALDH17 a ALDH20                   | 16 |
| <b>3 Experimentální část</b>                   | 18 |
| 3.1 Materiál                                   | 18 |
| 3.2 Metody                                     | 20 |
| 3.2.1 BLAST                                    | 20 |
| 3.2.1.1 Program BLAST                          | 21 |
| 3.2.1.2 Verze programu BLAST                   | 22 |
| 3.2.1.3 Algoritmus BLAST                       | 24 |
| 3.2.1.4 Princip algoritmu                      | 24 |
| 3.2.1.5 Dynamické programování                 | 26 |
| 3.2.1.6 BLAST alternativy                      | 27 |
| 3.2.1.7 Využití BLASTu                         | 28 |
| 3.2.2 MUSCLE                                   | 28 |
| 3.2.2.1 MUSCLE alternativy                     | 31 |
| 3.2.3 Gblocks                                  | 32 |
| 3.2.4 PhyML                                    | 33 |
| 3.2.5 NGPhylogeny                              | 35 |
| 3.2.6 Tvorba webového rozhraní                 | 36 |
| 3.2.6.1 PHP (PHP: Hypertext Preprocessor)      | 37 |
| 3.2.6.2 MySQL                                  | 37 |
| 3.2.6.3 JavaScript (JS)                        | 38 |
| 3.2.6.4 Ajax (Asynchronous JavaScript and XML) | 38 |

|                                      |    |
|--------------------------------------|----|
| <b>4 Výsledky</b>                    | 39 |
| 4.1 Vstup a funkce webového rozhraní | 39 |
| 4.2 Výstup programu                  | 40 |
| 4.3 Databáze                         | 43 |
| 4.4 Fylogenetická analýza            | 61 |
| <b>5 Diskuse</b>                     | 73 |
| <b>6 Závěr</b>                       | 77 |
| <b>7 Literatura</b>                  | 78 |

## **CÍLE PRÁCE**

- Vypracování literární rešerše na téma klasifikace rodiny ALDH v eukaryotních organismech
- Hledání ALDH genů v dostupných genomech rostlin od jednoduchých řas po vyšší rostliny zahrnující zástupce jednoděložných či dvouděložných
- Roztřídění genů do rodin a vytvoření databáze
- Fylogenetická analýza
- Tvorba webové aplikace s cílem správné identifikace neznámé rostlinné sekvence ALDH



## Seznam užitých zkratk

|         |  |
|---------|--|
| 10-FTHF | 10-formyltetrahydrofolát   |
| 2-AM    | 2-aminomukonát   |
| 2-AMS   | 2-aminomukonát semialdehyd   |
| 3-PG    | 3-fosfoglycerát  |
| AASAL   | $\alpha$ -aminoadipát semialdehyd  |
| ABAL    | 4-aminobutyraldehyd  |
| ACAPAL  | 3-acetamidopropionaldehyd  |
| AIC     | Akaike Information Criterion (metoda hodnocení statistických modelů)                                       |
| AJAX    | Asynchronous JavaScript and XML (webové technologie, využívané pro tvorbu asynchronních webových aplikací) |
| ALDH    | aldehyddehydrogenasa   |
| AMADH   | aminoaldehyddehydrogenasa  |
| AMSHD   | 2-aminomukonát-6-semialdehyddehydrogenasa  |
| APAL    | 3-aminopropionaldehyd  |
| API     | Application Programming Interface (rozhraní pro práci s datovými strukturami)                              |
| ATP     | Adenosintrifosfát  |
| BADH    | betainaldehyddehydrogenasa   |
| BAL     | $\beta$ -alanin  |
| BIC     | Bayesian information criterion (metoda hodnocení statistických modelů)                                     |
| BLAST   | Basic Local Alignment Search Tool (algoritmus pro porovnávání sekvencí)                                    |
| BLAT    | Blast Like Alignment Tool (alternativa programu BLAST)   |
| BLOCKS  | bioinformatická databáze obsahující multiple alignmenty konzervovaných částí proteinů                      |
| BLOSUM  | BLOcks SUBstitution Matrix (typ substituční matice)  |
| CLUSTAL | software pro mnohočetné porovnání sekvencí, či formát souboru sekvencí                                     |
| CoA     | koenzym A  |

|        |   |
|--------|---|
| CSS    | Cascading Style Sheets (technologie využívaná pro úpravu vzhledu webových stránek)                      |
| DMSP   | dimethylsulfonylpropionát   |
| DMSPAL | 3-dimethylsulfonylpropionaldehyd  |
| DOM    | Document Object Model (interface, který pracuje s HTML či XML dokumenty jako se stromovými strukturami) |
| DOPAL  | 3,4-dihydroxyfenylacetaldehyd   |
| FASTA  | algoritmus pro porovnávání sekvencí nebo též formát souboru sekvencí                                    |
| FDH    | 10-FTHFdehydrogenasa  |
| FPGA   | Field Programmable Gate Array (Programovatelné hradlové pole)   |
| GABA   | katabolismus $\gamma$ -aminobutyrate  |
| GAP    | glyceraldehyd-3-fosfát  |
| GAPN   | glyceraldehyd-3-fosfát dehydrogenasa  |
| GB     | glycin betain   |
| GBAL   | 4-guanidinobutyraldehyd   |
| GSAL   | glutamát $\gamma$ -semialdehyd  |
| GSALDH | glutamát- $\gamma$ -semialdehyddehydrogenasa  |
| HPRT1  | hypoxantin-guanin fosforibosyltransferasa   |
| HSP    | high-scoring segment pairs (párové segmenty sekvencí s vysokým skóre)                                   |
| HTML   | HyperText Markup Language (značkovací jazyk využíván pro tvorbu webových stránek)                       |
| IUPAC  | International Union of Pure and Applied Chemistry   |
| JS     | JavaScript (programovací jazyk)   |
| JSON   | JavaScript Object Notation (datový formát využíván JavaScriptem)  |
| MAFFT  | multiple alignment using fast Fourier transform (software pro multiple alignment)                       |
| MDA    | malondialdehyd  |
| MGC    | konsorciium Mammalian Gene Collection   |
| MMS    | methylmalonylsemialdehyd  |
| MMSDH  | methylmalonylsemialdehyddehydrogenasa   |
| MUSCLE | MUltiple Sequence Comparison by Log-Expectation   |

|                          |  |
|--------------------------|--|
| NAD <sup>+</sup> /NADH   | nikotinamidadenindinukleotid   |
| NADP <sup>+</sup> /NADPH | nikotinadenindinukleotidfosfát   |
| NCBI                     | The National Center for Biotechnology Information  |
| NIH                      | National Institutes of Health  |
| ORF                      | open reading frame (Otevřený čtecí rámeček)  |
| P5C                      | pyrrolin-5-karboxylát  |
| P5CDH                    | pyrrolin-5-karboxylát dehydrogenasa  |
| P5CS                     | Δ <sup>1</sup> -pyrrolin-5-karboxylátsynthasa  |
| P6CDH                    | Δ <sup>1</sup> -piperidin-6-karboxylátdehydrogenasa  |
| PAFF                     | Plant ALDH Family Finder (webové rozhraní)   |
| PAM                      | Point Accepted Mutation (typ substituční matice)   |
| PHP                      | programovací jazyk, též známý jako PHP: Hypertext Preprocessor   |
| PhyML                    | Phylogenetic estimation using Maximum Likelihood (software pro fylogenetickou analýzu)   |
| PHYLIP                   | formát souboru sekvencí  |
| PLAST                    | parallel local alignment search tool (alternativa programu BLAST)  |
| RA                       | kyselina retinová  |
| ROS                      | reactive oxygen species (reaktivní formy kyslíku)  |
| SQL                      | Structured Query Language (jazyk pro práci s databázemi)   |
| 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  |
| UPGMA                    | unweighted pair group method with arithmetic mean („shlukovací analýza“, nejjednodušší metoda konstrukce fylogenetických stromů) |
| XAMPP                    | soubor nástrojů pro vývoj webových aplikací (Apache HTTP Server, MySQL, PHP, Pearl)  |
| XML                      | Extensible Markup Language (značkovací jazyk, využíván mimo jiné pro výstup programu BLAST)                                      |

# 1 ÚVOD

Nadrodina aldehyddehydrogenas (ALDH) představuje rozmanitou skupinu enzymů podílejících se na metabolismu aldehydů endogenního i exogenního původu za pomoci  $\text{NAD}^+$  nebo  $\text{NADP}^+$  jako kofaktorů. V posledních letech probíhá na naší katedře v Olomouci výzkum zaměřený především na rostlinné ALDH, konkrétně 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 rostlinných genomů se objevuje stále více možností pro studium rostlinných ALDH genů a jejich evoluce od vodních rostlin jako jsou jednoduché řasy, až po vyšší suchozemské rostliny i vlivu těchto genů na schopnosti rostliny vypořádat se se stresovými podmínkami. Tato práce se zaměřuje na studium sekvencí ALDH v rozmanitých rostlinných zástupcích a tvorbu webového rozhraní, které uživateli umožní identifikovat neznámou sekvenci, pokud tato sekvence náleží do nadrodiny ALDH. Zařazení sekvencí v databázi našeho programu je pak podpořeno fylogenetickou analýzou.

## 2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

### 2.1 Aldehyddehydrogenasy

Nadrodina aldehyddehydrogenas (ALDH) představuje rozmanitou skupinu enzymů vyskytujících se jak u prokaryotických, tak i eukaryotických organismů. ALDH se podílejí na metabolismu aldehydů endogenního i exogenního původu za pomoci NAD<sup>+</sup> nebo NADP<sup>+</sup> jako kofaktorů a zabraňují tedy jejich akumulaci. Tyto enzymy jsou také užitečné při syntéze mnoha důležitých biomolekul z aldehydových meziproductů, jakými jsou například retinová kyselina, folát nebo betain (Marchitti *et al.*, 2008). ALDH proteiny se mohou vyskytovat v různých subcelulárních oblastech včetně mitochondrií nebo cytosolu a vykazují rozličnou substrátovou specifitu (Vasiliou *et al.*, 2004).

Způsob rozřazení a pojmenování neboli nomenklatura, těchto enzymů se řídí pravidly „Human Gene Nomenclature“ a byla zavedena roku 1999 (Vasiliou *et al.*, 1999). Názvy všech sekvencí obsahují stejný základ „ALDH“, následovaný číslicí označující její příslušnost do jednotlivých rodin. Následuje písmeno abecedy, značící podrodinu a případně se dodává další číslice, pomocí které se rozlišují jednotlivé geny dané podrodiny. ALDH se třídí na základě párového porovnávání aminokyselinových sekvencí. Geny se sekvenční podobností 40 a více procent, řadíme do stejné rodiny. Pokud je identita sekvencí vyšší nebo rovna 60 %, náleží také do společné podrodiny (Brocker *et al.*, 2013).

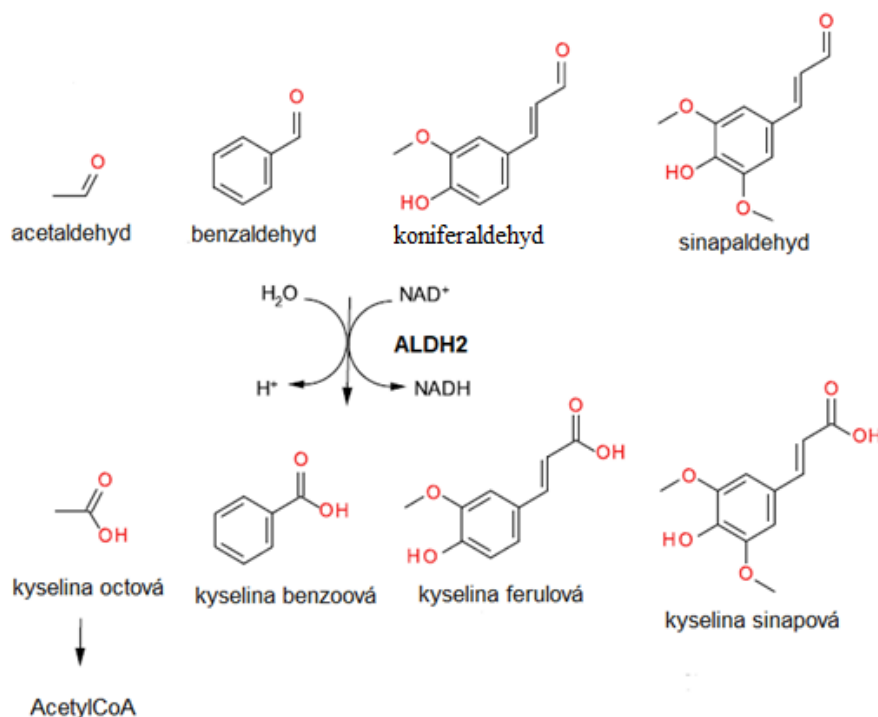
### 2.2 Rostlinné ALDH

Momentálně existuje 24 známých rodin aldehyddehydrogenas, mnohé specifické pouze pro určité organismy. V rostlinách bylo nalezeno 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 v rostlinách je různé, avšak převažuje trend, že komplexnější rostliny obsahují zpravidla větší množství genů. Ty se při přechodu z vodního prostředí na souš musely adaptovat vůči suchozemským stresorům a mnoho aldehyddehydrogenas odpovídá právě na stresové podmínky, jakými jsou zvýšená salinita prostředí, nepřiměřené teplo a zima nebo třeba oxidativní stres (Sunkar *et al.*, 2003, Cronk, 2001). Rodiny ALDH10, ALDH12, ALDH19, ALDH21, ALDH22, ALDH23 a ALDH24, jsou specifické tím, že jsou typické výhradně pro rostliny, kdy u rodin ALDH21, ALDH23 a ALDH24 se předpokládá, že se vyskytují pouze u nižších rostlin. Geny rodiny ALDH22 se pak s největší pravděpodobností vyvinuly až u vyšších rostlin (Brocker *et al.*, 2013).

## 2.2.1 Rodina ALDH2

Jedná se o skupinu rozmanitých mitochondriálních a cytosolických enzymů, s relativně širokou substrátovou specifikou, které jsou aktivní jako homotetramery (Marchitti *et al.*, 2008). Tato rodina se momentálně dělí do pěti podrodin ALDH2A, 2B, 2C, 2D a 2E. Mezi ALDH2A se řadí především živočišné enzymy, kdežto rostlinné ALDH se dělí do zbylých 4 podrodin.

ALDH2 se podílí na metabolismu ethanolu tak, že katalyzují oxidaci acetaldehydu na acetát. Ten se dále využívá při syntéze koenzymu A za pomoci acetyl-CoA synthasy, v procesu zvaném „pyruvátdehydrogenasová dráha“. V *Arabidopsis* obstarává tuto reakci především ALDH2B4 (Wei *et al.*, 2009). V tabáku byly tyto ALDH identifikovány v reprodukčních tkáních, kde hrají roli při vývoji pylu (op den Camp and Kuhlemeier 1997; Wei *et al.*, 2009). Studie genu ALDH2C4 pak naznačují, že má vliv na biosyntézu kyseliny ferulové a sinapové, látek významných pro pevnost buněčných stěn (Nair *et al.*, 2004). Lidská ALDH2 hraje důležitou roli v metabolismu neurotoxického 3,4-dihydroxyfenylacetaldehydu (DOPAL). Bylo vypořádáno potlačení oxidace DOPAL za přítomnosti acetaldehydu, tyto substráty tedy soutěží o to, který bude metabolizován stejným enzymem (MacKerell *et al.*, 1986).



**Obr. 1** Reakce katalyzované pomocí enzymů ALDH2, které využívají NAD<sup>+</sup> jako kofaktor. Acetaldehyd se přeměňuje na kyselinu octovou, ta se pak využívá k přípravě Acetyl-CoA, benzaldehyd na kyselinu benzoovou, koniferaldehyd na kyselinu ferulovou, sinapaldehyd na kyselinu sinapovou. Upraveno dle Končítíková *et al.*, (2015).

### 2.2.2 Rodina ALDH3

Tato rodina enzymů se u živočichů vyskytuje převážně v mitochondriích a cytosolu, kdežto u rostlin bylo zjištěno, že její enzymy mohou být obsaženy v rozličných organelách, jakými jsou cytosol, chloroplast nebo endoplazmatické retikulum (Marchitti *et al.*, 2008; Kirch *et al.*, 2004; Jimenez-Lopez *et al.*, 2010; Stiti *et al.*, 2011). Většina genů této rodiny má expresi regulovanou pomocí kyseliny abscisové.

Jedná se o velmi rozmanitou rodinu, jejíž zástupce je možné rozdělit do deseti podrodin. Živočišné geny se řadí převážně mezi podrodiny ALDH3A a ALDH3B, ale byly anotovány také ALDH3C, 3D a 3G. U rostlin dochází ke značné divergenci do šesti podrodin, jmenovitě ALDH3E, 3F, 3H, 3I, 3J a 3K. Podrodina ALDH3A obsahuje dioxinem indukovatelné enzymy, které se pak podílí na oxidaci středně dlouhých a dlouhých řetězců alifatických i aromatických aldehydů (Pappa *et al.*, 2003). Geny podrodiny ALDH3B kódují enzymy, které metabolizují aldehydy vznikající při peroxidaci lipidů. V lidském organismu tuto funkci zastává ALDH3B1. Byla identifikována také ALDH3B2 s 83 % identitou k první sekvenci. Její funkce však zatím není známa (Marchitti *et al.*, 2010).

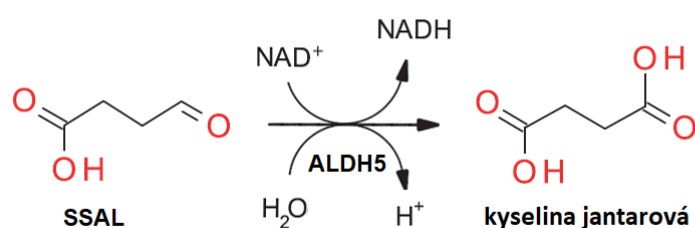
Co se týče rostlinných zástupců, tak v *Arabidopsis thaliana* byly identifikovány tři podrodiny (3F, 3H a 3I), kdy pro gen *ALDH3II* platí, že k jeho největší expresi dochází v listech a odpovídá na stresové podmínky jako dehydratace, pesticidy, těžké kovy nebo zvýšená salinita stejně jako *ALDH3HI*, jehož exprese je však obecně nižší (Kirch *et al.*, 2004; Jimenez-Lopez *et al.*, 2010; Stiti *et al.*, 2011; Missihoun *et al.*, 2012). *ALDH3FI* naopak na stresové podmínky vůbec nereaguje a její exprese je konstitutivní a bylo tedy vyvozeno, že tato podrodina vznikla nejspíše na základě specializace určitých tkání či organel (Kirch *et al.*, 2004).

### 2.2.3 Rodina ALDH5

Do rodiny ALDH5 řadíme sukcinátsemialdehyddehydrogenasy (SSALDH), které katalyzují  $\text{NAD}^+$  dependentní přeměnu sukcinát semialdehydu (SSAL) na sukcinát, jakožto součást katabolismu  $\gamma$ -aminobutyrate (GABA). Předpokládá se, že se jedná o mitochondriální proteiny, jelikož sukcinát se pak zapojuje do cyklu trikarboxylových kyselin (Bouche *et al.*, 2003; Gao and Han 2009). Zvířecí ALDH se zpravidla řadí do podrodiny ALDH5A, naopak většina rostlin obsahuje enzymy zařazené do podrodiny ALDH5F. Existuje však několik výjimek v podobě řas jako je například *Chlamydomonas*

*reinhardtii*, jejichž genové sekvence se značně liší od ostatních zástupců podrodiny 5F a byly zařazeny do podrodiny ALDH5G (Brocker *et al.*, 2013).

GABA hraje důležitou roli jako neurotransmitter v organismech savců a u rostlin pomáhá při opylování, oxidativním stresu nebo hypoxii (Fait *et al.*, 2008; Palanivelu *et al.*, 2003). Mutace *ALDH5* mohou v rostlinách způsobovat akumulaci meziproductů reaktivních forem kyslíku a buněčnou smrt během světelného či tepelného stresu (Bouche *et al.*, 2003). Naopak u člověka jsou tyto mutace zodpovědné za  $\gamma$ -hydroxymáselnou acidurii, vzácnou autosomální chorobu, při níž dochází k neurologickým poškozením nebo k poškození kognitivních funkcí, nahromaděním GABA,  $\gamma$ -hydroxymáselné kyseliny nebo SSAL v tkáních či tělních tekutinách (Akaboshi *et al.*, 2003).



**Obr. 2** Reakce sukcinátsemialdehyddehydrogenasy (ALDH5). Z SSAL při ní vzniká kyselina jantarová a NAD<sup>+</sup>, jakožto koenzym, se přeměňuje na NADH. Upraveno dle Kopečná *et al.*, (2015)

#### 2.2.4 Rodina ALDH6

Rodina ALDH6 je skupina mitochondriálních tetramerních enzymů nazývaných methylmalonylsemialdehyddehydrogenasy (MMSDH). Ty se účastní katabolismu valinu a pyrimidinu a katalyzují koenzym A – dependentní oxidativní dekarboxylaci malonátsemialdehydu na acetyl-CoA a také přeměnu methylmalonylsemialdehydu (MMS) na propionyl-CoA. Jedná se tak o jedinou rodinu aldehyddehydrogenas, která využívá koenzym A jako kofaktor (Marchitti *et al.*, 2008). Všechny doposud zkoumané rostliny obsahovaly alespoň jeden gen náležející do rodiny ALDH6 s poměrně velkou sekvenční identitou, řadí se tedy do podrodiny ALDH6B. Živočišné sekvence se naopak řadí do podrodiny ALDH6A.

Bylo zjištěno, že při ošetření hormony auxinem a giberelinem a také při zvýšené salinitě či dehydrataci dochází k upregulaci rostlinných *ALDH6* genů (Oguchi *et al.*, 2004; Zhang *et al.*, 2012). Přesná funkce této rodiny v rostlinách však stále není detailně popsána.

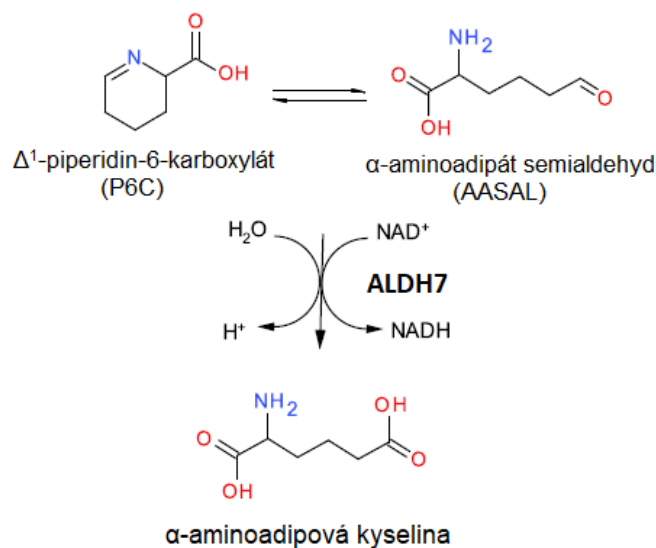


### 2.2.5 Rodina ALDH7

Rodina ALDH7 obsahuje enzymy známé jako  $\Delta^1$ -piperidin-6-karboxylátdehydrogenasy (P6CDH),  $\alpha$ -aminoadipátsemialdehyddehydrogenasy nebo antiqitiny. Jedná se o velice konzervovanou rodinu. Míra podobnosti sekvencí i z evolučně velmi vzdálených druhů je značně vysoká a je tedy velmi pravděpodobné, že i jejich funkce bude konzervovaná, a významná pro funkci buňky. Rodinu ALDH7 dělíme do podrodin ALDH7A a ALDH7C pro živočišné a ALDH7B a ALDH7D pro rostlinné zástupce. Lidská ALDH7A má pak například téměř 60% identitu s rostlinnými geny z rodiny 7B (Brocker *et al.*, 2013). Liší se však lokalizací, zatímco u rostlin se vyskytují tyto geny pouze v cytosolu, u savců byly nalezeny také v mitochondriích (Wong *et al.*, 2010).

Pro savčí ALDH7A byla zjištěna poměrně široká substrátová specifita, kdy lidská aldehyddehydrogenasa preferuje substrát  $\alpha$ -aminoadipát semialdehyd, stejně jako například enzym z kukuřice (*Zea mays*) (Končítíková *et al.*, 2015). Tyto enzymy se pak účastní metabolismu lysinu a katalyzují přeměnu  $\alpha$ -aminoadipových semialdehydů (AASAL) na  $\alpha$ -aminoadipát. Mimo to metabolizují také mnoho aldehydů vzniklých peroxidací lipidů (Mills *et al.*, 2006).

Genová exprese rostlinných enzymů z rodiny ALDH7, probíhá s největší pravděpodobností jako odpověď na stresové faktory, jelikož byla pozorována zvýšená exprese *ALDH7B* například při dehydrataci, zvýšené salinitě nebo po ozáření ultrafialovým zářením (Kotchoni *et al.*, 2006; Rodrigues *et al.*, 2006). Jejich obecná funkce není zatím známa, ale předpokládá se, že může být obdobná jako u savců, kvůli vysoké konzervovanosti sekvencí. Bylo však zjištěno, že například *ALDH7B* v rýži, pomáhá při dozrávání semen nebo u hrachu (*Pisum sativum*) je součástí procesů pro regulaci turgoru. Rekombinantní *ALDH7B* je pak schopná metabolizovat malondialdehyd (MDA), acetaldehyd nebo glycerinaldehyd (Shin *et al.*, 2009).



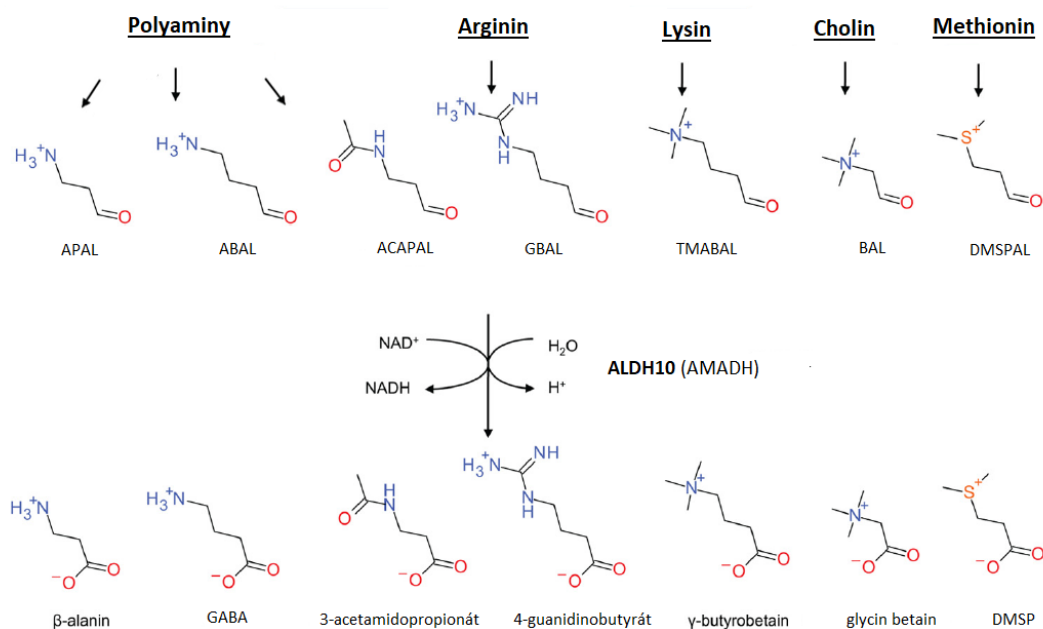
**Obr. 3** Oxidace  $\alpha$ -aminoadipát semialdehydu na  $\alpha$ -aminoadipovou kyselinu. Reakce je katalyzována pomocí ALDH7, která využívá  $\text{NAD}^+$  koenzym. Upraveno dle Končítiková *et al.*, (2015).

### 2.2.6 Rodina ALDH10

Enzymy rodiny ALDH10 nazýváme aminoaldehyddehydrogenasy (AMADH), 4-aminobutyraldehyddehydrogenasy, 4-guanidinobutyraldehyddehydrogenasy nebo také betainaldehyddehydrogenasy (BADH). Jejich hlavní rolí je deaktivace  $\omega$ -aminoaldehydů, jakými jsou 3-aminopropionaldehyd (APAL) nebo 4-aminobutyraldehyd (ABAL), které vznikají během katabolismu polyaminů rostlinnými aminooxidasami. Toho je docíleno jejich oxidací na netoxické metabolity jako  $\beta$ -alanin a  $\gamma$ -aminobutyrate (GABA) (Šebela *et al.*, 2000). Mutace genu pro AMADH může způsobit, že začne docházet k acylaci volných ABAL, při níž vzniká 2-acetyl-D<sub>1</sub>-pyrolin, který je aromatický a dodává vůni některým druhům rýže (Arikiti *et al.*, 2011; Bradbury *et al.*, 2008). Gen *ALDH10* také odpovídá na stresové podmínky, specificky na vyšší sucho, tím, že enzym produkuje osmoprotektivní glycin betain (GB). Ten funguje jako buněčný osmolyt a chaperon, tedy pomáhá udržovat osmotickou rovnováhu organismu a je využíván také pro stabilizaci struktury a funkce proteinů u některých rostlin (Allakhverdiev *et al.*, 2008).

Geny této rodiny můžeme dále dělit do podrodin *ALDH10A*, *ALDH10B* a *ALDH10C*, přičemž většina rostlinných zástupců se řadí do podrodiny 10A na základě jejich vysoké sekvenční homologie. Výjimku tvoří například sekvence z vinné révy (*Vitis vinifera*), jež byla zařazena do podrodiny ALDH10B a řasa *Volvox carteri*, kde byl identifikován gen, který měl nízkou podobnost s ostatními zástupci podrodiny 10A i 10B

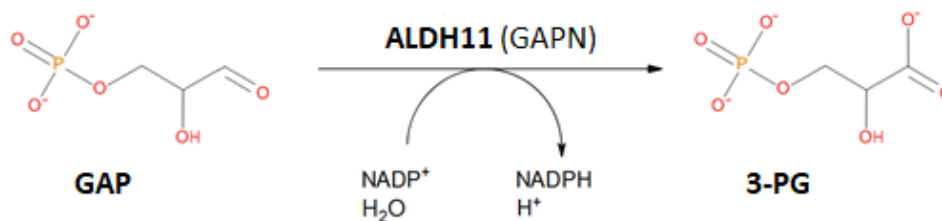
a byl tedy zařazen do nové podrodiny 10C (Brocker *et al.*, 2013). Savčí genomy pak ALDH10 neobsahují. Jejich funkci z velké části zastávají ALDH9, jež jsou hlavními producenty GB a  $\gamma$ -butyrobetainu savců (Brocker *et al.*, 2010; Vaz *et al.*, 2000; Končítíková *et al.*, 2019).



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

## 2.2.7 Rodina ALDH11

Rodina genů ALDH11 je zastoupena cytosolickými a nefosforylujícími glyceraldehyd-3-fosfátdehydrogenasami (GAPN). Ty se podílí na tzv. glykolytickém bypassu, reakci specifické pro fotosyntetizující rostliny a řasy, jež umožňuje přeskočit první fosforylační krok glykolýzy (Plaxton 1996; Valverde *et al.*, 1999). Zde katalyzují jednosměrnou oxidaci GAP na 3-fosfoglycerát (3-PG) za přítomnosti NADP<sup>+</sup> jako kofaktoru. Tím produkuje také NADPH, který se dále využívá při biosyntéze mannitolu (Valverde *et al.*, 1999; Gao a Loescher, 2000). Většina ALDH11 se řadí do podrodiny ALDH11A. Výjimku tvoří například *Vitis vinifera*, kde identifikované enzymy byly zařazeny do podrodiny ALDH11B (Brocker *et al.*, 2013).

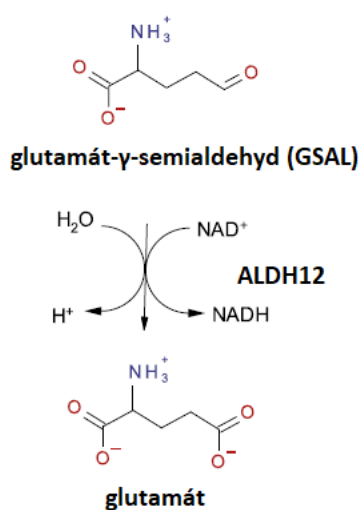


**Obr. 5** NADP<sup>+</sup> dependentní oxidace glyceraldehyd-3-fosfátu na 3-fosfoglycerát. Nefosforylující reakce enzymů rodiny ALDH11.

## 2.2.8 Rodina ALDH12

$\Delta$ -1-pyrrolin-5-karboxylát dehydrogenasy nebo-li NAD<sup>+</sup>-dependentní glutamát- $\gamma$ -semialdehyddehydrogenasy (GSALDH) jsou mitochondriální enzymy, které náleží do rodiny ALDH12. Jejich hlavní funkcí v organismu je přeměna glutamát  $\gamma$ -semialdehydu (GSAL) na glutamát, čímž se podílí na procesu katabolismu prolinu a argininu. Během stresových podmínek jako je sucho nebo zvýšená slanost prostředí, pak dochází ke snížené expresi těchto genů, nejspíše z důvodu udžení hladiny prolinu.

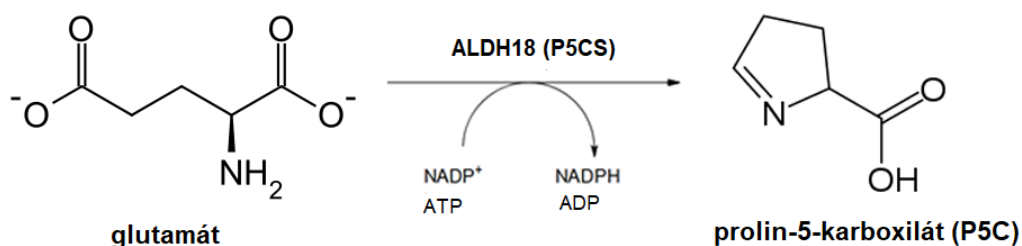
Rostlinné GSALDH se vyskytují u všech *Embryophyt*, zelených řas a *Heterokont*. Mají nízkou sekvenční podobnost (<24%) s živočišnými či bakteriálními sekvencemi, které se tak řadí do odlišné rodiny ALDH4. Tyto dvě rodiny však sdílí podobnou funkci, jelikož jejich aktivní místo je téměř totožné. Na základě fylogenetické analýzy můžeme soudit, že se *ALDH12* oddělily od genů *ALDH4* již při vývoji endosymbiotických předků rostlinných zástupců (Korasick *et al.*, 2019).



**Obr. 6** Oxidace GSAL, při níž vzniká glutamát a NADH. Tato reakce, v některých organismech též zprostředkována rodinou ALDH4, je součástí rostlinných drah pro degradaci prolinu a argininu. Upraveno dle Korasick *et al.*, (2019).

### 2.2.9 Rodina ALDH18

Rodinu ALDH18 tvoří  $\Delta$ -1-pyrrolin-5-karboxylátsynthasy (P5CS). Jedná se o bifunkční mitochondriální enzymy. Obsahují dvě domény, aldehyddehydrogenasovou na C-konci a kinasovou na N-konci. ALDH18A1 identifikovaná v genomu savců, k jejíž vysoké expresi dochází například ve slinivce či ledvinách, zprostředkovává přeměnu glutamátu na pyrrolin-5-karboxylát (P5C) za přítomnosti ATP a NADPH. P5C je dále přeměněn na ornithin, který se využívá při biosyntéze prolinu a argininu (Marchitti *et al.*, 2008). Poškození genů *ALDH18A1* pak může vést k mnoha metabolickým či neurologickým problémům, jakými jsou například hypoprolinemie nebo hypoornithinemie (Baumgartner *et al.*, 2005). Rostlinné ALDH18 jsou poměrně konzervované a reagují na dehydrataci organismu. Jejich upregulace zrychluje syntézu prolinu a jeho akumulaci, který pak stejně jako například betain, slouží k udržování osmotické rovnováhy organismu (Yoshida *et al.*, 1997).



**Obr. 7** Reakce, kterou katalyzuje savčí ALDH18. Za přítomnosti NADP<sup>+</sup> a ATP dochází k přeměně glutamátu na pyrrolin-5-karboxylát (P5C).

### 2.2.10 Rodina ALDH19

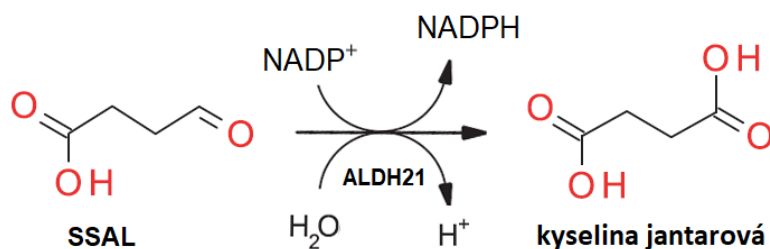
Enzymy rodiny ALDH19 byly identifikovány v rostlinách a bakteriích. Jsou známy také jako  $\gamma$ -glutamylfosfátreduktasy podobné lidským ALDH18A1 (Vasiliou *et al.*, 2013). Jedná se o proteiny indukované osmotickým stresem, katalyzující přeměnu L-glutamátu na L-glutamát-5-semialdehyd a fosfát, čímž se podílejí na biosyntéze L-prolinu z L-glutamátu (García-Ríos *et al.*, 1997). Sekvence ALDH19 v genomu *Solanum lycopersicum*, jediném doposud zkoumaném rostlinném zástupci, který tuto rodinu obsahuje, se skládá ze dvou ORF, které kódují doménu pro  $\gamma$ -glutamylkinasu a  $\gamma$ -glutamylfosfátreduktasu. Tyto ORF připomínají uspořádáním například polycistronické

operony v prokaryotách. Porovnání této sekvence s glutamylfosfátreduktasou z bakterie *Leuconostoc lactis* ukázalo, že se jedná o totožné sekvence. Je tedy pravděpodobné, že rostlinný zástupce této rodiny vznikl horizontálním genovým přenosem z bakterie *Leuconostoc*, do genomu *Solanum lycopersicum* (Stiti *et al.*, 2021).

### 2.2.11 Rodina ALDH21

Rodina ALDH21, vyskytující se pouze v určitých nižších rostlinách, jako jsou zelené řasy, Bryophyta, plavuně anebo kapradiny, obsahuje cytosolické tetramerní enzymy známé též jako sukcinátsemialdehyddehydrogenasy (SSALDH). Tyto enzymy využívají  $\text{NADP}^+$  jako koenzym a mají silnou preferenci pro sukcinát semialdehyd (SSAL) jakožto substrát. Účastní se tedy metabolismu  $\gamma$ -aminomáselné kyseliny, kdy oxidují SSAL na sukcinát. Pomáhají také rostlinám zvládat stresové podmínky jako je sucho či zvýšená salinita prostředí (Yang *et al.*, 2015).

Funkčně nejpříbuznější rodinou těchto ALDH je ALDH5, jejíž enzymy využívají  $\text{NAD}^+$  jako koenzym. Aktivita ALDH21 s tímto koenzymem je však pouze 3 %. Rostliny obsahující obě zmíněné rodiny jsou pak schopny oxidovat SSAL na sukcinát v cytosolu i mitochondriích. Přičemž vyšší rostliny, které zpravidla postrádají rodinu ALDH21 mají tuto funkčnost pouze v mitochondriích (Kopečná *et al.*, 2017).



**Obr. 8** Sukcinátsemialdehyddehydrogenasa (ALDH21), přeměňující SSAL na kyselinu jantarovou. Tato reakce využívá  $\text{NADP}^+$  jako koenzym na rozdíl od ALDH5. Upraveno dle Kopečná *et al.*, (2015).

## 2.2.12 Rodina ALDH22, ALDH23 a ALDH24

Tyto rodiny rostlinných ALDH byly identifikovány relativně nedávno a jejich podrobná funkce není zatím známa. *ALDH22A1*, jejíž přítomnost byla zjištěna v kukuřici *Zea mays*, reaguje vyšší expresí na stresové podmínky, jako nedostatek vody či zvýšená salinita prostředí (Huang *et al.*, 2008). Zvýšená přítomnost těchto enzymů v organismu pak pomáhá rostlině překonat nepříznivé podmínky.

Geny rodiny *ALDH23* byly doposud identifikovány pouze v nižších rostlinách jako jsou například mech *Physcomitrella patens* nebo plavuně *Selaginella moellendorffii*. Jejich funkce je neznámá stejně jako funkce enzymů z rodiny *ALDH24*. Byla však zjištěna sekvenční identita až 30% mezi geny *ALDH24* z řasy *Chlamydomonas reinhardtii* a většinou zástupců z rodiny *ALDH23* (Brocker *et al.*, 2013).

## 2.3 Živočišné a jiné ALDH

Většina rostlinných ALDH má také živočišné orthology. Rodinami specifickými pro živočišnou říši jsou však *ALDH1*, *ALDH4*, *ALDH8*, *ALDH9* a *ALDH16*. Mnohé živočišné ALDH pak mohou vykazovat vlastnosti mimo své enzymatické funkce. Některé se nacházejí v rohovce a čočce oka savců v podobě krystalinů, zde pak slouží k rozličným účelům. V první řadě pohlcují ultrafialové (UV) záření, čímž chrání oko před oxidativním poškozením. Přispívají také k jeho transparentnosti, zvýšením přenosu světla a jeho lomu na sítnici. Dále mohou také vázat hormony a jiné malé molekuly, jako jsou androgeny, cholesterol nebo acetaminofen (Estey *et al.*, 2007; Lassen *et al.*, 2006; Marchitti *et al.*, 2008).

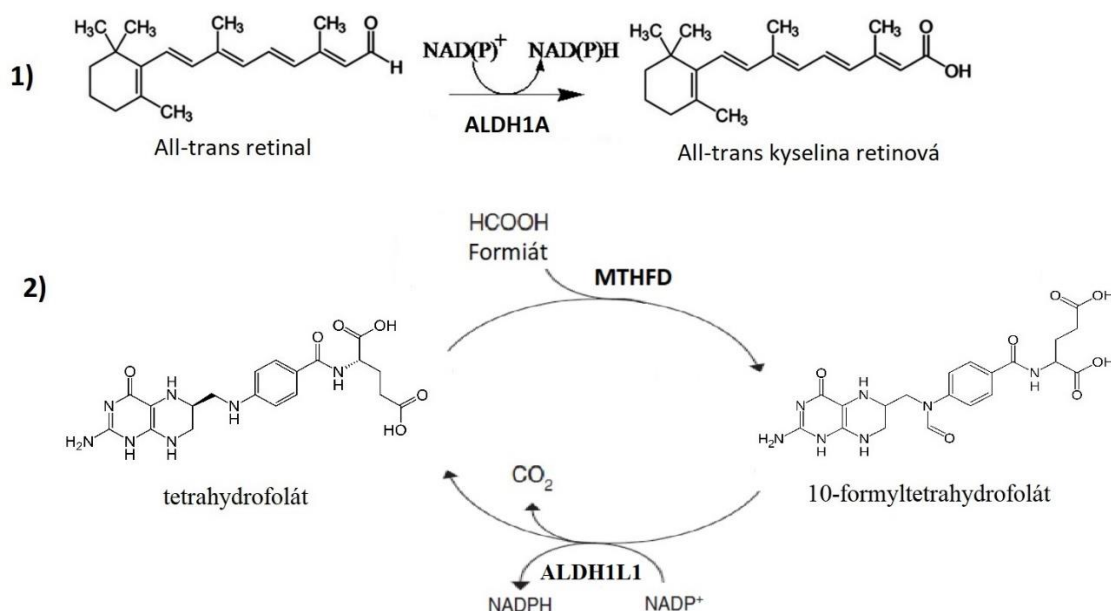
Zbylé rodiny ALDH, a to konkrétně *ALDH13*, *ALDH14*, *ALDH15*, *ALDH17* a *ALDH20*, jsou poměrně neprozkoumané a jejich přesná funkce není ve většině případů známa. Jedná se o rodiny vyskytující se u kvasinek a hub či měňavek a diplomonád.

### 2.3.1 Rodina ALDH1

První rodina ALDH je specifická pro živočišné genomy a obecně se dělí do tří největších podrodin, *ALDH1A*, *ALDH1B* a *ALDH1L*. Podrodina *ALDH1A* obsahuje enzymy, které se podílejí na přeměně retinalaldehydu na kyselinu retinovou (RA) a jsou tedy důležité pro regulaci signalizace této kyseliny. Ta pak ovlivňuje například funkci vitamínu A1, potřebného pro růst organismu (Vasiliou *et al.*, 1999). Tyto enzymy mají vysokou afinitu jak k *all-trans* retinalu, tak i *9-cis* retinalu (G. Duester, 2000).

Mitochondriální enzymy náležející do podrodiny ALDH1B mají vysokou podobnost s rodinou ALDH2, kdy mohou dosahovat až 75% identity sekvencí. Tyto proteiny, se taktéž podílejí na metabolismu acetaldehydu, avšak s nižší afinitou a aktivitou než právě ALDH z rodiny 2 (Stagos *et al.*, 2010). Tato podrodina je zastoupena u savců, avšak u ryb či ptáků tyto geny chybí (Jackson *et al.*, 2011). Lidská ALDH1B1 pak může také sloužit jako biomarker pro rakovinu tlustého střeva. (Chen *et al.*, 2011).

Podrodina ALDH1L obsahuje mitochondriální i cytosolické enzymy. Cytosolická ALDH1L1 funguje především jako 10-FTHFdehydrogenasa (FDH) a přeměňuje 10-formyltetrahydrofolát (10-FTHF) na tetrahydrofolát (Tsybovsky a Krupenko, 2011). Gen *ALDH1L2* je velmi podobný *ALDH1L1* a kóduje mitochondriální FDH, s podobnou funkcí. Na rozdíl od ALDH1L1 však nemetabolizuje pouze krátké aldehydy (Strickland *et al.*, 2011).



**Obr. 9 Reakce katalyzované enzymy rodin ALDH1A a ALDH1L.** Dochází k přeměně all-trans retinalu na kyselinu retinovou za pomoci NAD<sup>+</sup> nebo NADP<sup>+</sup> jako kofaktoru (1). ALDH1L1 je součástí metabolismu metanolu, kde přeměňuje 10-formyltetrahydrofolát na tetrahydrofolát a využívá NADP<sup>+</sup> jako kofaktor (2). Upraveno dle Seung-Hye *et al.*, (2016) a Marchitti *et al.*, (2008).

### 2.3.2 Rodina ALDH4

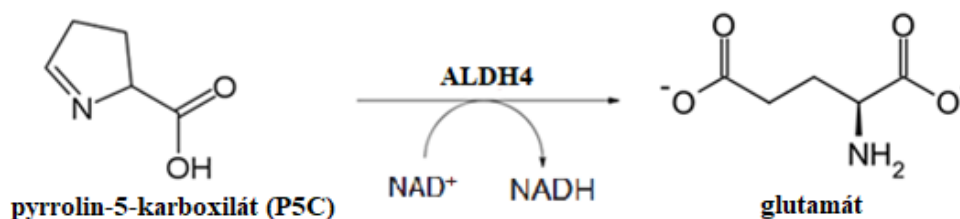
Rodina ALDH4 obsahuje mitochondriální enzymy známé jako pyrrolin-5-karboxylát dehydrogenasy (P5CDH) nebo také glutamát- $\gamma$ -semialdehyd dehydrogenasy, které katalyzují jednosměrnou NAD<sup>+</sup> dependentní přeměnu pyrrolin-5-karboxylátu (P5C)



na glutamát a zabraňují tedy jeho zvýšené akumulaci (Forte-McRobbie a Pietruszko, 1986).

ALDH4A1 oxiduje krátké a středně dlouhé alifatické aldehydy, které vznikají při peroxidaci lipidů (Farres *et al.*, 1988). Mimo to se podílí také na degradaci prolinu a hraje roli při oxidativním stresu, jelikož v laboratorních podmínkách bylo dokázáno, že buňky se zvýšenou produkcí tohoto enzymu, vytvářejí nižší množství reaktivních forem kyslíku (ROS) během stresových podmínek, jako je například ozáření UV zářením. Při poškození DNA dále dochází k upregulaci exprese tohoto genu, a hraje tedy určitou roli také při reparačních mechanismech DNA (Yoon *et al.*, 2004).

Lidská ALDH4A1 je aktivní jako homodimer a k její expresi dochází převážně v játrech, kosterní svalovině nebo ledvinách. (Hu *et al.*, 1996) Mutace tohoto genu může způsobit hyperprolinémii a zvýšenou koncentraci P5C v tělních tekutinách (Geraghty *et al.*, 1998).



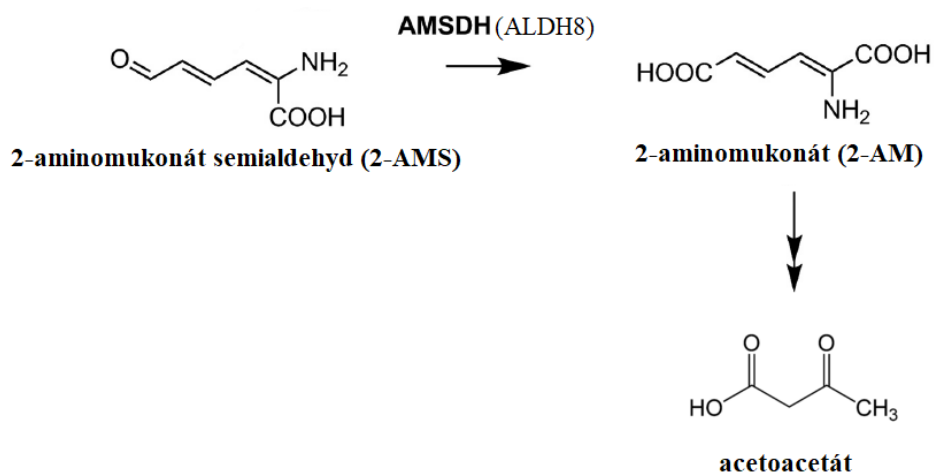
**Obr. 10** Reakce savčí ALDH4. Přeměna pyrrolin-5-karboxylátu (P5C) na glutamát, za užití NAD<sup>+</sup> jako kofaktoru.

### 2.3.3 Rodina ALDH8

Tato rodina obsahuje cytosolické enzymy náležející do jedné podrodiny ALDH8A, známé také jako 2-aminomukonát-6-semialdehyddehydrogenasy (AMSHD), které hrají roli v kynureninové dráze, což je primární proces pro degradaci tryptofanu u savců. Zde katalyzují NAD<sup>+</sup> dependentní oxidaci 2-aminomukonát semialdehydu (2-AMS) na 2-aminomukonát (2-AM). Meziprodukty této dráhy se využívají při imunitních procesech a mohou být spojeny i s neurodegenerativními chorobami (Davis *et al.*, 2018).

Tyto enzymy se také podílejí na tvorbě kyseliny retinové oxidací retinalu. Mimo to metabolizují alifatické aldehydy jako třeba acetaldehyd, decanal nebo octanal a jejich aktivita roste s délkou řetězce daného aldehydu. Jsou však aktivní i pro metabolické aldehydy jako sukcinátsemialdehyd a glutaraldehyd. K expresi genů *ALDH8* dochází

převážně v játrech a ledvinách, ale mírná exprese probíhá také v mozku a jiných orgánech (Lin a Napoli, 2000).

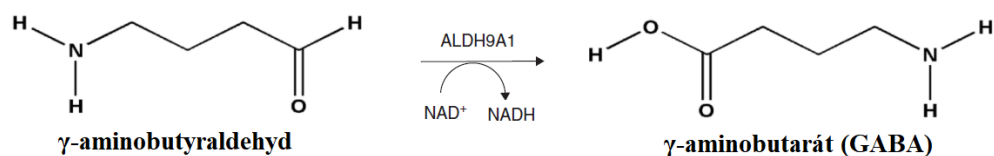


**Obr. 11** Reakce kynureninové dráhy zprostředkovaná AMSDH (ALDH8). 2-aminomukonát semialdehydu (2-AMS) se přeměňuje na 2-aminomukonát (2-AM), ten je pak dále metabolizován na acetoacetát. Upraveno dle Davis *et al.*, (2018).

### 2.3.4 Rodina ALDH9

Zástupci rodiny ALDH9 jsou tetramerní enzymy řadící se do podrodiny ALDH9A. K expresi jejich genů dochází u člověka v játrech, kosterní svalovině, ledvinách a v mozku (Lin *et al.*, 1996). Podílí se na alternativní biosyntéze GABA, kdy katalyzují oxidaci  $\gamma$ -aminobutyraldehydu na GABA, mimo to také slouží k oxidaci betain aldehydu,  $\gamma$ -trimethylaminobutyraldehydu a acetaldehydu (Kurys *et al.*, 1989; Chern a Pietruszko, 1995; Vaz *et al.*, 2000). ALDH9A hrají také určitou roli v metabolismu aldehydů odvozených od katecholaminu, jako je například toxický DOPAL (Marchitti *et al.*, 2007).

Kvůli vysoké afinitě těchto enzymů vůči  $\gamma$ -aminobutyraldehydu, DOPAL i acetaldehydu, mohou jejich substráty soupeřit o to, který z nich bude metabolizován a následně tedy ovlivňovat GABA cyklus nebo také dopaminové dráhy a vývoj mozku. U modelové ryby, byla pozorována vývojová regulace *ALDH9A1* a také její downregulace při zvýšené koncentraci ethanolu během embryogeneze. Tento enzym musí tedy hrát také určitou roli v teratogenních účincích ethanolu (Wang *et al.*, 2007).



**Obr. 12 Alternativní biosyntéza GABA pomocí ALDH9.** Dochází k  $\text{NAD}^+$  dependentní oxidaci  $\gamma$ -aminobutyraldehydu na  $\gamma$ -aminobutarát (GABA). Upraveno dle Marchitti *et al.*, (2008).

### 2.3.5 Rodiny ALDH13, ALDH14 a ALDH15

Tyto rodiny jsou zatím poměrně neprozkoumané. Rodina ALDH13 obsahuje pouze jeden enzym nalezen u organismu *Entamoeba histolytica*, který se podílí na metabolismu acetaldehydů a podobá se savčím ALDH3 (Samuelson *et al.*, 1992). Rodiny ALDH14 a ALDH15 obsahují enzymy z kvasinek a hub (Sophos a Vasiliou, 2003). Mezi ALDH14 patří mitochondriální enzymy podobné mikrosomální ALDH3A2 (Jackson *et al.*, 2011). Geny z rodiny ALDH15 jsou podobné rostlinným ALDH21 a zřejmě hrají určitou roli v ochraně proti oxidativnímu stresu (Brocker *et al.*, 2013).

### 2.3.6 Rodina ALDH16

Rodina ALDH16 byla objevena poměrně nedávno. Prvně byla sekvenována ALDH16A1 z lidské cDNA sekvenční knihovny konsorciem Mammalian Gene Collection (MGC) (Strausberg *et al.*, 2002). Orthologní geny byly následně identifikovány v ostatních organismech jako jsou šimpanz, myš nebo ryba (*Danio rerio*). K expresi enzymů této rodiny dochází v mnoha tkáních například v srdci, plicích nebo ledvinách (Shmueli *et al.*, 2003) a na základě podobnosti se dělí do 4 podrodin, 16A (savci), 16B (obojživelníci), 16C (bakterie) and 16D (ryby) (Vasiliou *et al.*, 2013).

ALDH16A1 je poměrně specifický zástupce nadrodiny ALDH, jelikož na rozdíl od všech ostatních, obsahuje dvě ALDH domény. Dále také obsahuje 4 transmembránové domény a jednu coil-coil doménu. Aktivní místo ALDH16 pak u mnoha obojživelníků a bakterií obsahuje katalytický cystein (Cys-302), který se však nevyskytuje u savců či ryb, jejichž ALDH16 jsou tak obecně považovány za pseudoenzymy (Vasiliou *et al.*, 2013).

Funkce a význam těchto enzymů jsou stále nejasné. Bylo však zjištěno, že mutace *ALDH16A1* souvisí s nemocí známou jako „dna“ a také s hladinou kyseliny močové v krevním séru. Na základě molekulárního modelování totiž vyplynulo, že ALDH16A1 může interagovat s hypoxantin-guanin fosforibosyltransferasou (HPRT1), enzymem souvisejícím právě s touto nemocí (Vasiliou *et al.*, 2013).

### 2.3.7 Rodiny ALDH17 a ALDH20

Rodina ALDH17 obsahuje geny nalezené v *Drosophila melanogaster*, které kódují enzymy pravděpodobně využívané pro katalýzu, podobně jako například u rodiny ALDH4 (P5CDH) (Vasiliou *et al.*, 2013). Rodina ALDH20 obsahuje proteiny nalezené pouze u diplomonád a měňavek, jako jsou například *Giardia lamblia* a *Entamoeba histolytica*. Kromě ALDH domény obsahují tyto multidoménné proteiny také alkoholdehydrogenázovou doménu (Sophos a Vasiliou, 2003).

## 3 EXPERIMENTÁLNÍ ČÁST

### 3.1 Materiál

Databázi programu PAFF tvoří ALDH sekvence z třiceti organismů roztríděné do tabulek dle jednotlivých rodin. Základem jsou sekvence Huseníčku rolního (*Arabidopsis thaliana*), kukuřice (*Zea mays*) a také lidské ALDH, vůči nimž program testuje, zda je uživatelská sekvence skutečně z nadrodiny ALDH. Rostlinné sekvence z *Arabidopsis thaliana*, řasy *Chlamydomonas reinhardtii* a *Volvox carteri*, plavuně *Selaginella moellendorffii*, rýže *Oryza sativa*, mechu *Physcomitrella patens*, topolu *Populus trichocarpa*, čiroku *Sorghum bicolor*, vinné révy *Vitis vinifera* a kukuřice *Zea mays*, byly nalezeny skrze databáze NCBI a Phytozome, na základě informací publikovaných v Brocker *et al.*, (2013).

**Tab. 1 Proteiny nadrodiny ALDH přítomné v *Arabidopsis thaliana*.** Tabulka obsahuje ID jednotlivých sekvencí zaznamenaných v databázích NCBI a Phytozome. Upraveno dle Brocker *et al.*, (2013).

| Family    | Gene name       | NCBI gene ID | Other names/aliases             | NCBI protein ID | Phytozome ID | Chrm | Exon # | AA # |
|-----------|-----------------|--------------|---------------------------------|-----------------|--------------|------|--------|------|
| Family 2  | <i>ALDH2B4</i>  | 823955       | ALDH2; ALDH2A; ALDH2B4          | NP_190383.1     | AT3G48000    | 3    | 11     | 538  |
|           | <i>ALDH2B7</i>  | 838991       | ALDH2B; F5O8.35; F5O8_35        | NP_564204.1     | AT1G23800    | 1    | 11     | 534  |
|           | <i>ALDH2C4</i>  | 822042       | ALDH1A; REF1                    | NP_566749.1     | AT3G24503    | 3    | 9      | 501  |
| Family 3  | <i>ALDH3F1</i>  | 829782       | F23E13.140; F23E13_140          | NP_195348.2     | AT4G36250    | 4    | 9      | 484  |
|           | <i>ALDH3H1</i>  | 841020       | ALDH4; T7O23.15; T7O23_15       | NP_175081.1     | AT1G44170    | 1    | 10     | 484  |
|           | <i>ALDH3I1</i>  | 829573       | ALDH3; F10M10.10; F10M10_10     | NP_567962.1     | AT4G34240    | 4    | 11     | 550  |
| Family 5  | <i>ALDH5F1</i>  | 844282       | SSADH; SSADHI                   | NP_178062.1     | AT1G79440    | 1    | 20     | 528  |
| Family 6  | <i>ALDH6B2</i>  | 815903       | T22C12.10; T22C12_10            | NP_179032.1     | AT2G14170    | 2    | 19     | 607  |
| Family 7  | <i>ALDH7B4</i>  | 841849       | F15I1.19; F15I1_19              | NP_175812.1     | AT1G54100    | 1    | 14     | 508  |
| Family 10 | <i>ALDH10A8</i> | 843831       | F25A4.11; F25A4_11              | NP_001185399.1  | AT1G74920    | 1    | 15     | 496  |
|           | <i>ALDH10A9</i> | 823972       | –                               | NP_190400.1     | AT3G48170    | 3    | 15     | 503  |
| Family 11 | <i>ALDH11A3</i> | 816962       | F27D4.18; F27D4_18; NP-GAPDH    | NP_001189589.1  | AT2G24270    | 2    | 9      | 496  |
| Family 12 | <i>ALDH12A1</i> | 836373       | K19B1.14; K19B1_14; P5CDH       | NP_568955.1     | AT5G62530    | 5    | 16     | 556  |
| Family 18 | <i>ALDH18B1</i> | 818566       | ATP5CS; P5CS1; T5I7.10; T5I7_10 | NP_181510.1     | AT2G39800    | 2    | 20     | 717  |
|           | <i>ALDH18B2</i> | 824727       | P5CS2                           | NP_191120.2     | AT3G55610    | 3    | 20     | 726  |
| Family 22 | <i>ALDH22A1</i> | 819849       | T8E24.4; T8E24_4                | NP_974242.1     | AT3G66658    | 3    | 14     | 596  |

**Tab. 2 ALDH geny přítomné v kukuřici *Zea mays*, jež byly nalezeny pomocí webových databází NCBI a Phytozome. Upraveno dle Brocker *et al.*, (2013).**

| Family    | Gene name       | NCBI gene ID | Other names/aliases   | NCBI protein ID | Phytozome ID                   | Chrm           | Exon #         | AA #             |
|-----------|-----------------|--------------|---|-----------------|--------------------------------|----------------|----------------|------------------|
| Family 2  | <i>ALDH2B2</i>  | 732806       | ALDH2B1 <sup>b</sup> ; ZmALDH2B1; rf2a                          | NP_001105891.1  | GRMZM2G058675                  | 9              | 11             | 549              |
|           | <i>ALDH2B5</i>  | 542567       | ALDH2B6 <sup>b</sup> ; ZmALDH2B5; rf2b; ac182825.4              | NP_001105576.1  | GRMZM2G125268                  | 4              | 9              | 550              |
|           | <i>ALDH2C1</i>  | 541913       | ALDH2C2 <sup>b</sup> ; ZmALDH2C1a; rf2c                         | NP_001105046.1  | GRMZM2G071021                  | 3              | 7              | 503              |
|           | <i>ALDH2C2</i>  | 541914       | ALDH2C3 <sup>b</sup> ; ALDHRF2D; rf2d; rf2d1; ALDH5; ZmALDH2C1b | NP_001105047.1  | GRMZM2G097706                  | 3              | 8              | 511              |
|           | <i>ALDH2C4</i>  | No entry     | rf2e  | No entry        | GRMZM2G380438G<br>RMZM2G407949 | 8 <sup>a</sup> | 8 <sup>a</sup> | 324 <sup>c</sup> |
|           | <i>ALDH2C5</i>  | No entry     | rf2f  | No entry        | GRMZM2G122172                  | 6 <sup>a</sup> | 8 <sup>a</sup> | 516              |
| Family 3  | <i>ALDH3E1</i>  | 100280692    | ZmALDH3E1   | NP_001147083.1  | GRMZM2G169458                  | 5              | 10             | 485              |
|           | <i>ALDH3E2</i>  | 100284774    | ZmALDH3E2   | NP_001151141.1  | GRMZM2G155502                  | 10             | 9              | 489              |
|           | <i>ALDH3H1</i>  | 100285809    | ZmALDH3H1   | NP_001152171.1  | GRMZM2G103546                  | 10             | 10             | 491              |
|           | <i>ALDH3H2</i>  | 100281700    | ZmALDH3H2   | NP_001148092.1  | GRMZM2G060800                  | 4              | 10             | 481              |
|           | <i>ALDH3H3</i>  | 100382449    | ZmALDH3H3   | NP_001168661.1  | GRMZM2G118800                  | 2              | 10             | 478              |
| Family 5  | <i>ALDH5F1</i>  | 100280779    | ZmALDH5F1   | NP_001147173.1  | GRMZM2G128114                  | 5              | 20             | 493              |
|           | <i>ALDH5F2</i>  | 100284047    | ZmALDH5F2   | NP_001150417.1  | GRMZM2G119482                  | 4              | 20             | 527              |
| Family 6  | <i>ALDH6B1</i>  | 100274311    | ZmALDH6B1   | NP_001142146.1  | GRMZM2G001898                  | 7              | 13             | 537              |
| Family 7  | <i>ALDH7B6</i>  | 100282748    | ZmALDH7B6   | NP_001149126.1  | GRMZM2G130440                  | 2              | 14             | 509              |
| Family 10 | <i>ALDH10A5</i> | 541949       | ZmALDH10A5; AMADH2  | ACS74868.1      | GRMZM2G135470                  | 10             | 15             | 506              |
|           | <i>ALDH10A8</i> | 100302679    | ZmALDH10A8; AMADH1a   | ACS74867.1      | GRMZM2G013214<br>GRMZM2G146754 | 4              | 15             | 505              |
|           | <i>ALDH10A9</i> | No entry     | ZmALDH10A9; AMADH1b   | AEP68091.1      | GRMZM2G016189                  | 1              | 15             | 506              |
| Family 11 | <i>ALDH11A3</i> | 542583       | ZmALDH11A3  | NP_001105589.1  | GRMZM2G035268                  | 4              | 9              | 498              |
| Family 12 | <i>ALDH12A1</i> | No entry     | ZmALDH12A1  | AAL70108.1      | GRMZM2G090087                  | 6              | 15             | 549              |
| Family 18 | <i>ALDH18B2</i> | 100280719    | ZmALDH18B2  | NP_001147111.1  | GRMZM2G028535                  | 8              | 21             | 731              |
|           | <i>ALDH18B1</i> | No entry     | ZmALDH18B1  | ACR33941.1      | GRMZM2G375504                  | 8              | 20             | 717              |
| Family 22 | <i>ALDH22A1</i> | 100125658    | ZmALDH22A1; AC212124.5  | NP_001106059.1  | GRMZM2G135341                  | 7              | 15             | 593              |

Dále databáze programu obsahuje také sekvence z řas *Ostreococcus lucimarinus*, *Chlorella variabilis*, *Auxenochlorella protothecoides*, *Chara braunii*, mechu *Sphagnum fallax*, smrku *Picea sitchensis*, rostliny *Amborella trichopoda*, vochy *Zostera marina* a špenátu *Spinacia oleracea*, taktéž uložené v databázích NCBI a Phytozome, jež byly publikovány ve Stiti *et al.*, (2021).

Pro nalezení sekvencí nadrodiny ALDH v genomech řas *Coccomyxa subellipsoidea* a *Chromochloris zofingiensis*, jabloně *Malus domestica*, broskvoně *Prunus persica* a rajčete *Solanum lycopersicum* bylo využito programu BLAST poskytovaného serverem Phytozome.

ALDH sekvence z kapradin *Azolla filiculoides* a *Salvinia cucullata*, byly vyhledány skrze software BLAST na serveru FernBase:

[https://www.fernbase.org/tools/blast?db\\_id=34](https://www.fernbase.org/tools/blast?db_id=34).

Sekvence z řasy *Klebsormidium nitens*, byly identifikovány pomocí oficiálního webu genomového projektu:

[http://www.plantmorphogenesis.bio.titech.ac.jp/~algae\\_genome\\_project/klebsormidium/](http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/)  
(release: 171026)

Databáze obsahuje také lidské (*Homo sapiens*), ptačí (*Gallus gallus*) a rybí (*Danio rerio*) sekvence, jež byly nalezeny pomocí softwaru Blast na serveru NCBI, na základě informací vydaných v Jackson *et al.*, (2011).

**Tab. 3 Geny ALDH obsaženy v lidském genomu.** Zaznamenáno je jejich ID v databázi NCBI a také označení jednotlivých sekvencí s předponou *Hum* v databázi programu PAFF.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvencí v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 1</i>  | <i>ALDH1A1</i>  | NP_000680.2     | <i>HumALDH1A1</i>            | 501                |
|                  | <i>ALDH1A2</i>  | NP_003879.2     | <i>HumALDH1A2</i>            | 518                |
|                  | <i>ALDH1A3</i>  | NP_000684.2     | <i>HumALDH1A3</i>            | 512                |
|                  | <i>ALDH1B1</i>  | NP_000683.3     | <i>HumALDH1B1</i>            | 517                |
|                  | <i>ALDH1L1</i>  | NP_036322.2     | <i>HumALDH1L1</i>            | 902                |
|                  | <i>ALDH1L2</i>  | NP_001029345.2  | <i>HumALDH1L2</i>            | 923                |
| <i>Rodina 2</i>  | <i>ALDH2A1</i>  | NP_000681.2     | <i>HumALDH2A1</i>            | 517                |
| <i>Rodina 3</i>  | <i>ALDH3A1</i>  | NP_000682.3     | <i>HumALDH3A1</i>            | 453                |
|                  | <i>ALDH3A2</i>  | NP_000373.1     | <i>HumALDH3A2</i>            | 485                |
|                  | <i>ALDH3B1</i>  | NP_000685.1     | <i>HumALDH3B1</i>            | 468                |
|                  | <i>ALDH3B2</i>  | NP_001026786.2  | <i>HumALDH3B2</i>            | 385                |
| <i>Rodina 4</i>  | <i>ALDH4A1</i>  | NP_003739.2     | <i>HumALDH4A1</i>            | 563                |
| <i>Rodina 5</i>  | <i>ALDH5A1</i>  | NP_001071.1     | <i>HumALDH5A1</i>            | 535                |
| <i>Rodina 6</i>  | <i>ALDH6A1</i>  | NP_005580.1     | <i>HumALDH6A1</i>            | 535                |
| <i>Rodina 7</i>  | <i>ALDH7A1</i>  | NP_001173.2     | <i>HumALDH7A1</i>            | 539                |
| <i>Rodina 8</i>  | <i>ALDH8A1</i>  | NP_072090.1     | <i>HumALDH8A1</i>            | 487                |
| <i>Rodina 9</i>  | <i>ALDH9A1</i>  | NP_000687.3     | <i>HumALDH9A1</i>            | 518                |
| <i>Rodina 16</i> | <i>ALDH16A1</i> | NP_699160.2     | <i>HumALDH16A1</i>           | 802                |
| <i>Rodina 18</i> | <i>ALDH18A1</i> | NP_002851.2     | <i>HumALDH18A1</i>           | 795                |

## 3.2 Metody

### 3.2.1 BLAST

Pro vytvoření sekvenční databáze ALDH genů pro webové rozhraní, byly prohledávány rostlinné genomy pomocí softwaru BLAST. Jednalo se převážně o verze *blastp* (protein-protein blast) a *tblastn* (protein-translated nucleotide blast). Z velké části jsem užíval tento software skrze weby NCBI (The National Center for Biotechnology

Information) a Phytozome, ale využil jsem také stránek Fernbase nebo Sol Genomic Network, což jsou stránky genomových projektů pro kapradiny a užitkové rostliny jako například rajče (*Solanum lycopersicum*).

The screenshot shows the NCBI BLAST search interface with the following elements and annotations:

- 1)** Input field for "Enter accession number(s), gi(s), or FASTA sequence(s)".
- 2)** "Query subrange" section with "From" and "To" input fields.
- 3)** "Or, upload file" section with a "Vybrat soubor" button and "Soubor nevybrán" text.
- 4)** "Align two or more sequences" checkbox.
- 5)** "Database" dropdown menu set to "Non-redundant protein sequences (nr)".
- 6)** "Organism" input field with "Add organism" button.
- 7)** "Program Selection" section with radio buttons for "Quick BLASTP", "blastp", "PSI-BLAST", "PHI-BLAST", and "DELTA-BLAST".

At the bottom, there is a "BLAST" button and a "Search database nr using Blastp (protein-protein BLAST)" option, with a checkbox for "Show results in a new window".

**Obr. 13 Rozhraní protein-protein BLASTu na stránkách NCBI.** Různé webové implementace BLAST se mohou graficky lišit, principiálně však fungují stejně. Software přijímá proteinové sekvence (1), kdy samotný alignment lze omezit jen na zvolenou část (2). Je možné také nahrát sekvence v podobě souboru ve formátu FASTA (3). Kromě volby databáze, vůči níž chceme blastovat (5), nabízí program možnost vlastního porovnávání sekvencí (4). Po zvolení databáze můžeme specifikovat organismus, jehož genom bude prohledáván (6). Nakonec existuje výběr alternativních verzí BLAST, které je možno využít mimo základní blastp (7).

### 3.2.1.1 Program BLAST

BLAST neboli Basic Local Alignment Search Tool je program, který přijímá jednotlivé sekvence nebo rovnou celé soubory sekvencí ve formátu FASTA. Dále většinou obsahuje možnost vybrat si databázi, vůči které chceme vyhledávat a další volitelné parametry. Výstup může program poskytovat v rozličných formátech, jako je textový soubor, ale i HTML nebo XML formát. Webový NCBI BLAST využívá HTML formátu, kde udává sekvence z databáze seřazené podle skóre podobnosti a také umožňuje uživateli si prohlédnout samotný alignment neboli srovnání vstupních a databázových sekvencí.



BLAST je tedy možné užívat skrze webový server NCBI, kde jsou k nalezení genomové databáze většiny osekvenovaných organismů, proti nimž lze následně prohledávat. Dostupný je také zdarma ke stažení v podobě aplikace, pro uživatele, kteří si chtějí například vytvořit vlastní databázi, ve které budou dále hledat podobné sekvence. Tato aplikace funguje lokálně skrze příkazový řádek. Jelikož je BLAST open-source, tak během let bylo vytvořeno mnoho modifikací základního programu, které umožňují práci s rozdílnými sekvencemi, ale i různé způsoby porovnávání.

### aldehyde dehydrogenase [Coccomyxa subellipsoidea C-169]

Sequence ID: [XP\\_005642981.1](#) Length: 486 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 6 to 481 [GenPept](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

| Score          | Expect  | Method                       | Identities   | Positives    | Gaps      |
|----------------|---|------------------------------|--------------|--------------|-----------|
| 636 bits(1641) | 0.0   | Compositional matrix adjust. | 303/477(64%) | 370/477(77%) | 1/477(0%) |
| Query 59       | QLLINGNFVDSASGKTFPTLDPRTEVIAHVAEGDAEDINRAVKAARTAFDEGPWPKMSA   |                              |              |              | 118       |
| Sbjct 6        | QLLING F D++ GKTF T+DPRTGE + VAE AED++RAVKAAR AFD GPWP+MS<br>QLLINGKFEDASGGKTFETMDPRTGEPLMTVAEAQAEDVDRAVKAARQAFDHGPWPRMSG   |                              |              |              | 65        |
| Query 119      | YERSRVLLRFADLVEKHSEELASLETWDNGKPYQQLTAEIPMFARLFRYYAGWADKIHG   |                              |              |              | 178       |
| Sbjct 66       | +R ++ + A L+EK++EELA+LE+ DNGK Y + + ++PM RYYAGWADKI+G<br>RQRGNIMHKLATLMEKNTTELATLESLDNGKAYSAAF SIDVPMAVEHLRYYAGWADKIYG      |                              |              |              | 125       |
| Query 179      | LTIPADGNYQVHTLHEPIGVAGQIIPWNFP LLMFAWKVGPALACGNTIVLKTAEQTPLTA   |                              |              |              | 238       |
| Sbjct 126      | TIP DG Q +TL EP+GV GQIIPWNFP+LM AWK+GPALA GNTIV+K AEQTPL+A<br>QTIPTDGKMQAYTLKEPLGVVQIIPWNFPILMQAWKLGPALAAGNTIVMKVAEQTPLSA   |                              |              |              | 185       |
| Query 239      | FYAGKLFLEAGLPPGVNLIVSFGFATAGAALASHMDVDKLAFTGSDTGTGKVIILGLAANSN  |                              |              |              | 298       |
| Sbjct 186      | G+L LEAGLPPGVNLNI+ G G AGAALASH +DK+AFTGST+ GK+I+ AA N<br>LRVGEALALEAGLPPGVNLNIIPGDGPVAGAALASHKIDKIAFTGSTEVGKIIMRQAAE-N     |                              |              |              | 244       |
| Query 299      | LKPVTLELGGKSPFIVFEDADIDKAVELAHFALFFNQGCCAGSRTFVHEKVVDEFVEK  |                              |              |              | 358       |
| Sbjct 245      | + PVTLELGGKS I+ DAD+D+AV AH ALFFN GCC AGSRTFVHE +YDEFV +<br>VIPVTLELGGKSACIICPDADLDEAVRGAHEALFFNHGCC TAGSRTFVHESIYDEFVAR    |                              |              |              | 304       |
| Query 359      | SKARALKRVVGDPPFRKGI EQGPQIDLKQFEKVMKYIKSGIESNATLECGGDQIGDKGYFI  |                              |              |              | 418       |
| Sbjct 305      | + A R VGDPP + +QGPQ+ +Q++K+M I +G+E A LE GG + G++GYF+<br>AAKLASGRRVGDPPDRATQGPQVQSQEYQDKIMGLISTGVEQGAKLETGGKRRHGERGYFV      |                              |              |              | 364       |
| Query 419      | QPTVFSNVKDDMLIAQDEIFGVPVQSILKFSVDVEVIKRANETKYGLAAGVFTKNLDTANR   |                              |              |              | 478       |
| Sbjct 365      | QPTVFSNV DDM IA DEIFGVPVQSILK+S VDEVI+RAN T+YGLAAGVFT+NL AN<br>QPTVFSNVTDMMIATDEIFGVPVQSILKWSVDEVIIRANNTTEYGLAAGVFTQNLAMANT |                              |              |              | 424       |
| Query 479      | VSRALKAGTVWVNCDFVDAIIPFGGKMSGNGREKGIYSLNLYLQIKAVVTALNKP   |                              |              |              | 535       |
| Sbjct 425      | +SRALKAGTVWVN ++ FDA +PFGGK SG GREKG +L++Y Q KAV + +P<br>ISRALKAGTVWVNTWNQFDAGVPFGGKYKSGIGREKGEAALSHYTQTKAVYMPMAEP          |                              |              |              | 481       |

**Obr.14 Výstup NCBI BLAST v podobě alignmentu dvou sekvencí.** Tento výstup lze zobrazit pro všechny softwarem nalezené významné shody. V tomto případě byla prohledávána databáze sekvencí řasy *Coccomyxa subellipsoidea* pomocí ALDH2B4 z *Arabidopsis thaliana*.

### 3.2.1.2 Verze programu BLAST

Nejběžnějšími verzemi programu BLAST jsou nukleotid-nukleotid BLAST (blastn) a protein-protein BLAST (blastp). Tyto verze umožňují porovnávat uživatelem zadané a v databázi uložené nukleotidové, respektive proteinové sekvence.

Dalším odvětvím softwarů jsou ty, které pracují s abstraktními translačními produkty šesti čtecích rámců nukleotidových sekvencí. Jedná se o blastx (Nucleotide 6-

frame translation-protein BLAST), který porovnává tyto produkty zadané nukleotidové sekvence se sekvencemi z proteinové databáze a tblastn (Protein-nucleotide 6-frame translation BLAST), který naopak porovnává proteinovou sekvenci se čtecími rámci sekvencí nukleotidové databáze. Existuje také tblastx (Nucleotide 6-frame translation-nucleotide 6-frame translation BLAST), ten pracuje s šesti čtecími rámci jak u hledané sekvence, tak u sekvencí nukleotidové databáze, vůči níž je spuštěn. Tato verze je schopná odhalit i vzdálené vztahy mezi sekvencemi, je však značně pomalejší než jiné alternativy. Programy, které pracují s nukleotidovými sekvencemi a jejich čtecími rámci, poskytují obecně přesnější výsledky, jelikož tyto sekvence bývají méně konzervované oproti sekvencím proteinovým.

Verze programu BLAST, které pozměňují samotný proces prohledávání jsou následující.

**Position-Specific Iterative BLAST (PSI-BLAST, blastpgp)** – tato verze programu je jednou z nejpřesnějších, jelikož dokáže identifikovat i velmi vzdálenou evoluční příbuznost jednotlivých proteinů. Princip je takový, že první běh algoritmu probíhá obdobně jako u ostatních verzí BLASTu, kdy jsou nalezeny blízké příbuzné proteiny, pro uživatelskou sekvenci. Z nich se pak zformuje profilová sekvence, obsahující zásadní znaky všech sekvencí identifikovaných při prvním běhu algoritmu. Proběhne druhý běh algoritmu, tentokrát s vytvořenou profilovou sekvencí a celý proces se opakuje.

**Context-Specific BLAST (CS-BLAST)** – verze BLAST, která umožňuje nalézt až dvojnásobné množství vzdáleně příbuzných proteinových sekvencí než základní verze tohoto programu. Zachovává si však stejnou rychlost či chybovost. Je založen na lokálním kontextu sekvence, na němž závisí pravděpodobnost mutace jednotlivých aminokyselin.

Existují však také akcelerované verze programu BLAST. Například společnost TimeLogic poskytuje Tera-BLAST, kdy se jedná o FPGA-akcelerovanou implementaci tohoto algoritmu, která využívá logické integrované obvody a v rychlosti stokrát předčí základní algoritmus. Další verzí je například G-BLASTN, což je GPU-akcelerovaná implementace blastn, jejíž zrychlení není tak markantní a hlavním omezením je nutnost vměstnat celou prohledávanou databázi do paměti grafické karty (Zhao a Chu, 2014). Mezi CPU-akcelerované implementace se pak řadí například QuickBLAST, vytvořený

NCBI, který využívá předfiltrování pomocí odhadů Jaccard indexu. Toto filtrování algoritmus značně urychlí, za cenu snížené citlivosti (Maden a Boratyn, 2017).

### 3.2.1.3 Algoritmus BLAST

Samotný algoritmus BLAST byl navržen v USA Stephenem Alchulem ve spolupráci s Warrenem Gishem, Webbem Millerem, Eugenem Myersem, a Davidem J. Lipmanem z NIH (National Institutes of Health) a publikován roku 1990 v *Journal of Molecular Biology* (Altschul *et al.*, 1990). Pro jeho úspěšný průběh je nutné určit hledanou sekvenci a cílovou databázi sekvencí, která bude prohledávána. BLAST hledá high-scoring segment páry (HSP) neboli alignment částí sekvencí z databáze a částí dotazované sekvence s vysokou identitou. Jelikož je BLAST užíván na prohledávání z pravidla obrovských genových databází, využívá heuristické přístupy vycházející ze Smith-Watermanova algoritmu, který by byl kvůli své časové složitosti nevhodný pro práci s větší databází. Pracují více jak 50krát rychleji než původní algoritmus, ale přesnost jejich alignmentu může být nižší (Mount, D. W., 2004).

### 3.2.1.4 Princip algoritmu

Proces algoritmu můžeme rozdělit na několik částí. Jako první dojde k odstranění tzv. „low-complexity regions“ a repetice z dotazované sekvence. To značí části sekvencí skládající se z malého množství rozdílných prvků a tandemové repetice, které by mohly dosahovat falešně vysokého skóre a algoritmus tak zpomalit nebo ovlivnit přesnost vyhledávání. Tyto oblasti jsou označeny znaky X (pro proteiny) a N (pro nukleotidy). Program je pak ignoruje.

Následně se vytvoří seznam k-prvkových podsekvencí neboli slov, které budou vyhledávány v databázi. Pro proteinové sekvence se většinou jedná o 3-prvková slova, pro nukleotidy je pak užíváno slov 11-prvkových. Pro ty je provedeno vzájemné porovnání a jsou určena jejich skóre za pomoci substitučních matic. Pro nukleové kyseliny se užívá matice IUPAC, též zvaná „identity matrix“, která hodnotí shody kladnou konstantní hodnotou, naopak neshody nulou či zápornou hodnotou. Proteinové sekvence jsou porovnávány na základě matic PAM nebo BLOSUM, které berou ohled například i na konzervované substituce jednotlivých aminokyselin. Slova, jejichž skóre dosahuje určité hranice jsou dále užita pro hledání v databázi, ostatní jsou odstraněna.

Program pak hledá přesné shody slov se sekvencemi v databázi. Takto nalezené úseky jsou rozšiřovány a hodnoceny opět pomocí matic. Dokud skóre alignmentu roste, pokračuje se v jeho rozšiřování. Úseky, které dosáhnou vyššího skóre, než je zadaná hranice se potom nazývají high-scoring páry (HSP). Když se skóre přestane zvyšovat, algoritmus končí a vydá všechny HSP nalezené v dané databázi (Mount, D. W., 2004; Xiong, J., 2006).

### **Maticе PAM („Point Accepted Mutation“)**

Tuto matici poprvé vytvořila Margaret Dayhoff a její tým roku 1978. Jedná se o nejstarší substituční matici, v bioinformatice užívanou na porovnávání proteinových sekvencí. Je vhodná pro alignment příbuzných proteinů, jelikož nezohledňuje příliš mnoho evolučních předpokladů. V matici každý sloupec a řádek představuje jednu z proteinogenních aminokyselin a každý záznam této matice pak udává pravděpodobnost, s jakou může být daná řádková aminokyselina nahrazena sloupcovou pomocí mutace během daného evolučního intervalu. Různé PAM pak korespondují s odlišnými časovými úseky. První matice byla vytvořena pro proteiny o sekvenční identitě 85% a vyšší. Byla nazvána PAM1 a slouží jako základ pro další matice. V praxi se nejčastěji užívají matice PAM30, PAM70 nebo PAM250 (Dayhoff *et al.*, 1978).

### **Maticе BLOSUM („BLOcks SUBstitution Matrix“)**

Empirické matice užívané pro alignment evolučně vzdálenějších proteinových sekvencí, založené na lokálním přiřazení konzervativních domén. Byly vytvořeny na základě skenu databáze BLOCKS, kdy se identifikovaly konzervované oblasti proteinových rodin a vypočítala se pravděpodobnost substituce jednotlivých aminokyselin. Číslování těchto matic pak bylo stanoveno na základě počtu konzervovaných domén v porovnávaných sekvencích. Algoritmus BLAST používá při svých výpočtech 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  1

```

**Obr. 15 Substituční matice BLOSUM62** užívaná softwarem BLAST. Jednotlivé aminokyseliny jsou ohodnoceny na základě podobností a pravděpodobnosti jejich substituce. Hodnota mezery (\*) je -4.

### 3.2.1.5 Dynamické programování

Algoritmus BLAST byl inspirován principy dynamického programování užívaných Smith-Watermanovým algoritmem, který stejně jako BLAST provádí lokální alignment sekvencí. I když BLAST nedosahuje vždy optimálního alignmentu, jako je tomu u Smith-Watermanova algoritmu, vynahrazuje to svou rychlostí, které se téměř žádná implementace původního algoritmu nevyrovná.

### Algoritmus Smith-Waterman

Tento algoritmus vznikl modifikací algoritmu pro globální porovnání sekvencí, zvaného Needleman-Wunch algoritmus, jež publikovali Saul B. Needleman a Christian D. Wunsch v roce 1970 (Needleman a Wunsch, 1970). V principu jsou tyto algoritmy téměř totožné, největšími rozdíly jsou, že Needleman-Wunch pracuje i se zápornými hodnotami ve svých výpočetních tabulkách a jeho výstup, tedy výsledné porovnání sekvencí, je vždy stejně dlouhý jako nejdelší ze zadaných sekvencí. Naopak Smith-Watermanův algoritmus, který vytváří lokální alignment, pracuje sice s celými sekvencemi, ale jeho výstupem může být i alignment různě dlouhých podsekvencí.

Princip algoritmu je tedy následující. Na vstupu přijímá dvě porovnávané sekvence a substituční matici, která slouží ke skórování podobnosti sekvencí, a tedy k dosažení optimálního alignmentu. Často se také specifikuje hodnota sankce za mezery. Následuje vytvoření tabulky o velikosti  $((n+1) \times (m+1))$ , kdy první z porovnávaných sekvencí délky  $n$ , je zapsána do prvního řádku tabulky, zatímco druhá sekvence o délce  $m$ , se nachází v prvním sloupci tabulky. Ta se následně vyplňuje podle vzorce:

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

Tedy hodnota určité buňky tabulky je buďto hodnotou její levé sousední buňky + sankce za vložení mezery (inzerce), nebo hodnotou její horní sousední buňky + sankce za vložení mezery (delece), nebo hodnota její levé horní diagonálně sousední buňky + skóre shody nebo neshody sekvenčních znaků na dané pozici. Pokud by mělo dojít k tomu, že skóre bude záporné, zapíše se místo něj nula. Při grafickém znázornění algoritmu se uvádí také směr průchodu tabulkou pomocí šipek.

|   |   |   |    |    |   |   |   |
|---|---|---|----|----|---|---|---|
| - | - | 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. 16 Výpočetní tabulka Smith-Watermanova algoritmu.** Šipkami je znázorněno, ze kterých sousedních buněk vychází ohodnocení každého nenulového políčka tabulky. Pro nukleotidové sekvence ATCGAA a CATAC je optimální alignment zaznačen červenými šipkami s maximální hodnotou skóre 11.

### 3.2.1.6 BLAST alternativy

Pro hledání podobnosti mezi proteinovými i DNA sekvencemi, lze využít mnoho dalších programů. FASTA, jakožto předchůdce BLASTu, který byl vytvořen Davidem J. Lipmanem and Williamem R. Pearsonem roku 1985, disponuje podobnými funkcemi (Lipman a Pearson, 1985). Využívá také heuristické přístupy pro dosažení vhodného alignmentu. Na rozdíl od BLASTu však neprohledává pouze high-scoring segmenty, a tedy i když dosahují oba algoritmy srovnatelné přesnosti, BLAST je časově úspornější.

FASTA však navíc poskytuje větší výběr substitučních matic a umožňuje tedy prohledávání s širším rozsahem evolučních vzdáleností.

BLAT (Blast Like Alignment Tool) je rychlejší, ale méně přesná alternativa BLASTu. Místo lineárního prohledávání využívá indexování databáze pomocí k-merů, čímž může nalézt HSP rychleji (Kent WJ, 2002). Pro porovnávání databází je pak nejvhodnější PLAST (parallel local alignment search tool), jelikož je schopný porovnat velké množství sekvencí za využití malého množství výpočetní paměti (Lavenier D., 2009). V metagenomice se využívá DIAMOND, pro porovnávání milionů sekvencí současně, který je až 20 000krát rychlejší než BLASTX, ale stále si dokáže udržet relativně vysokou míru senzitivity (Buchfink *et al.*, 2015).

### 3.2.1.7 Využití BLASTu

BLAST může být součástí mnoha procesů, jako jsou například identifikace neznámých druhů, kdy nám program umožňuje najít homologní druhy k DNA neznámého druhu. Dále můžeme BLAST využít pro lokalizaci známých domén zvolené proteinové sekvence nebo pro mapování DNA, kdy, pokud chceme sekvenovat gen známého druhu na neznámé chromozomální pozici, je možné pomocí BLASTu srovnat pozici zadané sekvence se sekvencemi z databáze. Program lze také užít při stanovení fylogeneze a tvorbě fylogenetických stromů, jelikož například NCBI BLAST má tuto funkci zabudovanou. Takto získaná fylogenetická data však nemusí být nejpřesnější, a proto je vhodné využívat specializovaný software jako například PhyML.

### 3.2.2 MUSCLE

Pomocí programu MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>) byl proveden multiple alignment neboli mnohonásobné porovnání všech sekvencí uložených ve vytvořené databázi, jako příprava pro fylogenetickou analýzu. Software MUSCLE neboli MULTiple Sequence Comparison by Log-Expectation byl poprvé publikován roku 2004 Robertem C. Edgarem skrze Nucleic Acids Research (Edgar R. C., 2004). Slouží k mnohočetnému přiřazení proteinových i nukleotidových sekvencí, které musí uživatel zadat ve formátu FASTA. Jelikož se jedná o software pro globální alignment, sekvence by měly být evolučně příbuzné a přibližně stejně dlouhé, pro dosažení optimálních výsledků. Kromě využití v procesu fylogenetické analýzy může být MUSCLE použit také pro predikci struktury nebo identifikaci residuí.

**STEP 1 - Enter your input sequences**

Enter or paste a set of sequences in any supported format:

1)

Or upload a file:  ALDH5\_M...CLE\_obr.fas 2)

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

---

**STEP 2 - Set your Parameters**

OUTPUT FORMAT: 3)

ClustalW

The default settings will fulfill the needs of most users. 4)

(Click here, if you want to view or change the default settings.)

---

**STEP 3 - Submit your job**

Be notified by email (Tick this box if you want to be notified by email when the results are available)

**Obr. 17 Interface programu MUSCLE.** Stejně jako je tomu u programu BLAST, i zde je možné zadat vstupní sekvence buďto manuálně (1) nebo v podobě FASTA souboru (2). Dále nám program dává volbu výstupního formátu (3). Po stisknutí tlačítka „More options“ (4) se objeví také možnost zobrazit pomocné stromy, které algoritmus vytváří během své první či druhé fáze (tzv. guide tree).

Algoritmus lze rozdělit do tří kroků. Cílem první fáze je velmi rychle vytvořit konceptuální mnohočetné přiřazení i za cenu snížené přesnosti. Nejdříve je vytvořena vzdálenostní matice k-merů každého porovnávaného páru sekvencí. Na základě této matice algoritmus UPGMA vytvoří binární strom a podle vzniklého větvení je zkonstruován prvotní alignment. Ten slouží jako základ pro další kroky algoritmu. Při druhé fázi se konceptuální alignment zpřesní pomocí Kimurovi dvouparametrové vzdálenosti a je vytvořena nová vzdálenostní matice. Z té je stejným způsobem vyprodukován sekundární alignment. Toto nové porovnání pak respektuje také transnitní a transversní substituci purinu. V poslední etapě se multiple alingment ještě zpřesní za cenu časové složitosti. Binární strom z druhé fáze algoritmu je rozdělen na dva, odstraněním vybrané hrany. Pro tyto dva stromy se vytvoří jejich profily, jež se znovu porovnájí mezi sebou. Pokud výsledný alignment všech sekvencí dosahuje určitého skóre, pracuje se s ním dále a kroky třetí fáze se opakují, dokud algoritmus nedosáhne terminační podmínky. MUSCLE tvoří alignmenty z velké části přesněji a rychleji než rozšířenější software Clustal a je tedy vhodnou alternativou tohoto programu (Edgar R. C., 2004).



Program MUSCLE nabízí uživateli dva typy výstupních formátů. Jedná se o tzv. „interleaved“ formáty, mezi které se řadí například CLUSTAL. Sekvence jsou pak zapsány v podobě přiřazení pod sebou. Další variantou jsou „sequential“ formáty jako je FASTA nebo PHYLIP 3.2, zde jsou celé sekvence uloženy odděleně i s nově přidanými mezerami. Pro následnou analýzu sekvencí jsem využíval formát FASTA.

## Results for job muscle-I20210323-125527-0191-80544514-p1m

Alignments Result Summary Guide Tree Phylogenetic Tree Results Viewers Submission Details

Download Alignment File Show Colors

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

AfALDH5H1 -----MIRDRAARGTSAQGRGRSRRGSPL
AfALDH5H2 -----QDFLIIVTRSKV-----
CsALDH5F1 -----MAGLHRLSLTSLRHRQMMLTSI
KnALDH5F1 -----MNVKQARIILGFRQLFCHRSAGATSIASQTSGARYSRSC
AfALDH5F1 MMMVFYSRRGWDGFLLVQKESACVPHPDLIHSSSFLLTAAAAVPAALNPSSTPPLSSSST
SmALDH5F1 -----MATIASA-----
AtALDH5F1 -----MVIGAAARVAIGGCRKLISSTLSL
MdALDH5F1 -----MTFRASSMAGRSFKFLSHLPYMHAPT
PrpALDH5F1 -----MTFRASRMSARSKLFLTHLPSMQAPP
ZomALDH5F1 -----MATALYRVILHGSKTSSYTRFLC
ZmALDH5F1 -----
ZmALDH5F2 -----MAMAMMAMRRAVALGARHIPAAAASS

AfALDH5H1 ERKTATSRSLTFPCLVSWG-----LTRPSRVSPRERRGHKERPRIAPRPFEDRLR
AfALDH5H2 -----FNRGIPMDIVK
CsALDH5F1 SRLQQFVGTQAFHV-----TAAEELLNKVV
KnALDH5F1 VVPSNIQDVHSTLHSRCGSGGGGAGKRDLIGRFTQRTFHLTPTCQMALDGSNSVVKLK
AfALDH5F1 VMQASLSKLRTLHHASKAITTGFAQLRYLST-----PSDEAVIHKIK
SmALDH5F1 -----GASQAVLDKIN
AtALDH5F1 LLVSSQCRQ-----MS-----MDAQSVSEKLR
MdALDH5F1 PLSRQLQHPHTLNCEQAALL-----IS-----MDAKSFVSQLN
PrpALDH5F1 LLSRQ-----IS-----MDVKSFAARLS
ZomALDH5F1 RRMMSV-----GS-----VVSEDAISRLLR
ZmALDH5F1 -----MS-----ADASAAMEKIR
ZmALDH5F2 FRVSLRH-----MS-----ADAGAAMEKIR
:

AfALDH5H1 DHSDAQAEILDGVVIEAVFRAFAALAGLLDATERGDFGGDQAGVSDSHAIFEFLGNPED
AfALDH5H2 DASLFREA---FYS DGAWHKG-GADVIQVTPATGKVIQVTPKATTEQTRAAIEAAHAA
CsALDH5F1 DKDLVKTL---GFINGEWVSASDGSTIPVDNPATGEVIANVAACLKASDTKSAIAAASAA
KnALDH5F1 DPSLLKDL---GLIGGKWVGAHDGKTPVHNIPATGEVVAEVLPGAQETAIAAIAADEA
AfALDH5F1 EAGLLQSK---GLIGGKWVGAHDGKTLP-----EAKYFKLRA
SmALDH5F1 NAGLLRAK---GLISGEWVSAEDGRTLVPYNPATGEFLTEVPLMGERETLSAIDSAHNA
AtALDH5F1 SSGLLRDQ---GLIGGKWLDSDYDNKTIKVNNPATGEIADVACMGTKETNDAIASSYEA

```

**Obr. 18 Výstup programu MUSCLE s alignmenty v ClustalW „interleaved“ formátu.** Kromě samotného alignmentu program nabízí možnost zobrazit „guide tree“ nebo dokonce fylogenetický strom, který však nemusí být nejpřesnější. Sekvence rodiny ALDH5 jsou částí sekvencí z databáze. Předpony odpovídají rostlinným zástupcům. Použité zkratky: Af – *Azolla filiculoides* (kapradina), At – *Arabidopsis thaliana* (huseníček rolní), Cs – *Coccomyxa subellipsoidea* (řasa), Kn – *Klebsormidium nitens* (řasa), Md – *Malus domestica* (jabloň), Prp – *Prunus persica* (broskvoň), Sm – *Selaginella moellendorffii*, Zm – *Zea mays* (kukuřice), Zom – *Zostera marina* (vocha mořská).

```

>AfALDH5H1
-----MIRDRAARGTSAQGRGRSRRGSPL
ERKTATSRLTFPCLVSWSG-----LTRPSRVSPRERRGHKERPRIAPRPFEDRLR
DHSDAQAEILDLGVVIEAVFRAFAALAGLLDATERGDFGGDQAGVSDHAIFFELGNPED
FRDVARVEIGRQAEI-----
-----GVVGHGNDLFLGVELEQRRDGAEGFLAGDLHLSGDAGEDGRLEEQAALRIRL
AADRHSGALGHGIGDVF LDLVDGGAGDQRTGIDAGRKAVADLHRRHLGGQLFGEGIIDAG
LDIEAVGA---DAGLAVVTVFGDHRAFDGGIKIGVVKDDERRIAAQFQAEHLDLVGALA
HQDLADTGRAGKQLADLVILGELLADRRRILRSDDVENAGGQAGAFGQQRQRQRCL
GGGLGDHGAAGGHRRRELAGQHGVGEVPRGDRGDDADRLDDDDALVRLVAGNGVAIDAL
GFLAEPLDIGGAI EHLALGLGHRLAHFHGQDGGEVIGIGDHQVVE-----
LAQHGGAFLAGLLGPDFLRGVGGIDGAGGLCAVHV-----GQRGDDST-----
-----IGRIGDIKGLAAVALGNPFARDQRLVSSQRR
VLEL-----GQQSFGNRHGSGLRSAFAQCVFENIHRFGKHCILDGDRRQEADDIAIDAAG
QEDQPLGQRLF AHLVGEIGIRGLGARNRLQALLEAVAERHGPFVEALFLDDVEHGIAGSD
ADRVAAI GAAEAAGAGRIHDVGAAGDGS-----
-----
---REAAAEALGHGGDVGDNVAMFHRKQFSRAGKA-----
-----

```

**Obr. 19** Příklad sekvence z FASTA souboru vytvořeného programem MUSCLE. Soubor obsahuje celé sekvence včetně mezer, zaznamenané postupně za sebou.

### 3.2.2.1 MUSCLE alternativy

Nejvyužívanějšími programy pro multiple alignment, jsou softwary Clustal. Ty dosahují optimálních výsledků ve většině případů, avšak excelují při porovnávání sekvencí s různou divergencí. Vzniklo mnoho jejich variant, princip algoritmu však zůstává stejný. Program nejdříve vytvoří párové alignmenty pomocí progresivních srovnávacích metod. Následně zkonstruuje strom těchto alignmentů a pomocí něj tvoří finální mnohočetné porovnání sekvencí. V současné době se nejčastěji využívají verze ClustalW, která je součástí například softwaru Bioedit a Clustal Omega, který je obecně považován za jednu z nejrychlejších variant tohoto programu, speciálně pro velké datasey (Higgins a Sharp, 1988). Mimo tuto řadu softwarů se však používají i další alternativy.

#### **T-coffee** (Tree-based Consistency Objective Function for Alignment Evaluation)

Jedná se o software pro multiple alignment, který nejprve generuje knihovnu párových porovnání všech zadaných sekvencí a na jejím základě potom sestavuje výsledné mnohočetné porovnání. Existuje mnoho variant tohoto programu, které umožňují například vzájemné porovnávání proteinových a RNA sekvencí, práci s 3D strukturami nebo také sloučení dvou a více dříve vytvořených multiple alignmentů. Tyto možnosti jsou hlavní výhodou T-coffee v porovnání s programem MUSCLE. Dosahuje

však také vyšší rychlosti a přesnosti, pokud pracuje s evolučně příbuznými sekvencemi (Notredame *et al.*, 2000).

### MAFFT (multiple alignment using fast Fourier transform)

Software pro vícečetné porovnání nukleotidových i proteinových sekvencí. Původní algoritmus využíval rychlé Fourierovy transformace pro seřazení sekvencí a byl dále obohacován o nové funkce. Nyní je MAFFT využíván převážně pro alignment velmi velkého množství sekvencí, kdy rychlostí předčí ostatní programy (Kato *et al.*, 2002).

### 3.2.3 Gblocks

Vzniklý multiple alignment byl dále upraven, pomocí programu Gblocks ([http://molevol.cmima.csic.es/castresana/Gblocks\\_server.html](http://molevol.cmima.csic.es/castresana/Gblocks_server.html)). Ten umožňuje odstranit ze sekvencí úseky s nízkou podobností a dlouhé mezery, a tedy zpřesnit následnou fylogenetickou analýzu. Program pracuje se sekvencemi ve formátu FASTA. (Castresana, 2000)

**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. 20 Rozhraní programu Gblocks.** Program přijímá alignment sekvencí ve formátu FASTA, který je možné zadat manuálně (1) nebo v podobě připraveného souboru (2). Pro správnou funkci softwaru je nutné zvolit typ sekvence (3). Nakonec je také možné povolit méně striktní porovnávání či možnost ponechávat ve finálním souboru i kratší signifikantní alignmenty nebo mezery, a tedy odfiltrovat pouze opravdu nevhodné části sekvencí (4).

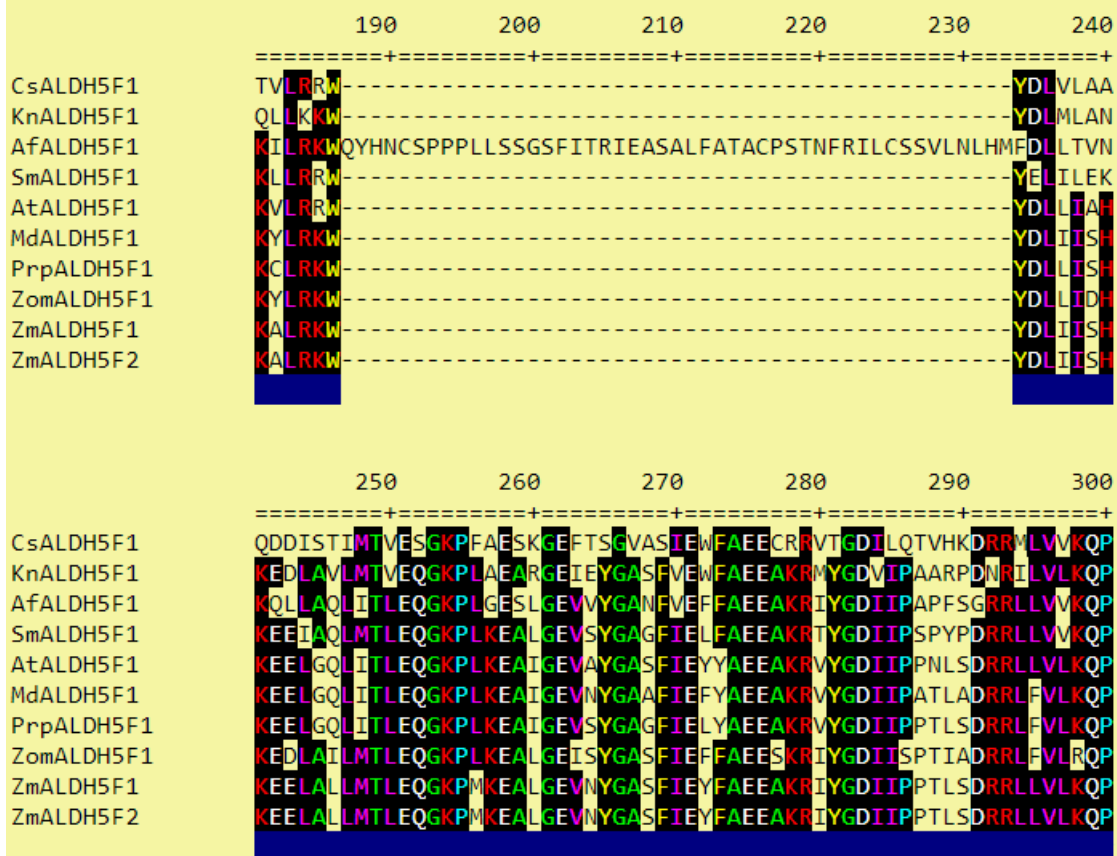
## Gblocks 0.91b Results

Processed file: ALDH5\_fasta2.fas

Number of sequences: 10

Alignment assumed to be: Protein

New number of positions: 428 (selected positions are underlined in blue)



**Obř. 21 Grafický vřstup programu Gblocks.** Modřými řarami jsou zaznařeny řásti sekvencí, které program vydá ve vřstupnřm FASTA souboru. Do vřslednřho souboru se tedy dostanou pouze ty řásti, které dosahujř dostateřnř podobnosti.

### 3.2.4 PhyML

Samotnř fylogenetickř analřza sekvencí z databřze, byla po jejich předchozř úpravř provedena pomocř programu PhyML (Phylogenetic estimation using Maximum Likelihood). Program přijimř soubory formřtu PHYLIP a na zřkladř multiple alignmentu sekvencí z třchto souborř pak konstruuje fylogenetickř stromy metodou maximum likelihood. Software nejdřive vygeneruje matici evoluřnřch vzdřlenostř, na zřkladř zadanřch sekvencí. Pomocř třto matice sestavř poãateãnř strom, pro kterř se vypoãitř jeho pravdřpodobnost a takř pravdřpodobnost vřch jeho podstromř. Souãastř třchto

výpočtů jsou i hodnoty zvolených substitučních modelů. Program poté zkouší všechny možné změny dendrogramu. Algoritmus vydá finální strom a končí, pokud již není možné vyzkoušet jiné uspořádání a délka jeho větví je stabilní (Guindon a Gascuel, 2003).

**PhyML online execution**

---

**Input Data**

Sequences (PHYLIP format)  Soubor nevybrán **1)** File  Example (DNA file) (from Phylogenetic Handbook)

Data Type **2)** DNA  Amino-Acids

---

**Substitution Model**

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

If you use SMS, please cite:  
"SMS: Smart Model Selection in PhyML."  
Vincent Lefort, Jean-Emmanuel Longueville, Olivier Gascuel.  
Molecular Biology and Evolution, 34(9):2422-2424, 2017.

---

**Set by user**

**4)** Substitution model  optimized  empirical

Equilibrium frequencies fixed  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

Number of random starting tree  yes  no

---

**Branch Support**

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

Perform bootstrap **5)**  yes  no

Name of your analysis

Your email

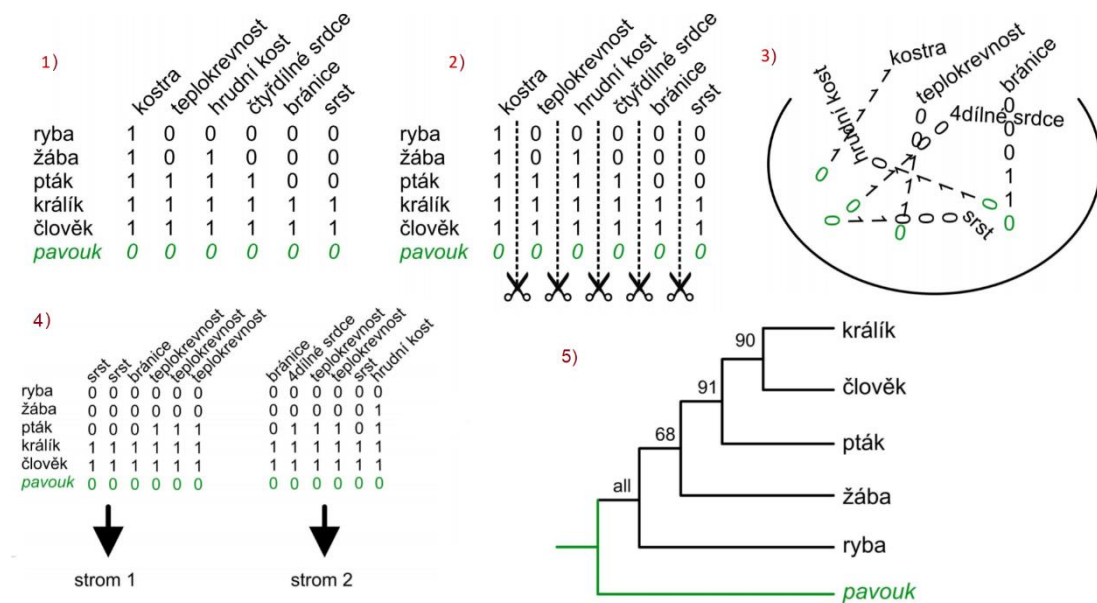
Please confirm your email

**Obr. 22 Interface programu PhyML.** Program přijímá soubory sekvencí ve formátu PHYLIP (1). Dále je nutné zvolit, zda se jedná o nukleotidovou či proteinovou sekvenci (2). Program nabízí také výběr automatických voleb substitučního modelu AIC a BIC (3), ale je možné zvolit model také manuálně s možností bližších specifikací (4). Nakonec máme také na výběr z několika možností ohodnocení větví tvořených fylogenetickými stromy. Nejvyužívanější je bootstrapping nebo test přibližného poměru pravděpodobností (5).

Program nabízí mnoho substitučních modelů, které se využívají pro výpočet pravděpodobnosti substituce v jednotlivých větvích stromu, a jsou tedy součástí výpočtu pravděpodobnosti celého dendrogramu. Jedná se jak o velmi rychlé metody s nižší přesností, tak i metody značně pomalejší, avšak mnohonásobně spolehlivější. V minulosti bylo využíváno modelů jako WAG a JTT, avšak v současnosti je nejpoužívanějším substitučním modelem LG, který je značně přesnější oproti svým předchůdcům (Le a Gascuel, 2008). PhyML však nabízí také možnost nechat výběr vhodného modelu na programu. Tyto možnosti jsou AIC (Akaike Information Criterion) a BIC (Bayesian Information Criterion), které využívají penalizační hodnoty pro určení ideálního modelu pro fylogenezi, aby nedocházelo ke ztrátě dat. Pro fylogenezi databázových sekvencí jsem využil možnosti AIC. Ta vybírá model na základě výpočtu definovaného následujícím vzorcem:  $AIC = 2k - 2\ln(L)$ . Očekávaný počet parametrů v daném modelu

je značen  $k$  a  $L$  označuje hodnotu jeho pravděpodobnostní funkce. Pokud máme tedy množinu kandidátních modelů pro určitá data, preferuje se model s nejnižší hodnotou AIC. Princip výpočtu BIC je obdobný, pouze využívá odlišné penalizační hodnoty, a to  $\ln(n)k$  místo  $2k$  (Burnham a Anderson, 2004).

Pro ohodnocení větví vzniklých fylogenetických stromů využívá program metody bootstrapping (obr.23) nebo test přibližného poměru pravděpodobností (the approximate likelihood ratio test). Pro analýzu sekvencí v této práci bylo použito metody bootstrapping.

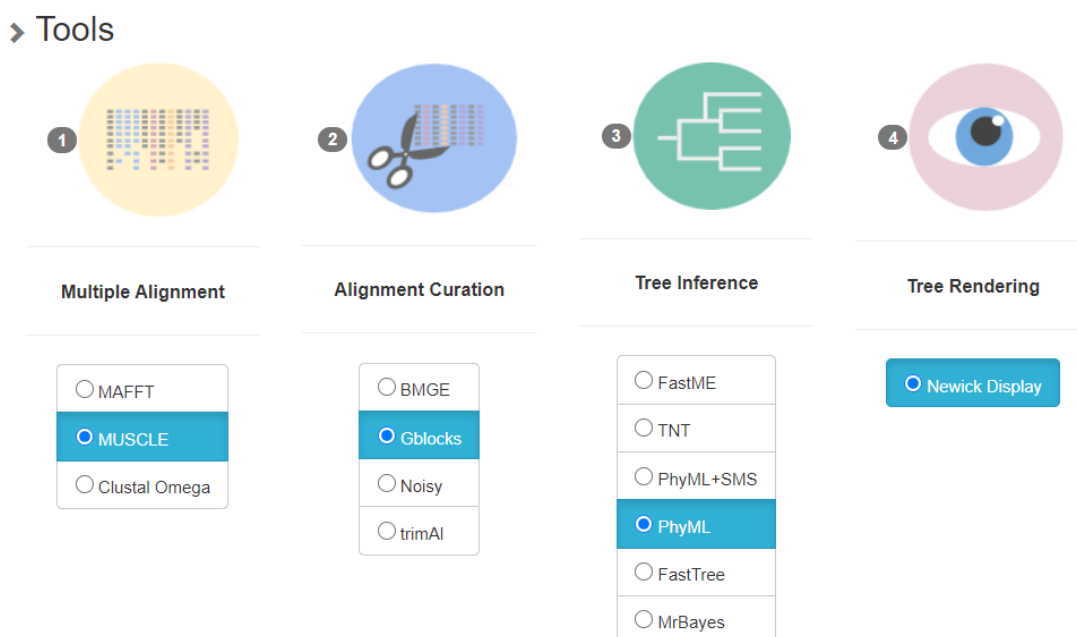


**Obr. 23 Princip bootstrappingu.** Alignment sekvencí můžeme přirovnat k tabulce živočišných druhů, kde každý druh obsahuje určité znaky (1). Nejprve dojde k rozdělení jednotlivých sloupců těchto znaků (2) a bude probíhat náhodný výběr z jejich množiny (3). Vybráním stejného počtu znaků jako měly původní tabulky vznikají „pseudomatrix“, v nichž se mohou jednotlivé znaky opakovat (4). Takových matic se vytvoří zadaný počet, společně s odpovídajícími fylogenetickými stromy. Procento zastoupení jednotlivých znaků v každé z těchto matic je pak zaznamenáno ve výstupním stromě, který program pracující s bootstrappingem vydá (5).

### 3.2.5 NGPhylogeny

NGPhylogeny je software, usnadňující a urychlující proces fylogenetické analýzy, který je dostupný na webu <https://ngphylogeny.fr/>. Stránka nabízí přehledný interface, který uživateli umožňuje provést fylogenetickou analýzu pro dané data, od multiple alignmentu až po tvorbu fylogenetických stromů, s možnostmi výběru rozličných metod pro jednotlivé kroky analýzy. Uživatel si může také zobrazit výstupy jednotlivých mezikroků celého procesu (Lemoine *et al.*, 2019). Mimo realizaci hlavního postupu fylogenetické analýzy, kdy byly sekvence z databáze analyzovány pomocí MUSCLE, GBlocks a následně PhyML, byl tento software využit také pro testování alternativních

metod a jejich přesnosti vzhledem k námi zvolené analýze. Kromě programu MUSCLE byla využívána také možnost MAFFT a pro fylogenetické stromy byla provedena analýza pomocí TNT. Ta, na rozdíl od PhyML, užívá metody maximum parsimony, která vydá dendrogram s nejmenším počtem mutací potřebných k dosažení pozorovaného stavu (Goloboff a Catalano, 2016). FastME byla další testovaná možnost, která vytváří fylogenetické stromy na základě vzdálenostních matic, metodou minimální evoluce, jež se snaží minimalizovat délky větví (Desper a Gascuel, 2002).



**Obr. 24** Interface programu NGPhylogeny pro „A la Carte“ analýzu. Zde si uživatel může zvolit metody, využívané v jednotlivých krocích analýzy, které jsou pro něj nejvhodnější.

### 3.2.6 Tvorba webového rozhraní

Pro vývoj webového rozhraní PAFF, byl využit programovací jazyk PHP a Javascript (JS) společně s databázovým systémem MySQL. Webová stránka je realizována pomocí HTML (HyperText Markup Language) a CSS (Cascading Style Sheets). Po zadání sekvence do rozhraní a stisknutí tlačítka „Find Family“, je zavolán porovnávací algoritmus v jazyce PHP, pomocí technologie Ajax. Ten získává sekvence z MySQL databáze pomocí příkazů v jazyce SQL. Stránka také využívá metody „on key up událostí“, kdy po každém stisknutí klávesy je zavolán JS skript, který provádí různé změny. Pro testování funkčnosti webové aplikace jsem využíval softwarový balíček XAMPP. Ten obsahuje Apache HTTP Server, MySQL databázi přístupnou skrze rozhraní

phpMyAdmin, a interprety skriptů programovacího jazyka PHP či Perl (<https://sourceforge.net/projects/xampp/>). Pro inspiraci a vyladění implementace programu jsem využíval webových stránek <https://www.php.net/manual>, <https://www.w3schools.com/> a <https://stackoverflow.com/>.

### 3.2.6.1 PHP (PHP: Hypertext Preprocessor)

PHP je univerzální programovací jazyk. První verzi vytvořil a vydal Rasmus Lerdorf roku 1994 jako „Personal Home Page“. Jedná se o velmi flexibilní jazyk a je možné s ním pracovat téměř na jakémkoli operačním systému či platformě. Obsahuje komplexní sadu nástrojů pro práci s relačními databázemi skrze softwary jako MySQL. Tento jazyk byl koncipován především pro vývoj webových aplikací, ale může být použit i v mnoha jiných oblastech, například při vývoji grafických aplikací nebo jako skriptovací jazyk využívaný z příkazového řádku (php-cli). Nejčastěji se tedy PHP využívá na webových serverech, kdy jsou skripty spouštěny na straně serveru. Takové kódy pak dynamicky upravují webovou stránku a vytváří nový obsah (Lerdorf *et al.*, 2006). PHP bývá součástí různých balíčků pro vývoj webových aplikací. Nejznámějším je LAMP, jež kombinuje operační systém Linux, webový server Apache a MySQL databázový systém společně s PHP, někdy může PHP nahradit Python či Pearl (<https://whatis.techtarget.com/definition/LAMP-Linux-Apache-MySQL-PHP>).

### 3.2.6.2 MySQL

Jedná se o open-source software pro správu relačních databází, vytvořený pomocí jazyka C a C++, který využívá jazyk SQL (Structured Query Language) pro přidávání, úpravu a odebrání informací z těchto databází (Sheldon a Moes, 2005). Relační databáze sestávají z tabulek s daty, kdy jejich řádky obsahují dané záznamy a sloupce obsahují jednotlivé hodnoty či vlastnosti záznamů, mohou reprezentovat mimo jiné i relace mezi jednotlivými tabulkami (Codd, 1970). MySQL je víceuživatelský a obsahuje tedy nástroje pro komplexní správu přístupových práv. Program může být používán samostatně, kdy umožňuje uživateli práci s databázemi pomocí SQL příkazů, avšak častěji se využívá ve spojení s dalšími softwary, například pro vývoj aplikací s přístupem k relačním databázím (Sheldon a Moes, 2005). MySQL je například součástí softwarového balíčku XAMPP, skrze který je jeho používání poměrně jednoduché, mimo jiné i díky administračnímu nástroji phpMyAdmin.



### 3.2.6.3 JavaScript (JS)

JavaScript je esenciálním skriptovacím jazykem, pracujícím v prohlížeči na straně klienta, který umožňuje uživateli vytvářet interaktivní webové aplikace. Vyhodnocení kódu tedy zprostředkovává prohlížeč uživatele stránky na rozdíl od PHP, kde se kód interpretuje přímo na serveru. Jedná se o multiparadigmatický jazyk, který podporuje objektově orientované i událostmi řízené programování. JavaScript skripty bývají často součástí HTML kódu stránky a interagují s DOM (Document Object Model). Využívají se například pro animace objektů nebo pro interaktivní objekty jako videa či hry (Flanagan, 2006). Nejrozšířenější knihovnou tohoto jazyka je jQuery, která byla vytvořena pro usnadnění práce s HTML DOM elementy, CSS animací nebo Ajax aplikacemi (<https://jquery.com/>). Existují také tzv. API (Application Programming Interface), neboli rozhraní pro práci s rozličnými datovými strukturami či DOM (Flanagan, 2006).

### 3.2.6.4 Ajax (Asynchronous JavaScript and XML)

Jedná se o soubor webových technologií, využívaných pro tvorbu asynchronních webových aplikací. Což znamená, že aplikace mohou interagovat s daty na serveru na pozadí a neovlivňovat tak danou stránku a její chování. Je tedy možné dynamicky měnit obsah webu, aniž by bylo nutné celou stránku znovu načítat (Ullman a Dykes, 2007). V minulosti se pro práci s daty takovýmto způsobem využívalo výhradně XML, avšak v současnosti převládá formát JSON (JavaScript Object Notation) (<https://www.json.org/json-en.html>). Ajax obsahuje technologie HTML, CSS a DOM, pro dynamické zobrazení a práci s daty. Dále JSON, pomocí něž se uskutečňuje výměna dat se serverem, XMLHttpRequest pro asynchronní komunikaci a JavaScript, pomocí něhož tyto technologie interagují. Mnoho knihoven jako jQuery pak obsahuje abstrakce pro usnadnění práce s Ajax (Ullman a Dykes, 2007).

## 4. VÝSLEDKY

### 4.1 Vstup a funkce webového rozhraní

Bylo vytvořeno webové rozhraní PAFF (Plant ALDH Family Finder), jež přijímá proteinové či nukleotidové sekvence ve formátu FASTA. Pro vloženou sekvenci program automaticky vypočítá její délku na základě počtu znaků, kdy ignoruje header, molekulovou hmotnost, pomocí uložených hodnot pro každou proteinogenní aminokyselinu a izoelektrický bod, který počítá metodou bisekce. Software je schopen určit, zda se jedná o proteinovou či nukleotidovou sekvenci, je však také možné, nechat toto rozhodnutí na uživateli. Pro nukleotidové sekvence program dále nabízí možnost výběru ze tří čtecích rámců a intuitivního určení startovního kodonu. Pokud by sekvence obsahovala nepodporované znaky, zobrazí se uživateli upozornění. Rozhraní pak nabízí možnost vybrat z různých rostlinných oddělení a tříd, které jsou obsaženy v databázi, vůči nimž je možné prohledávat. Po stisknutí tlačítka „Find Family“ program prohledá databázi rostlinných sekvencí, pomocí sekvence uživatelské, na základě Smith-Watermannova algoritmu. Pokud by na vstupu nebyla nalezena žádná sekvence, program se nespustí a vydá chybovou hlášku.



1)

Mass: 0  
pI (Isoel.point): 0  
Length: 0  
2)

Choose type of sequence:  Protein  cDNA/DNA\* 3)

\*Choose your preferred translation method

Intuitive: Couldn't Find Start Codon

Frame 1: Sequence length: 0 AA

Frame 2: Sequence length: 0 AA 4)

Frame 3: Sequence length: 0 AA

Choose plant species: All ▾ 5)

Find Family 6)

**Obr. 25 Interface webového rozhraní PAFF pro přiřazení sekvencí ALDH do rodin.** Program přijímá jednotlivé sekvence ve formátu FASTA (1). Na straně pak můžeme pozorovat dodatečné informace jako délka sekvence, hmotnost či izoelektrický bod, které se dynamicky přepočítávají pro každou změnu sekvence (2). Je možné zvolit typ vkládané sekvence, program však dokáže rozlišit sekvence také automaticky (3). Pokud se jedná o sekvenci nukleotidů, je možné dále vybrat vhodný překladový rámec „Frame 1 – Frame 3“, nebo nalezení počátečního kodonu nechat na programu „Intuitive“ (4). Je zde také možnost zvolit pouze část prohledávané databáze, a to s možnostmi pro řasy, mechy, kapradiny, jednoděložné a dvouděložné rostliny (5). Nakonec je zde tlačítko pro spuštění samotného porovnávání sekvencí (6).

## 4.2 Výstup programu

Program může vydat tři možné výstupy. Pokud zjistí, na základě tří prvotních organismů, jejichž sekvence porovná se zadaným řetězcem, že uživatelská sekvence skutečně náleží do nadrodiny ALDH, prohledá celou databázi. A to buď rostlinnou, nebo zvířecí část, na základě podobnosti vstupu vůči prvotním sekvencím. V těchto případech je výstupem seznam rodin s maximálně pěti nejpodobnějšími sekvencemi k zadanému vstupu. Pokud byla na vstupu zadána rostlinná sekvence, program zahlásí „*Your sequence belongs most likely to the Plant ALDH Family XY*“, naopak pokud se jedná o sekvenci podobnou zvířecím zástupcům, vydá program „*Your sequence most likely doesn't belong to any known Plant ALDH Family, but into Animal ALDH Family XY*“. Pokud by neexistovala podobnost mezi zkoumanými sekvencemi vyšší než 60%, výstupem bude „*Your sequence belongs most likely to the ALDH Family XY but is not similar enough to any known subfamily. You might have discovered a new subfamily*“. Program poskytuje informace o identitě či podobnosti nalezených sekvencí a je schopen také zobrazit příslušný alignment výstupní a uživatelské sekvence. Mimo to, dokáže na základě tohoto porovnání, také zaznačit pozice katalyticky aktivních aminokyselin.

Your sequence belongs most likely to the Plant ALDH Family 2 B

- ▼ **Family 2** (closest homolog sequence identity is 79%)
  - ▶ **ZmALDH2B2**: Cover - 534/534, Id. - 79%, Sim. - 90% **NCBI ID**: NP\_001105891.1
  - ▶ **OsALDH2B1**: Cover - 534/534, Id. - 78%, Sim. - 90% **NCBI ID**: NP\_001057358.1
  - ▶ **SbALDH2B1**: Cover - 534/534, Id. - 78%, Sim. - 89% **NCBI ID**: BAB92019.1
  - ▶ **PtALDH2B7**: Cover - 534/534, Id. - 76%, Sim. - 90% **NCBI ID**: XP\_002301540.1
  - ▶ **VvALDH2B9**: Cover - 534/534, Id. - 76%, Sim. - 90% **NCBI ID**: XP\_002274863.1
- ▶ **Family 5** (closest homolog sequence identity is 36%)
- ▶ **Family 10** (closest homolog sequence identity is 34%)
- ▶ **Family 6** (closest homolog sequence identity is 27%)
- ▶ **Family 23** (closest homolog sequence identity is 24%)
- ▶ **Family 11** (closest homolog sequence identity is 24%)
- ▶ **Family 22** (closest homolog sequence identity is 24%)
- ▶ **Family 21** (closest homolog sequence identity is 23%)
- ▶ **Family 7** (closest homolog sequence identity is 20%)
- ▶ **Family 3** (closest homolog sequence identity is 19%)
- ▶ **Family 12** (closest homolog sequence identity is 15%)
- ▶ **Family 24** (closest homolog sequence identity is 13%)
- ▶ **Family 18** (closest homolog sequence identity is 10%)
- ▶ **Family 19** (closest homolog sequence identity is 7%)

**Obr. 26 Výstup programu PAFF pro sekvenci ALDH2B z chřestu (*Asparagus officinalis*).** Program určí příslušnost sekvence a vypíše všechny rodiny ALDH uloženy v rostlinné databázi a pět jejich vstupu nejpodobnějších sekvencí, seřazených sestupně. Uvedeno je pokrytí alignmentu, identita a podobnost sekvencí a také příslušné ID sekvencí z databáze NCBI.

▼ **ZmALDH2B2**: Cover - 534/534, Id. - 79%, Sim. - 90% **NCBI ID**: NP\_001105891.1

```

1: MARRAASSLV--SRCLLARAPAGAPPAAPSAPRRTVPADGMHRLLPGLVLRQFSTAAAVEEPI-TPSVHVNKYTKLLINGNF
  ||  +  |||| |  ++  ++++ |+  ++  |||||+|+  ||+| |  ||| |||||
  MAAARRITSSIIISRCLASPRISILLRGGASSGIQR-----Y-----GTAAAVDEQIPPPVKVEYTKLLINGNF

81: VDSASGKTFPTLDPRTGEVIAHVAEGDAEDINRAVAARKAFDEGPWPKMTAYERSRILLRFADLIEKHNDLAALETWD
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  VDSASGKTFPTLDPRTGEVIAHVAEGDAEDVNHAVAARNAFDEGPWPKMPAYERSRILLRFADLIEKHNDLAALETWD

161: NGKPYEQAQIEVPMVARLMRYAGWADKIHGLIVPADGPHHVQILHEPIGVAGQIIPWIFPLLMYAWKVGVPALACGNTL
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  NGKPYEQAQVEIPMLTRFMRYAGWADKIHGLVVPADGPHHVQVLEHEPIGVAGQIIPWIFPLLMFAWKVGVPALACGNTI

241: VLKTAEQTPLSALYISKLLHEAGLPEGVNVVSGFGPTAGAAASHMDVDKIAFTGSDTGTGKIIELAAKSNLKVTLLEL
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  VLKTAEQTPLSALYVAKLLHEAGIPEGVLNIVSGFGPTAGAAASHMEVDKLAFTGSDTGTGKIVLEMAAKSNLKVTLLEL

321: GGSFPFIIMDDADVDHAVELAHFALFFNQGC[CAGSRTFVHERVYDEFVEKAKARALKRVVGDPPFRKGVEQGPQIDDEQ
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  GGSFPMIVLDDADVDKAVELAHSALFFNQGC[CAGSRTFVHESIYDEFVEKSKARALKRVVGDPPFRKGVEQGPQIDDEQ

401: FNKILRYIRYVDGGATLVTGGDRLGDKGFYIQPTIFSDVQDGMKIAQEEIFGPVQSILKFDLNEVIKRANASQYGLAA
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  FNKILRCIRSGVDSGANLVXRYG--SKGFYIQPTIFSDVQDDMAIAKEEIFGPVQSILKFRDLDEVIKRANTRRYGLAA

481: GVFTNSLDTANTLTRALRAGTVWVNCDFVDAIIPFGGKMSGIGREKGVDSLKNYLQVKAVVTPKNAAWL
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  GVFTSNLDNANTLIRALKVGTVWVNCDFVDAIIPFGGKMSGQGREKIDGLKAYLQVKAVVTPKNAAWL
  
```

**Obr. 27 Příklad alignmentu sekvencí.** Ten je možné zobrazit pro každou databázovou sekvenci na výstupu. Zde se jedná o alignment sekvence ALDH2B2 z kukuřice *Zea mays* a ALDH2B z chřestu (*Asparagus officinalis*). Úplné shody aminokyselin jsou značeny |, přičemž konzervativní

substituce znázorňuje znak +. Jsou zaznačeny také tři katalytické aminokyseliny: asparagin (N) – modře, kyselina glutamová (E) – červeně a cystein (C) - hnědě.

Your sequence most likely doesn't belong to any known Plant ALDH Family but into Animal ALDH Family 16 A

- ▼ **Family 16** (closest homolog sequence identity is 72%)
  - ▶ **HumALDH16A1**: Cover - 733/733, Id. - 72%, Sim. - 85% **NCBI ID**: NP\_699160.2
  - ▶ **DrALDH16A1**: Cover - 733/733, Id. - 21%, Sim. - 53% **NCBI ID**: NP\_001073423.2
- ▶ **Family 18** (closest homolog sequence identity is 10%)
- ▶ **Family 1** (closest homolog sequence identity is 9%)
- ▶ **Family 2** (closest homolog sequence identity is 8%)
- ▶ **Family 3** (closest homolog sequence identity is 8%)
- ▶ **Family 5** (closest homolog sequence identity is 8%)
- ▶ **Family 9** (closest homolog sequence identity is 8%)
- ▶ **Family 4** (closest homolog sequence identity is 8%)
- ▶ **Family 7** (closest homolog sequence identity is 8%)
- ▶ **Family 6** (closest homolog sequence identity is 7%)
- ▶ **Family 8** (closest homolog sequence identity is 7%)

**Obr. 28** Vzhled výstupu po prohledávání zvířecí databáze pomocí sekvenční ALDH16A z Kaloně Egyptského (*Rousettus aegyptiacus*).

Pokud by byla na vstupu zadána sekvenční, jež obsahuje povolené znaky, ale nepatří do žádné z ALDH rodin, program vydá hodnoty podobnosti nejbližší nalezené sekvenční z databáze a zahlásí „Your sequence does not seem to belong to the ALDH superfamily“.

Your sequence does not seem to belong to the ALDH superfamily

- ▼ **Closest homolog sequence was AtALDH2B4 (9%)**
  - ▶ **AtALDH2B4**: Cover - 607/607, Id. - 9%, Sim. - 35% **NCBI ID**: NP\_190383.1

**Obr. 29** Výstup při prohledávání databáze sekvencí BSA. Nenalezena dostatečná shoda.

Program nakonec umožňuje také stáhnout si dosažené výsledky v textové podobě, kde jsou zaznamenány všechny vypsané rodiny a jejich zástupci, společně s dodatečnými informacemi jako jsou pokrytí alignmentu, identita sekvencí či jejich NCBI ID. Mimo to je program také schopný předat uživatelskou sekvenci na web NCBI Blastu, pro další analýzu.

You can download your results here:

For further research, you can Blast your sequence with the NCBI Blast against a bigger database:

Or look for N-terminal presequences via the TargetP server:

Or use Cofactory for prediction of cofactor specificity:

**Obr. 30** Další možnosti programu, které se zobrazí po dokončení analýzy. Tyto možnosti se nacházejí na konci webové stránky a umožňují uživateli stáhnout si svoje výsledky v textové podobě pomocí tlačítka „TXT“. Tlačítko „NCBI“ odkáže uživatele na webové stránky NCBI

Blast a předá serveru vstupní sekvenci. Tlačítka „TargetP“ a „Cofactory“ odkazují na server Center for Biological Sequence Analysis (CBS), kde je možné nalézt N-terminální presekvenční sekvencí nebo identifikovat Rossmannův fold a určit kofaktorovou specifitu sekvence.

### 4.3 Databáze

Databáze programu obsahuje sekvence ze třiceti organismů. Zkoumané genomy rostlin potvrzují předpokládaný trend, že nadrodina ALDH značně expandovala s přechodem vodních rostlin na souš, a to především její rodiny ALDH2 a ALDH3. V tabulce (Tab. 4) pak můžeme pozorovat zastoupení jednotlivých rodin ve všech rostlinných zástupcích v databázi.

**Tab. 4** Počet genů jednotlivých rodin ALDH ve všech rostlinných druzích uložených v databázi.

| Druh                                  | ALDH rodina |   |   |   |   |    |    |    |    |    |    |    |    |    |
|---------------------------------------|-------------|---|---|---|---|----|----|----|----|----|----|----|----|----|
|                                       | 2           | 3 | 5 | 6 | 7 | 10 | 11 | 12 | 18 | 19 | 21 | 22 | 23 | 24 |
| <i>Auxenochlorella protothecoides</i> | 2           | - | 1 | 1 | - | 1  | -  | 1  | 2  | -  | -  | 1  | -  | -  |
| <i>Coccomyxa subellipsoidea</i>       | 2           | - | 1 | 1 | - | 1  | 1  | 1  | 1  | -  | 1  | 1  | -  | -  |
| <i>Chara braunii</i>                  | -           | 1 | 2 | 1 | - | 1  | -  | -  | -  | -  | -  | -  | 1  | -  |
| <i>Chlamydomonas reinhardtii</i>      | 1           | - | 1 | 1 | - | 1  | 1  | 1  | 1  | -  | -  | 1  | -  | 1  |
| <i>Chlorella variabilis</i>           | 2           | - | 1 | 1 | - | -  | 1  | 1  | 1  | -  | -  | 1  | -  | -  |
| <i>Chromochloris zofingiensis</i>     | 2           | - | 1 | 1 | - | 1  | 1  | 1  | 1  | -  | -  | 1  | -  | -  |
| <i>Klebsormidium nitens</i>           | 2           | 1 | 1 | 1 | 1 | 2  | 1  | 1  | 1  | -  | 1  | 1  | 1  | -  |
| <i>Ostreococcus lucimarinus</i>       | -           | 1 | 1 | - | - | 1  | 1  | 1  | 1  | -  | -  | 1  | -  | -  |
| <i>Volvox carteri</i>                 | 1           | - | - | 1 | - | 1  | 1  | 1  | 1  | -  | -  | 1  | -  | -  |
| <i>Physcomitrella patens</i>          | 2           | 5 | 2 | 1 | 1 | 1  | 5  | 1  | 1  | -  | 1  | -  | 1  | -  |
| <i>Sphagnum fallax</i>                | 1           | 7 | 1 | 1 | 1 | 1  | 3  | 1  | 3  | -  | 1  | 1  | 2  | -  |
| <i>Selaginella moellendorffii</i>     | 6           | 2 | 1 | 1 | 1 | 1  | 6  | 1  | 1  | -  | 1  | 1  | 2  | -  |
| <i>Azolla filiculoides</i>            | 5           | 6 | 3 | 2 | 1 | 1  | 3  | 1  | 3  | -  | 1  | 2  | -  | -  |
| <i>Salvinia cucullata</i>             | 3           | 4 | 1 | 1 | 1 | 1  | 3  | 1  | 2  | -  | 1  | 2  | -  | -  |
| <i>Picea sitchensis</i>               | 4           | 2 | - | - | 1 | 2  | 1  | -  | 2  | -  | -  | 1  | -  | -  |
| <i>Oryza sativa</i>                   | 5           | 5 | 1 | 1 | 1 | 2  | 1  | 1  | 2  | -  | -  | 1  | -  | -  |
| <i>Sorghum bicolor</i>                | 5           | 4 | 1 | 1 | 1 | 2  | 1  | 1  | 2  | -  | -  | 1  | -  | -  |

|                             |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|-----------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <i>Zea mays</i>             | 6 | 5 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | - | - | 1 | - | - |
| <i>Zostera marina</i>       | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | - | - | 1 | - | - |
| <i>Amborella trichopoda</i> | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | 1 | - | - |
| <i>Arabidopsis thaliana</i> | 3 | 3 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | - | - | 1 | - | - |
| <i>Malus domestica</i>      | 9 | 6 | 2 | 2 | 2 | 4 | 2 | 2 | 5 | - | - | 2 | - | - |
| <i>Populus trichocarpa</i>  | 4 | 6 | 1 | 4 | 2 | 2 | 3 | 1 | 2 | - | - | 1 | - | - |
| <i>Prunus persica</i>       | 3 | 3 | 1 | 3 | 1 | 3 | 1 | 1 | 2 | - | - | 1 | - | - |
| <i>Solanum lycopersicum</i> | 5 | 6 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | - | 1 | - | - |
| <i>Spinacia oleracea</i>    | 2 | 6 | 1 | 3 | 2 | 1 | 1 | 2 | 1 | - | - | 1 | - | - |
| <i>Vitis vinifera</i>       | 5 | 4 | 3 | 3 | 2 | 2 | 2 | 1 | 2 | - | - | 1 | - | - |

Z části se jedná se o geny rostlin identifikované v Brocker *et al.*, (2013). Jejich zaznačení v databázi se řídí nomenklaturou určenou v tomto článku s dodatečným rozlišením jednotlivých rostlin pomocí předpon. Pro tyto rostliny, *Arabidopsis thaliana* (*At*), *Chlamydomonas reinhardtii* (*Chr*), *Volvox carteri* (*Vc*), *Selaginella moellendorffii* (*Sm*), *Oryza sativa* (*Os*), *Physcomitrella patens* (*Pp*), *Populus trichocarpa* (*Pt*), *Sorghum bicolor* (*Sb*), *Vitis vinifera* (*Vv*) a *Zea mays* (*Zm*), bylo užito tabulek vytvořených právě v Brocker *et al.*, (2013).

Dále byly identifikovány sekvence v rostlinách *Ostreococcus lucimarinus* (*Ol*) (Tab. 4), *Chlorella variabilis* (*Cv*) (Tab. 5), *Auxenochlorella protothecoides* (*Ap*) (Tab. 6), *Chara braunii* (*Cb*) (Tab. 7), *Picea sitchensis* (*Ps*) (Tab. 8), *Sphagnum fallax* (*Sf*) (Tab. 9), *Amborella trichopoda* (*Amt*) (Tab. 10), *Zostera marina* (*Zom*) (Tab. 11) a *Spinacia oleracea* (*So*) (Tab. 12) na základě informací publikovaných ve Stiti *et al.*, (2021). Pro tyto rostliny však byly vytvořeny nové tabulky a jejich sekvence byly v případech, kdy to bylo možné, pojmenovány na základě nomenklatury v Brocker *et al.*, (2013).

**Tab. 5 Geny ALDH přítomné v genomu řasy *Ostreococcus lucimarinus*.** Zaznamenáno je označení sekvencí v databázi PAFF, stejně jako ID odpovídajících záznamů z databází NCBI a Phytozome.

| Rodina ALDH      | Název genu      | NCBI protein ID | Phytozome ID | Značení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|--------------|-----------------------------|--------------------|
| <i>Rodina 3</i>  | <i>ALDH3A1</i>  | XP_001415915.1  | 13873        | <i>OIALDH3A1</i>            | 471                |
| <i>Rodina 5</i>  | <i>ALDH5A1</i>  | XP_001419272.1  | 42872        | <i>OIALDH5A1</i>            | 503                |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | XP_001420154.1  | 50750        | <i>OIALDH10A1</i>           | 523                |
| <i>Rodina 11</i> | <i>ALDH11A1</i> | XP_001418445.1  | 12614        | <i>OIALDH11A1</i>           | 482                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | XP_001417108.1  | 30910        | <i>OIALDH12A1</i>           | 521                |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | XP_001420000.1  | 1184         | <i>OIALDH18B1</i>           | 420                |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | XP_001418857.1  | 16034        | <i>OIALDH22B1</i>           | 510                |

**Tab. 6 Geny ALDH přítomné v genomu řasy *Chlorella variabilis*.** Tabulka obsahuje označení sekvencí v databázi programu PAFF s odpovídající předponou. Je zaznačeno také ID sekvencí z databáze NCBI, skrze kterou byl genom prohledáván.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2E1</i>  | XP_005852174.1  | <i>CvALDH2E1</i>             | 497                |
|                  | <i>ALDH2E2</i>  | XP_005850564.1  | <i>CvALDH2E2</i>             | 597                |
| <i>Rodina 5</i>  | <i>ALDH5G1</i>  | XP_005851079.1  | <i>CvALDH5G1</i>             | 514                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | XP_005848526.1  | <i>CvALDH6B1</i>             | 463                |
| <i>Rodina 11</i> | <i>ALDH11A1</i> | XP_005842757.1  | <i>CvALDH11A1</i>            | 423                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | XP_005844734.1  | <i>CvALDH12A1</i>            | 542                |
| <i>Rodina 18</i> | <i>ALDH18D1</i> | XP_005849463.1  | <i>CvALDH18D1</i>            | 405                |
| <i>Rodina 22</i> | <i>ALDH22B1</i> | XP_005849371.1  | <i>CvALDH22B1</i>            | 377                |



**Tab. 7 Geny ALDH přítomné v genomu řasy *Auxenochlorella protothecoides*.** Tabulka obsahuje záznamy sekvencí z databáze NCBI a jejich odpovídající označení v naší databázi.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B1</i>  | XP_011398903.1  | <i>ApALDH2B1</i>             | 481                |
|                  | <i>ALDH2B2</i>  | XP_011396510.1  | <i>ApALDH2B2</i>             | 732                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | XP_011398214.1  | <i>ApALDH5F1</i>             | 523                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | XP_011398183.1  | <i>ApALDH6B1</i>             | 496                |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | XP_011398091.1  | <i>ApALDH10A1</i>            | 482                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | XP_011398835.1  | <i>ApALDH12A1</i>            | 548                |
| <i>Rodina 18</i> | <i>ALDH18D1</i> | XP_011399237.1  | <i>ApALDH18D1</i>            | 523                |
|                  | <i>ALDH18D2</i> | XP_011399238.1  | <i>ApALDH18D2</i>            | 311                |
| <i>Rodina 22</i> | <i>ALDH22B1</i> | XP_011400850.1  | <i>ApALDH22B1</i>            | 541                |

**Tab. 8 Geny ALDH přítomné v genomu řasy *Chara braunii*.** Je uvedeno jejich označení v databázi NCBI, stejně tak jejich název s odpovídající předponou z databáze programu PAFF.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 3</i>  | <i>ALDH3L1</i>  | GBG89202.1      | <i>CbALDH3L1</i>             | 369                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | GBG72965.1      | <i>CbALDH5F1</i>             | 884                |
|                  | <i>ALDH5F2</i>  | GBG60803.1      | <i>CbALDH5F2</i>             | 332                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | GBG79441.1      | <i>CbALDH6B1</i>             | 1111               |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | GBG76733.1      | <i>CbALDH10A1</i>            | 505                |
| <i>Rodina 23</i> | <i>ALDH23A1</i> | GBG88766.1      | <i>CbALDH23A1</i>            | 552                |

**Tab. 9 Geny ALDH přítomné v genomu smrku *Picea sitchensis*.** Zaznamenáno je označení sekvencí v databázi PAFF, společně s jejich ID z NCBI. Pomlčkou jsou značeny sekvence, které se v databázi programu nenacházejí, buďto z důvodu parciálnosti sekvence, nebo duplikace.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvencí v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B6</i>  | ABK24267.1      | <i>PsALDH2B6</i>             | 544                |
|                  | <i>ALDH2B1</i>  | ABR16485.1      | <i>PsALDH2B1</i>             | 500                |
|                  | <i>ALDH2B2</i>  | ABK24518.1      | -                            | 248                |
|                  | <i>ALDH2B3</i>  | ABK27130.1      | -                            | 200                |
| <i>Rodina 3</i>  | <i>ALDH3H1</i>  | ABR16899.1      | <i>PsALDH3H1</i>             | 556                |
|                  | <i>ALDH3H2</i>  | ABK24239.1      | <i>PsALDH3H2</i>             | 478                |
| <i>Rodina 7</i>  | <i>ALDH7B4</i>  | ABR18443.1      | <i>PsALDH7B4</i>             | 511                |
| <i>Rodina 10</i> | <i>ALDH10A8</i> | ABK24463.1      | <i>PsALDH10A8</i>            | 503                |
|                  | <i>ALDH10A9</i> | ABK24261.1      | -                            | 503                |
| <i>Rodina 11</i> | <i>ALDH11A3</i> | ABK27107.1      | -                            | 293                |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | ABR18097.1      | <i>PsALDH18B1</i>            | 749                |
|                  | <i>ALDH18B2</i> | ABR16169.1      | <i>PsALDH18B2</i>            | 501                |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | ADE75649.1      | <i>PsALDH22A1</i>            | 596                |

**Tab. 10 Geny ALDH přítomné v genomu mechu *Sphagnum fallax*. Zaznamenáno je označení sekvence v databázi PAFF, stejně jako ID odpovídajících záznamů v Phytozome.**

| Rodina ALDH      | Název genu      | Phytozome ID          | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B1</i>  | Sphfalx06G044000.1.p  | <i>SfALDH2B1</i>             | 559                |
| <i>Rodina 3</i>  | <i>ALDH3K1</i>  | Sphfalx17G095500.1.p  | <i>SfALDH3K1</i>             | 480                |
|                  | <i>ALDH3K2</i>  | Sphfalx10G097100.2.p  | <i>SfALDH3K2</i>             | 486                |
|                  | <i>ALDH3K3</i>  | Sphfalx12G067100.1.p  | <i>SfALDH3K3</i>             | 475                |
|                  | <i>ALDH3K4</i>  | Sphfalx15G058300.1.p  | <i>SfALDH3K4</i>             | 486                |
|                  | <i>ALDH3K5</i>  | Sphfalx09G106400.1.p  | <i>SfALDH3K5</i>             | 485                |
|                  | <i>ALDH3K6</i>  | Sphfalx06G091600.2.p  | <i>SfALDH3K6</i>             | 483                |
|                  | <i>ALDH3H1</i>  | Sphfalx13G118100.1.p  | <i>SfALDH3H1</i>             | 494                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | Sphfalx19G075900.1.p  | <i>SfALDH5F1</i>             | 524                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | Sphfalx02G156800.1.p  | <i>SfALDH6B1</i>             | 587                |
| <i>Rodina 7</i>  | <i>ALDH7B4</i>  | Sphfalx11G042000.1.p  | <i>SfALDH7B4</i>             | 511                |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | Sphfalx02G135400.1.p  | <i>SfALDH10A1</i>            | 505                |
| <i>Rodina 11</i> | <i>ALDH11A1</i> | Sphfalx17G041200.1.pS | <i>SfALDH11A1</i>            | 497                |
|                  | <i>ALDH11A2</i> | phfalx12G091900.1.p   | <i>SfALDH11A2</i>            | 496                |
|                  | <i>ALDH11A3</i> | Sphfalx12G053400.1.p  | <i>SfALDH11A3</i>            | 497                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | Sphfalx15G038000.1.p  | <i>SfALDH12A1</i>            | 561                |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | Sphfalx12G082900.1.p  | <i>SfALDH18B1</i>            | 768                |
|                  | <i>ALDH18B2</i> | Sphfalx09G099000.1.p  | <i>SfALDH18B2</i>            | 766                |
|                  | <i>ALDH18B3</i> | Sphfalx06G107600.1.p  | <i>SfALDH18B3</i>            | 767                |
| <i>Rodina 21</i> | <i>ALDH21A1</i> | Sphfalx13G127900.1.p  | <i>SfALDH21A1</i>            | 497                |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | Sphfalx13G102000.1.p  | <i>SfALDH22A1</i>            | 596                |
| <i>Rodina 23</i> | <i>ALDH23A1</i> | Sphfalx02G014100.1.p  | <i>SfALDH23A1</i>            | 492                |
|                  | <i>ALDH23A2</i> | Sphfalx17G014900.1.p  | <i>SfALDH23A2</i>            | 482                |

**Tab. 11 Geny ALDH přítomné v genomu rostliny *Amborella trichopoda*. Zaznamenáno je označení sekvencí v databázi PAFF, stejně jako ID odpovídajících záznamů z databází NCBI.**

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B7</i>  | XP_006844236.2  | <i>AmtALDH2B7</i>            | 535                |
|                  | <i>ALDH2C4</i>  | XP_006842238.2  | <i>AmtALDH2C4</i>            | 501                |
| <i>Rodina 3</i>  | <i>ALDH3E1</i>  | XP_006829153.1  | <i>AmtALDH3E1</i>            | 480                |
|                  | <i>ALDH3H1</i>  | XP_011624378.1  | <i>AmtALDH3H1</i>            | 524                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | XP_020525723.1  | <i>AmtALDH5F1</i>            | 529                |
| <i>Rodina 6</i>  | <i>ALDH6B2</i>  | XP_006852882.1  | <i>AmtALDH6B2</i>            | 533                |
| <i>Rodina 7</i>  | <i>ALDH7B4</i>  | XP_006849726.1  | <i>AmtALDH7B4</i>            | 510                |
| <i>Rodina 10</i> | <i>ALDH10A9</i> | XP_006856888.1  | <i>AmtALDH10A9</i>           | 503                |
| <i>Rodina 11</i> | <i>ALDH11A3</i> | XP_020530844.1  | <i>AmtALDH11A3</i>           | 486                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | XP_006842140.1  | <i>AmtALDH12A1</i>           | 551                |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | XP_011624153.1  | <i>AmtALDH18B1</i>           | 716                |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | XP_006840472.1  | <i>AmtALDH22A1</i>           | 596                |

**Tab. 12 Geny ALDH přítomné v genomu vochy *Zostera marina*.** Zaznamenáno je označení sekvencí v databázi PAFF, stejně jako ID odpovídajících záznamů z databází NCBI a Phytozome.

| Rodina ALDH      | Název genu      | NCBI protein ID | Phytozome ID  | Označení sekvencí v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|---------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B2</i>  | KMZ61155.1      | Zosma03g04330 | <i>ZomALDH2B2</i>            | 534                |
|                  | <i>ALDH2C1</i>  | KMZ59710.1      | Zosma05g05870 | <i>ZomALDH2C1</i>            | 495                |
|                  | <i>ALDH2C2</i>  | KMZ57568.1      | Zosma05g14430 | <i>ZomALDH2C2</i>            | 491                |
| <i>Rodina 3</i>  | <i>ALDH3H1</i>  | KMZ73588.1      | Zosma04g24070 | <i>ZomALDH3H1</i>            | 482                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | KMZ73718.1      | Zosma05g19330 | <i>ZomALDH5F1</i>            | 524                |
| <i>Rodina 6</i>  | <i>ALDH6B2</i>  | KMZ62240.1      | Zosma01g19780 | <i>ZomALDH6B2</i>            | 536                |
| <i>Rodina 7</i>  | <i>ALDH7B6</i>  | KMZ69947.1      | Zosma01g30690 | <i>ZomALDH7B6</i>            | 508                |
| <i>Rodina 10</i> | <i>ALDH10A8</i> | KMZ60546.1      | Zosma01g42820 | <i>ZomALDH10A8</i>           | 506                |
| <i>Rodina 11</i> | <i>ALDH11A1</i> | KMZ74191.1      | Zosma03g28230 | <i>ZomALDH11A1</i>           | 496                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | KMZ56434.1      | Zosma06g26290 | <i>ZomALDH12A1</i>           | 557                |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | KMZ73084.1      | Zosma06g24890 | <i>ZomALDH18B1</i>           | 717                |
|                  | <i>ALDH18B2</i> | KMZ65332.1      | Zosma04g15330 | <i>ZomALDH18B2</i>           | 719                |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | KMZ68070.1      | Zosma01g38810 | <i>ZomALDH22A1</i>           | 595                |

**Tab. 13 Geny ALDH přítomné v genomu špenátu *Spinacia oleracea*.** Tabulka obsahuje záznamy sekvencí z databází Phytozome a NCBI, a také jejich označení v databázi programu PAFF. Pomlčkou jsou značeny parciální sekvence, které se v databázi nenacházejí.

| Rodina<br>ALDH | Název genu      | NCBI protein ID | Phytozome ID      | Označení<br>sekvence v<br>databázi | Počet<br>aminokysel<br>lin |
|----------------|-----------------|-----------------|-------------------|------------------------------------|----------------------------|
| Rodina 2       | <i>ALDH2B4</i>  | XP_021850593.1  | Spov3_chr6.00694  | <i>SoALDH2B4</i>                   | 534                        |
|                | <i>ALDH2C4</i>  | XP_021837880.1  | Spov3_C0044.00009 | <i>SoALDH2C4</i>                   | 500                        |
| Rodina 3       | <i>ALDH3F1</i>  | XP_021849540.1  | Spov3_chr3.00727  | <i>SoALDH3F1</i>                   | 488                        |
|                | <i>ALDH3F2</i>  | XP_021862637.1  | Spov3_chr1.00108  | <i>SoALDH3F2</i>                   | 484                        |
|                | <i>ALDH3H1</i>  | XP_021858246.1  | Spov3_chr5.00178  | <i>SoALDH3H1</i>                   | 494                        |
|                | <i>ALDH3I1</i>  | XP_021863786.1  | Spov3_chr1.00762  | <i>SoALDH3I1</i>                   | 487                        |
|                | <i>ALDH3I2</i>  | XP_021863811.1  | Spov3_chr1.00760  | <i>SoALDH3I2</i>                   | 522                        |
|                | <i>ALDH3I3</i>  | XP_021861606.1  | Spov3_chr4.01290  | -                                  | 206                        |
| Rodina 5       | <i>ALDH5F1</i>  | XP_021864456.1  | Spov3_chr4.02986  | <i>SoALDH5F1</i>                   | 534                        |
| Rodina 6       | <i>ALDH6B1</i>  | XP_021862462.1  | Spov3_chr1.04616  | <i>SoALDH6B1</i>                   | 533                        |
|                | <i>ALDH6B2</i>  | XP_021838774.1  | Spov3_chr3.01514  | <i>SoALDH6B2</i>                   | 805                        |
|                | <i>ALDH6B3</i>  | XP_021862464.1  | Spov3_chr1.04616  | <i>SoALDH6B3</i>                   | 728                        |
| Rodina 7       | <i>ALDH7B4</i>  | XP_021865286.1  | Spov3_chr4.04545  | <i>SoALDH7B4</i>                   | 508                        |
|                | <i>ALDH7B1</i>  | XP_021858694.1  | -                 | -                                  | 198                        |
| Rodina 10      | <i>ALDH10A8</i> | XP_021837172.1  | Spov3_chr6.00773  | <i>SoALDH10A8</i>                  | 497                        |
| Rodina 11      | <i>ALDH11A3</i> | XP_021864704.1  | Spov3_chr4.03080  | <i>SoALDH11A3</i>                  | 496                        |
| Rodina 12      | <i>ALDH12A1</i> | XP_021850981.1  | Spov3_chr6.00884  | <i>SoALDH12A1</i>                  | 554                        |
|                | <i>ALDH12A2</i> | XP_021840333.1  | -                 | -                                  | 257                        |
| Rodina 18      | <i>ALDH18B1</i> | XP_021845616.1  | Spov3_chr4.03018  | <i>SoALDH18B1</i>                  | 719                        |
| Rodina 22      | <i>ALDH22A1</i> | XP_021852296.1  | Spov3_chr3.04496  | <i>SoALDH22A1</i>                  | 594                        |

Databáze dále obsahuje sekvence z kapradin *Azolla filiculoides* (*Af*) (Tab. 13) a *Salvinia cucullata* (*Scu*) (Tab. 14), které jsou jedinými sekvenovanými zástupci tohoto oddělení. Tabulky sekvencí ALDH z těchto rostlin byly taktéž publikovány ve Stiti *et al.*, (2021), avšak při prohledávání nukleotidové databáze těchto rostlin na serveru FernBase, byly nalezeny dodatečné sekvence a byly tedy vytvořeny nové tabulky. Při pojmenování jednotlivých genů jsem se opět řídil nomenklaturou z Brocker *et al.*, (2013).

**Tab. 14 Geny ALDH identifikovány v kapradině *Azolla filiculoides*.** Genom byl prohledáván skrze databázi projektu FernBase, jež obsahuje všechny sekvenované zástupce oddělení kapradin. Kromě FernBase ID tabulka obsahuje také označení sekvencí v naší databázi. Parciální sekvence do databáze ukládány nebyly a jsou značeny pomlčkou.

| Rodina ALDH      | Název genu      | FernBase protein ID | Označení sekvence<br>v databázi | Počet<br>aminokyselin |
|------------------|-----------------|---------------------|---------------------------------|-----------------------|
| <i>Rodina 2</i>  | <i>ALDH2B1</i>  | Azfi_s0083.g038897  | <i>AfALDH2B1</i>                | 497                   |
|                  | <i>ALDH2C1</i>  | Azfi_s0078.g038198  | <i>AfALDH2C1</i>                | 465                   |
|                  | <i>ALDH2C2</i>  | Azfi_s0049.g030778  | <i>AfALDH2C2</i>                | 468                   |
|                  | <i>ALDH2C3</i>  | Azfi_s0021.g015740  | <i>AfALDH2C3</i>                | 510                   |
|                  | <i>ALDH2F1</i>  | Azfi_s0002.g004343  | <i>AfALDH2F1</i>                | 480                   |
| <i>Rodina 3</i>  | <i>ALDH3H1</i>  | Azfi_s0587.g078563  | <i>AfALDH3H1</i>                | 516                   |
|                  | <i>ALDH3H2</i>  | Azfi_s0036.g025798  | <i>AfALDH3H2</i>                | 425                   |
|                  | <i>ALDH3H3</i>  | Azfi_s0109.g045319  | <i>AfALDH3H3</i>                | 457                   |
|                  | <i>ALDH3H4</i>  | Azfi_s0011.g012726  | <i>AfALDH3H4</i>                | 518                   |
|                  | <i>ALDH3H5</i>  | Azfi_s0636.g080483  | <i>AfALDH3H5</i>                | 573                   |
|                  | <i>ALDH3H6</i>  | Azfi_s0002.g001677  | <i>AfALDH3H6</i>                | 635                   |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | Azfi_s0033.g025123  | <i>AfALDH5F1</i>                | 672                   |
|                  | <i>ALDH5H1</i>  | Azfi_s0002.g003146  | <i>AfALDH5H1</i>                | 733                   |
|                  | <i>ALDH5H2</i>  | Azfi_s0002.g003388  | <i>AfALDH5H2</i>                | 636                   |
| <i>Rodina 6</i>  | <i>ALDH6B2</i>  | Azfi_s0012.g012829  | <i>AfALDH6B2</i>                | 567                   |
|                  | <i>ALDH6C1</i>  | Azfi_s0002.g003326  | <i>AfALDH6C1</i>                | 576                   |
| <i>Rodina 7</i>  | <i>ALDH7B4</i>  | Azfi_s0068.g036288  | <i>AfALDH7B4</i>                | 497                   |
| <i>Rodina 10</i> | <i>ALDH10A9</i> | Azfi_s0092.g043158  | <i>AfALDH10A9</i>               | 464                   |
| <i>Rodina 11</i> | <i>ALDH11A1</i> | Azfi_s0160.g054025  | <i>AfALDH11A1</i>               | 546                   |
|                  | <i>ALDH11A2</i> | Azfi_s0003.g007703  | <i>AfALDH11A2</i>               | 458                   |
|                  | <i>ALDH11A3</i> | Azfi_s0006.g010360  | <i>AfALDH11A3</i>               | 944                   |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | Azfi_s0027.g023626  | <i>AfALDH12A1</i>               | 571                   |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | Azfi_s0157.g053821  | <i>AfALDH18B1</i>               | 754                   |
|                  | <i>ALDH18B2</i> | Azfi_s0022.g015914  | <i>AfALDH18B2</i>               | 742                   |
|                  | <i>ALDH18B3</i> | Azfi_s1111.g097729  | <i>AfALDH18B3</i>               | 738                   |
| <i>Rodina 21</i> | <i>ALDH21A1</i> | Azfi_s0412.g068547  | <i>AfALDH21A1</i>               | 1140                  |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | Azfi_s0022.g016122  | <i>AfALDH22A1</i>               | 366                   |
|                  | <i>ALDH22A2</i> | Azfi_s0022.g016119  | -                               | 187                   |

**Tab. 15 Geny ALDH identifikovány v kapradině *Salvinia cucullata*.** Genom byl prohledáván skrze databázi projektu FernBase. Tabulka obsahuje označení sekvencí z databáze programu, společně s ID ze serveru Fernbase. Pomlčkou jsou značeny parciální sekvence, které do databáze zařazeny nebyly.

| Rodina<br>ALDH | Název genu      | Fernbase ID             | Označení v<br>databázi | Počet<br>aminokyselin |
|----------------|-----------------|-------------------------|------------------------|-----------------------|
| Rodina 2       | <i>ALDH2B1</i>  | Sacu_v1.1_s0061.g015297 | <i>ScuALDH2B1</i>      | 533                   |
|                | <i>ALDH2B2</i>  | Sacu_v1.1_s0069.g016540 | <i>ScuALDH2B2</i>      | 503                   |
|                | <i>ALDH2C1</i>  | Sacu_v1.1_s0045.g013293 | -                      | 248                   |
| Rodina 3       | <i>ALDH3H1</i>  | Sacu_v1.1_s0014.g006331 | <i>ScuALDH3H1</i>      | 516                   |
|                | <i>ALDH3H2</i>  | Sacu_v1.1_s0095.g019266 | <i>ScuALDH3H2</i>      | 539                   |
|                | <i>ALDH3H3</i>  | Sacu_v1.1_s1488.g028042 | <i>ScuALDH3H3</i>      | 488                   |
|                | <i>ALDH3K1</i>  | Sacu_v1.1_s0093.g019019 | <i>ScuALDH3K1</i>      | 539                   |
| Rodina 5       | <i>ALDH5F1</i>  | Sacu_v1.1_s0131.g022111 | <i>ScuALDH5F1</i>      | 582                   |
| Rodina 6       | <i>ALDH6B1</i>  | Sacu_v1.1_s0129.g021951 | <i>ScuALDH6B1</i>      | 551                   |
| Rodina 7       | <i>ALDH7B4</i>  | Sacu_v1.1_s0036.g011734 | <i>ScuALDH7B4</i>      | 461                   |
| Rodina 10      | <i>ALDH10A1</i> | Sacu_v1.1_s0099.g019543 | <i>ScuALDH10A1</i>     | 545                   |
| Rodina 11      | <i>ALDH11A1</i> | Sacu_v1.1_s0070.g016707 | <i>ScuALDH11A1</i>     | 518                   |
|                | <i>ALDH11A2</i> | Sacu_v1.1_s0112.g020728 | <i>ScuALDH11A2</i>     | 498                   |
|                | <i>ALDH11A3</i> | Sacu_v1.1_s0196.g025486 | <i>ScuALDH11A3</i>     | 546                   |
| Rodina 12      | <i>ALDH12A1</i> | Sacu_v1.1_s0068.g016417 | <i>ScuALDH12A1</i>     | 550                   |
| Rodina 18      | <i>ALDH18B1</i> | Sacu_v1.1_s0176.g024621 | <i>ScuALDH18B1</i>     | 633                   |
|                | <i>ALDH18B2</i> | Sacu_v1.1_s0225.g026315 | <i>ScuALDH18B2</i>     | 619                   |
| Rodina 21      | <i>ALDH21A1</i> | Sacu_v1.1_s0112.g020746 | <i>ScuALDH21A1</i>     | 1041                  |
| Rodina 22      | <i>ALDH22A1</i> | Sacu_v1.1_s0071.g016883 | -                      | 252                   |
|                | <i>ALDH22A2</i> | Sacu_v1.1_s0071.g016884 | <i>ScuALDH22A2</i>     | 391                   |

Zbylé databázové sekvence z rostlin odpovídají řasám *Coccomyxa subellipsoidea* (*Cs*) (Tab. 15), *Chromochloris zofingiensis* (*Chz*) (Tab. 16) a *Klebsormidium nitens* (*Kn*) (Tab. 17), dále jsou zde sekvence z jabloně *Malus domestica* (*Md*) (Tab. 18) a broskvoně *Prunus persica* (*Prp*) (Tab. 19), a také rajčete *Solanum lycopersicum* (*Sl*) (Tab. 20), jakožto jediného zástupce rostlinné říše, který obsahuje sekvence řadící se do rodiny ALDH19. Tabulky sekvencí z těchto rostlin nebyly doposud publikovány nebo neobsahují aktuální informace a jejich zařazení do rodin bylo realizováno opět pomocí nomenklatury stanovené v Brocker *et al.*, (2013).



**Tab. 16 ALDH geny nalezeny v genomu řasy *Coccomyxa subellipsoidea*.** Tabulka obsahuje záznamy těchto sekvencí v databázích serverů NCBI a Phytozome, stejně tak označení sekvencí v databázi programu PAFF.

| Rodina ALDH      | Název genu      | NCBI protein ID | Phytozome ID | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|--------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B2</i>  | XP_005642984.1  | 54894        | <i>CsALDH2B2</i>             | 565                |
|                  | <i>ALDH2B4</i>  | XP_005642981.1  | 31641        | <i>CsALDH2B4</i>             | 486                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | XP_005651730.1  | 34715        | <i>CsALDH5F1</i>             | 537                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | XP_005647038.1  | 29754        | <i>CsALDH6B1</i>             | 544                |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | XP_005645499.1  | 37718        | <i>CsALDH10A1</i>            | 498                |
| <i>Rodina 11</i> | <i>ALDH11A3</i> | XP_005650540.1  | 52649        | <i>CsALDH11A3</i>            | 496                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | XP_005643005.1  | 54900        | <i>CsALDH12A1</i>            | 542                |
| <i>Rodina 18</i> | <i>ALDH18D1</i> | XP_005645012.1  | 37765        | <i>CsALDH18D1</i>            | 778                |
| <i>Rodina 21</i> | <i>ALDH21A1</i> | XP_005644523.1  | 19137        | <i>CsALDH21A1</i>            | 500                |
| <i>Rodina 22</i> | <i>ALDH22B1</i> | XP_005650418.1  | 12562        | <i>CsALDH22B1</i>            | 517                |

**Tab. 17 ALDH geny nalezeny v genomu řasy *Chromochloris zofingiensis*.** Tabulka obsahuje označení všech sekvencí tohoto organismu uložených v databázi programu PAFF a jejich odpovídající ID v databázi serveru Phytozome.

| Rodina ALDH      | Název genu      | Phytozome ID  | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|---------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2E1</i>  | Cz03g20090.t1 | <i>ChzALDH2E1</i>            | 479                |
|                  | <i>ALDH2F1</i>  | Cz11g04290.t1 | <i>ChzALDH2F1</i>            | 490                |
| <i>Rodina 5</i>  | <i>ALDH5G1</i>  | Cz05g02090.t1 | <i>ChzALDH5G1</i>            | 559                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | Cz12g20220.t1 | <i>ChzALDH6B1</i>            | 524                |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | Cz03g10210.t1 | <i>ChzALDH10A1</i>           | 498                |
| <i>Rodina 11</i> | <i>ALDH11A1</i> | Cz06g05120.t1 | <i>ChzALDH11A1</i>           | 497                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | Cz05g18170.t1 | <i>ChzALDH12A1</i>           | 501                |
| <i>Rodina 18</i> | <i>ALDH18D1</i> | Cz04g39260.t1 | <i>ChzALDH18D1</i>           | 438                |
| <i>Rodina 22</i> | <i>ALDH22B1</i> | Cz09g10170.t1 | <i>ChzALDH22B1</i>           | 613                |

**Tab. 18 ALDH geny nalezeny v genomu řasy *Klebsormidium nitens*.** V tabulce jsou uvedeny ID těchto genů v databázích NCBI a projektu Plantmorphogenesis, společně s označením odpovídajících sekvencí v databázi našeho programu. Pomlčka značí parciální sekvence nezařazené do databáze.

| Rodina<br>ALDH   | Název genu      | NCBI<br>protein<br>ID | Plantmorphogenesis<br>ID | Označení<br>sekvence v<br>databázi | Počet<br>aminokyseliny |
|------------------|-----------------|-----------------------|--------------------------|------------------------------------|------------------------|
| <i>Rodina 2</i>  | <i>ALDH2B1</i>  | GAQ84628.1            | kfl00198_0100_v1.1       | <i>KnALDH2B1</i>                   | 495                    |
|                  | <i>ALDH2F1</i>  | GAQ83917.1            | kfl00168_0120_v1.1       | -                                  | 573                    |
| <i>Rodina 3</i>  | <i>ALDH3M1</i>  | GAQ82469.1            | kfl00112_0130_v1.1       | <i>KnALDH3M1</i>                   | 478                    |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | GAQ91062.1            | kfl00722_0010_v1.1       | <i>KnALDH5F1</i>                   | 580                    |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | GAQ92677.1            | kfl01091_0010_v1.1       | <i>KnALDH6B1</i>                   | 564                    |
| <i>Rodina 7</i>  | <i>ALDH7B4</i>  | GAQ80312.1            | kfl00051_0130_v1.1       | <i>KnALDH7B4</i>                   | 521                    |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | GAQ87015.1            | kfl00325_0190_v1.1       | <i>KnALDH10A1</i>                  | 544                    |
|                  | <i>ALDH10A2</i> | GAQ90986.1            | kfl00710_0030_v1.1       | -                                  | 123                    |
| <i>Rodina 11</i> | <i>ALDH11A3</i> | GAQ90687.1            | kfl00674_0010_v1.1       | <i>KnALDH11A3</i>                  | 497                    |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | GAQ87509.1            | kfl00358_0030_v1.1       | <i>KnALDH12A1</i>                  | 535                    |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | GAQ89957.1            | kfl00582_0010_v1.1       | <i>KnALDH18B1</i>                  | 746                    |
| <i>Rodina 21</i> | <i>ALDH21A1</i> | GAQ88729.1            | kfl00454_0140_v1.1       | <i>KnALDH21A1</i>                  | 575                    |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | GAQ84711.1            | kfl00202_0060_v1.1       | <i>KnALDH22A1</i>                  | 621                    |
| <i>Rodina 23</i> | <i>ALDH23A1</i> | GAQ77842.1            | kfl00004_0420_v1.1       | <i>KnALDH23A1</i>                  | 496                    |

**Tab. 19 Geny ALDH identifikovány v genomu jabloně *Malus domestica*.** Tabulka obsahuje záznamy sekvencí z databáze Phytozome, a také jejich označení v databázi programu PAFF. Pomlčkou jsou značeny duplikátní sekvence, které by databázi ničím nepřispěly a nebyly do ní tedy zařazeny.

| Rodina ALDH | Název genu      | Phytozome ID | Označení sekvence v databázi | Počet aminokyselin |
|-------------|-----------------|--------------|------------------------------|--------------------|
| Rodina 2    | <i>ALDH2B1</i>  | MD14G1183400 | <i>MdALDH2B4</i>             | 542                |
|             | <i>ALDH2B2</i>  | MD06G1177200 | -                            | 539                |
|             | <i>ALDH2B3</i>  | MD13G1090800 | <i>MdALDH2B7</i>             | 534                |
|             | <i>ALDH2B4</i>  | MD13G1090600 | -                            | 534                |
|             | <i>ALDH2C1</i>  | MD02G1049800 | <i>MdALDH2C1</i>             | 501                |
|             | <i>ALDH2C2</i>  | MD02G1049700 | <i>MdALDH2C2</i>             | 503                |
|             | <i>ALDH2C3</i>  | MD02G1050000 | <i>MdALDH2C3</i>             | 501                |
|             | <i>ALDH2C4</i>  | MD15G1187300 | <i>MdALDH2C4</i>             | 554                |
|             | <i>ALDH2C5</i>  | MD02G1049900 | <i>MdALDH2C5</i>             | 482                |
| Rodina 3    | <i>ALDH3F1</i>  | MD15G1250400 | <i>MdALDH3F1</i>             | 483                |
|             | <i>ALDH3F2</i>  | MD02G1136600 | <i>MdALDH3F2</i>             | 487                |
|             | <i>ALDH3H1</i>  | MD15G1345700 | <i>MdALDH3H1</i>             | 502                |
|             | <i>ALDH3H2</i>  | MD00G1029100 | <i>MdALDH3H2</i>             | 486                |
|             | <i>ALDH3I1</i>  | MD15G1137200 | <i>MdALDH3I1</i>             | 560                |
|             | <i>ALDH3I2</i>  | MD15G1136900 | -                            | 560                |
| Rodina 5    | <i>ALDH5F1</i>  | MD10G1216600 | <i>MdALDH5F1</i>             | 541                |
|             | <i>ALDH5F2</i>  | MD05G1238300 | <i>MdALDH5F2</i>             | 527                |
| Rodina 6    | <i>ALDH6B1</i>  | MD09G1234200 | <i>MdALDH6B2</i>             | 539                |
|             | <i>ALDH6B2</i>  | MD15G1309700 | -                            | 1077               |
| Rodina 7    | <i>ALDH7B1</i>  | MD02G1094300 | <i>MdALDH7B4</i>             | 508                |
|             | <i>ALDH7B2</i>  | MD15G1219200 | -                            | 509                |
| Rodina 10   | <i>ALDH10A1</i> | MD14G1187400 | <i>MdALDH10A1</i>            | 503                |
|             | <i>ALDH10A2</i> | MD06G1181500 | <i>MdALDH10A2</i>            | 503                |
|             | <i>ALDH10A3</i> | MD06G1181400 | <i>MdALDH10A3</i>            | 537                |
|             | <i>ALDH10A4</i> | MD14G1187300 | <i>MdALDH10A4</i>            | 504                |
| Rodina 11   | <i>ALDH11A1</i> | MD02G1009300 | <i>MdALDH11A3</i>            | 496                |
|             | <i>ALDH11A2</i> | MD15G1154800 | -                            | 496                |
| Rodina 12   | <i>ALDH12A1</i> | MD06G1189000 | <i>MdALDH12A1</i>            | 545                |
|             | <i>ALDH12A2</i> | MD14G1196000 | -                            | 468                |
| Rodina 18   | <i>ALDH18B1</i> | MD01G1140200 | <i>MdALDH18B1</i>            | 717                |
|             | <i>ALDH18B2</i> | MD12G1150700 | <i>MdALDH18B2</i>            | 700                |
|             | <i>ALDH18B3</i> | MD00G1141100 | <i>MdALDH18B3</i>            | 680                |
|             | <i>ALDH18B4</i> | MD04G1136200 | <i>MdALDH18B4</i>            | 634                |
|             | <i>ALDH18B5</i> | MD07G1204500 | -                            | 717                |
| Rodina 22   | <i>ALDH22A1</i> | MD13G1155400 | <i>MdALDH22A1</i>            | 596                |
|             | <i>ALDH22A2</i> | MD16G1154800 | -                            | 596                |

**Tab. 20 Geny ALDH identifikovány v genomu broskvoně *Prunus persica*.** Tabulka obsahuje záznamy sekvencí z databází serverů NCBI a Phytozome, a také jejich označení v databázi programu PAFF. ALDH6B3 nebyla do databáze zařazena z důvodu špatného C-terminálního splicingu.

| Rodina ALDH      | Název genu      | NCBI protein ID | Phytozome ID     | Označení sekvence v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------|------------------------------|--------------------|
| <i>Rodina 2</i>  | <i>ALDH2B6</i>  | XP_007210569.1  | Prupe.5G175900.1 | <i>PrpALDH2B6</i>            | 543                |
|                  | <i>ALDH2B2</i>  | XP_007221800.1  | Prupe.1G255200.1 | <i>PrpALDH2B2</i>            | 534                |
|                  | <i>ALDH2C4</i>  | XP_007201949.1  | Prupe.7G229100.1 | <i>PrpALDH2C4</i>            | 503                |
| <i>Rodina 3</i>  | <i>ALDH3F1</i>  | XP_020423668.1  | Prupe.7G163900.1 | <i>PrpALDH3F1</i>            | 501                |
|                  | <i>ALDH3H1</i>  | XP_007222253.1  | Prupe.1G493800.1 | <i>PrpALDH3H1</i>            | 493                |
|                  | <i>ALDH3I1</i>  | XP_007213817.2  | Prupe.4G279800.1 | <i>PrpALDH3I1</i>            | 549                |
| <i>Rodina 5</i>  | <i>ALDH5F1</i>  | XP_007213742.2  | Prupe.4G128100.1 | <i>PrpALDH5F1</i>            | 526                |
| <i>Rodina 6</i>  | <i>ALDH6B1</i>  | XP_007199787.1  | Prupe.8G073100.1 | <i>PrpALDH6B1</i>            | 538                |
|                  | <i>ALDH6B2</i>  | XP_007201739.1  | Prupe.8G073300.1 | <i>PrpALDH6B2</i>            | 698                |
|                  | <i>ALDH6B3</i>  | -               | Prupe.6G205000.1 | -                            | 1084               |
| <i>Rodina 7</i>  | <i>ALDH7B4</i>  | XP_007201943.1  | Prupe.7G196700.1 | <i>PrpALDH7B4</i>            | 508                |
| <i>Rodina 10</i> | <i>ALDH10A1</i> | XP_007209928.1  | Prupe.5G179800.1 | <i>PrpALDH10A1</i>           | 503                |
|                  | <i>ALDH10A2</i> | XP_020418929.1  | Prupe.5G179700.1 | <i>PrpALDH10A2</i>           | 503                |
|                  | <i>ALDH10A3</i> | XP_020415265.1  | Prupe.3G277000.1 | <i>PrpALDH10A3</i>           | 535                |
| <i>Rodina 11</i> | <i>ALDH11A3</i> | XP_007203871.1  | Prupe.7G260700.2 | <i>PrpALDH11A3</i>           | 497                |
| <i>Rodina 12</i> | <i>ALDH12A1</i> | XP_007209859.1  | Prupe.5G187700.1 | <i>PrpALDH12A1</i>           | 554                |
| <i>Rodina 18</i> | <i>ALDH18B1</i> | ALT55651.1      | Prupe.2G240800.1 | <i>PrpALDH18B1</i>           | 717                |
|                  | <i>ALDH18B2</i> | XP_020422510.1  | Prupe.6G262300.1 | <i>PrpALDH18B2</i>           | 722                |
| <i>Rodina 22</i> | <i>ALDH22A1</i> | XP_007221826.1  | Prupe.1G118500.1 | <i>PrpALDH22A1</i>           | 596                |

**Tab. 21 Geny ALDH identifikovány v genomu rajčete *Solanum lycopersicum*.** V tabulce jsou uvedeny záznamy sekvencí z databází serverů Phytozome a NCBI, a také jejich označení v databázi programu PAFF. Pomlčkou jsou značeny parciální sekvence, které do databáze zařazeny nebyly.

| Rodina ALDH | Název genu      | NCBI protein ID | Phytozome ID       | Označení sekvencí v databázi | Počet aminokyselin |
|-------------|-----------------|-----------------|--------------------|------------------------------|--------------------|
| Rodina 2    | <i>ALDH2B1</i>  | NP_001333536.1  | Solyc03g114150.3.1 | <i>SIALDH2B1</i>             | 535                |
|             | <i>ALDH2B2</i>  | NP_001333533.1  | Solyc05g005700.3.1 | <i>SIALDH2B2</i>             | 529                |
|             | <i>ALDH2B3</i>  | XP_004245256.1  | Solyc08g068190.3.1 | <i>SIALDH2B3</i>             | 534                |
|             | <i>ALDH2B4</i>  | XP_010316283.1  | Solyc02g086970.3.1 | <i>SIALDH2B4</i>             | 491                |
|             | <i>ALDH2C1</i>  | XP_004251601.1  | Solyc12g007030.2.1 | <i>SIALDH2C1</i>             | 500                |
| Rodina 3    | <i>ALDH3F1</i>  | XP_010316396.1  | Solyc02g084640.3.1 | <i>SIALDH3F1</i>             | 469                |
|             | <i>ALDH3F2</i>  | NP_001333537.1  | Solyc01g011510.3.1 | <i>SIALDH3F2</i>             | 500                |
|             | <i>ALDH3F3</i>  | -               | Solyc01g013890.2.1 | -                            | 206                |
|             | <i>ALDH3F4</i>  | -               | Solyc01g012700.3.1 | -                            | 310                |
|             | <i>ALDH3F5</i>  | -               | Solyc01g012760.1.1 | -                            | 207                |
|             | <i>ALDH3H1</i>  | NP_001333570.1  | Solyc06g060250.3.1 | <i>SIALDH3H1</i>             | 474                |
| Rodina 5    | <i>ALDH5F1</i>  | NP_001233841.1  | Solyc09g090700.1.1 | <i>SIALDH5F1</i>             | 522                |
| Rodina 6    | <i>ALDH6B1</i>  | NP_001333575.1  | Solyc01g106080.3.1 | <i>SIALDH6B1</i>             | 576                |
| Rodina 7    | <i>ALDH7B4</i>  | XP_004235550.1  | Solyc03g122310.3.1 | <i>SIALDH7B4</i>             | 524                |
| Rodina 10   | <i>ALDH10A1</i> | NP_001333606.1  | Solyc06g071290.3.1 | <i>SIALDH10A1</i>            | 504                |
|             | <i>ALDH10A2</i> | NP_001234235.2  | Solyc03g113800.3.1 | <i>SIALDH10A2</i>            | 505                |
| Rodina 11   | <i>ALDH11A3</i> | NP_001333610.1  | Solyc07g005390.3.1 | <i>SIALDH11A3</i>            | 496                |
| Rodina 12   | <i>ALDH12A1</i> | NP_001333614.1  | Solyc06g071000.3.1 | <i>SIALDH12A1</i>            | 555                |
| Rodina 18   | <i>ALDH18B1</i> | NP_001233907.1  | Solyc08g043170.3.1 | <i>SIALDH18B1</i>            | 720                |
|             | <i>ALDH18B2</i> | XP_004240687.1  | Solyc06g019170.3.1 | <i>SIALDH18B2</i>            | 717                |
| Rodina 19   | <i>ALDH19A1</i> | AAB67877.1      | -                  | <i>SIALDH19</i>              | 414                |
| Rodina 22   | <i>ALDH22A1</i> | XP_004252923.1  | Solyc12g099290.2.1 | <i>SIALDH22A1</i>            | 593                |

Kromě rostlin byly pro databázi prozkoumány také živočišné genomy a vybrané sekvence ALDH z člověka (*Homo sapiens*) (Tab. 21), slepice (*Gallus gallus*) (Tab. 22) a ryby (*Danio rerio*) (Tab. 23). Jejich pojmenování se řídí nomenklaturou z Jackson *et al.*, (2011).

**Tab. 22 ALDH geny přítomné v genomu člověka.** Všechny sekvence jsou v databázi zaznačené odpovídajícím názvem s předponou „Hum“. Délka je také pro každou uvedeno NCBI ID těchto proteinů.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvencí v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 1</i>  | <i>ALDH1A1</i>  | NP_000680.2     | <i>HumALDH1A1</i>            | 501                |
|                  | <i>ALDH1A2</i>  | NP_003879.2     | <i>HumALDH1A2</i>            | 518                |
|                  | <i>ALDH1A3</i>  | NP_000684.2     | <i>HumALDH1A3</i>            | 512                |
|                  | <i>ALDH1B1</i>  | NP_000683.3     | <i>HumALDH1B1</i>            | 517                |
|                  | <i>ALDH1L1</i>  | NP_036322.2     | <i>HumALDH1L1</i>            | 902                |
|                  | <i>ALDH1L2</i>  | NP_001029345.2  | <i>HumALDH1L2</i>            | 923                |
| <i>Rodina 2</i>  | <i>ALDH2A1</i>  | NP_000681.2     | <i>HumALDH2A1</i>            | 517                |
| <i>Rodina 3</i>  | <i>ALDH3A1</i>  | NP_000682.3     | <i>HumALDH3A1</i>            | 453                |
|                  | <i>ALDH3A2</i>  | NP_000373.1     | <i>HumALDH3A2</i>            | 485                |
|                  | <i>ALDH3B1</i>  | NP_000685.1     | <i>HumALDH3B1</i>            | 468                |
|                  | <i>ALDH3B2</i>  | NP_001026786.2  | <i>HumALDH3B2</i>            | 385                |
| <i>Rodina 4</i>  | <i>ALDH4A1</i>  | NP_003739.2     | <i>HumALDH4A1</i>            | 563                |
| <i>Rodina 5</i>  | <i>ALDH5A1</i>  | NP_001071.1     | <i>HumALDH5A1</i>            | 535                |
| <i>Rodina 6</i>  | <i>ALDH6A1</i>  | NP_005580.1     | <i>HumALDH6A1</i>            | 535                |
| <i>Rodina 7</i>  | <i>ALDH7A1</i>  | NP_001173.2     | <i>HumALDH7A1</i>            | 539                |
| <i>Rodina 8</i>  | <i>ALDH8A1</i>  | NP_072090.1     | <i>HumALDH8A1</i>            | 487                |
| <i>Rodina 9</i>  | <i>ALDH9A1</i>  | NP_000687.3     | <i>HumALDH9A1</i>            | 518                |
| <i>Rodina 16</i> | <i>ALDH16A1</i> | NP_699160.2     | <i>HumALDH16A1</i>           | 802                |
| <i>Rodina 18</i> | <i>ALDH18A1</i> | NP_002851.2     | <i>HumALDH18A1</i>           | 795                |

**Tab. 23 ALDH geny přítomné v ptačím genomu *Gallus gallus*.** Sekvence v databázi programu jsou značeny s předponou „Gg“ a ke každé je uvedeno také odpovídající ID ze serveru NCBI.

| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvencí v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 1</i>  | <i>ALDH1A1</i>  | NP_989908.1     | <i>GgALDH1A1</i>             | 509                |
|                  | <i>ALDH1A2</i>  | NP_990326.2     | <i>GgALDH1A2</i>             | 518                |
|                  | <i>ALDH1A3</i>  | NP_990000.1     | <i>GgALDH1A3</i>             | 512                |
|                  | <i>ALDH1L2</i>  | XP_416314.2     | <i>GgALDH1L2</i>             | 922                |
| <i>Rodina 2</i>  | <i>ALDH2A1</i>  | NP_001376401.1  | <i>GgALDH2A1</i>             | 519                |
| <i>Rodina 3</i>  | <i>ALDH3A2</i>  | NP_001006223.1  | <i>GgALDH3A2</i>             | 490                |
|                  | <i>ALDH3B1</i>  | XP_426371.2     | <i>GgALDH3B1</i>             | 496                |
| <i>Rodina 4</i>  | <i>ALDH4A1</i>  | NP_001239033.1  | <i>GgALDH4A1</i>             | 551                |
| <i>Rodina 5</i>  | <i>ALDH5A1</i>  | XP_418909.3     | <i>GgALDH5A1</i>             | 548                |
| <i>Rodina 6</i>  | <i>ALDH6A1</i>  | XP_004941836.1  | <i>GgALDH6A1</i>             | 538                |
| <i>Rodina 7</i>  | <i>ALDH7A1</i>  | XP_424422.2     | <i>GgALDH7A1</i>             | 536                |
| <i>Rodina 8</i>  | <i>ALDH8A1</i>  | XP_419732.1     | <i>GgALDH8A1</i>             | 487                |
| <i>Rodina 9</i>  | <i>ALDH9A1</i>  | XP_422248.4     | <i>GgALDH9A1</i>             | 579                |
| <i>Rodina 18</i> | <i>ALDH18A1</i> | XP_025007757.1  | <i>GgALDH18A1</i>            | 797                |

**Tab. 24 ALDH geny přítomné v genomu ryby *Danio rerio*.** Všechny sekvence tohoto organismu jsou značeny v databázi předponou „Dr“. V tabulce je pak uvedeno také ID ze serveru NCBI.

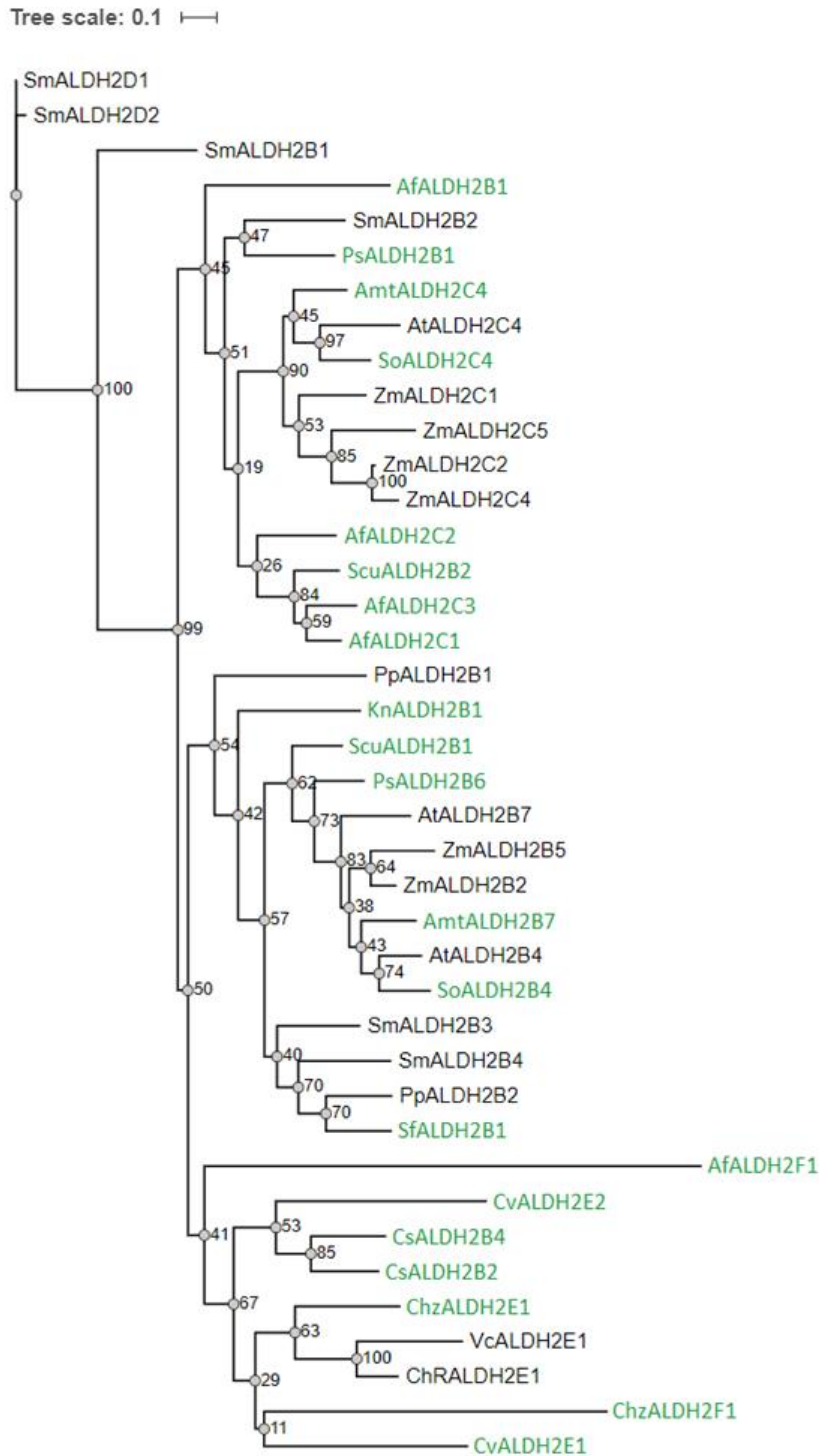
| Rodina ALDH      | Název genu      | NCBI protein ID | Označení sekvencí v databázi | Počet aminokyselin |
|------------------|-----------------|-----------------|------------------------------|--------------------|
| <i>Rodina 1</i>  | <i>ALDH1A2</i>  | AAL26232.1      | <i>DrALDH1A1</i>             | 518                |
|                  | <i>ALDH1A3</i>  | NP_001038210.1  | <i>DrALDH1A2</i>             | 513                |
|                  | <i>ALDH1L1</i>  | NP_001185701.1  | <i>DrALDH1L1</i>             | 903                |
|                  | <i>ALDH1L2</i>  | XP_002661418.2  | <i>DrALDH1L2</i>             | 923                |
| <i>Rodina 2</i>  | <i>ALDH2A1</i>  | AAM19352.1      | <i>DrALDH2A1</i>             | 516                |
| <i>Rodina 3</i>  | <i>ALDH3A2</i>  | NP_997814.1     | <i>DrALDH3A2</i>             | 488                |
|                  | <i>ALDH3B1</i>  | NP_775328.2     | <i>DrALDH3B1</i>             | 473                |
| <i>Rodina 4</i>  | <i>ALDH4A1</i>  | NP_957452.1     | <i>DrALDH4A1</i>             | 556                |
| <i>Rodina 5</i>  | <i>ALDH5A1</i>  | XP_009290813.1  | <i>DrALDH5A1</i>             | 522                |
| <i>Rodina 6</i>  | <i>ALDH6A1</i>  | NP_001002374.1  | <i>DrALDH6A1</i>             | 525                |
| <i>Rodina 7</i>  | <i>ALDH7A1</i>  | NP_997889.1     | <i>DrALDH7A1</i>             | 511                |
| <i>Rodina 8</i>  | <i>ALDH8A1</i>  | NP_001004540.1  | <i>DrALDH8A1</i>             | 487                |
| <i>Rodina 9</i>  | <i>ALDH9A1</i>  | NP_958879.1     | <i>DrALDH9A1</i>             | 508                |
| <i>Rodina 16</i> | <i>ALDH16A1</i> | NP_001073423.2  | <i>DrALDH16A1</i>            | 795                |
| <i>Rodina 18</i> | <i>ALDH18A1</i> | NP_001077015.1  | <i>DrALDH18A1</i>            | 782                |

#### 4.4 Fylogenetická analýza

Byla provedena fylogenetická analýza pro všechny rostlinné rodiny ALDH vyskytující se v databázi s výjimkou ALDH24 a ALDH19, které obsahovaly pouze jednu sekvenci. Délka stromových větví charakterizuje evoluční vzdálenost mezi jednotlivými geny.

Rostlinná rodina ALDH2 je velmi početnou zvláště u suchozemských rostlin a její nejobsáhlejší podrodiny jsou ALDH2B, ALDH2C a ALDH2D. Mnoho zástupců z oddělení řas se však podobnostně poměrně liší a řadí se tedy do samostatné rodiny ALDH2E, typické převážně právě pro řasy. ALDH, jež nejevily značnou podobnost vůči žádnému z dříve specifikovaných podrodin, byly zařazeny do podrodiny ALDH2F, kterou lze jasně pozorovat i na vytvořeném fylogenetickém stromě. (Obr. 31)

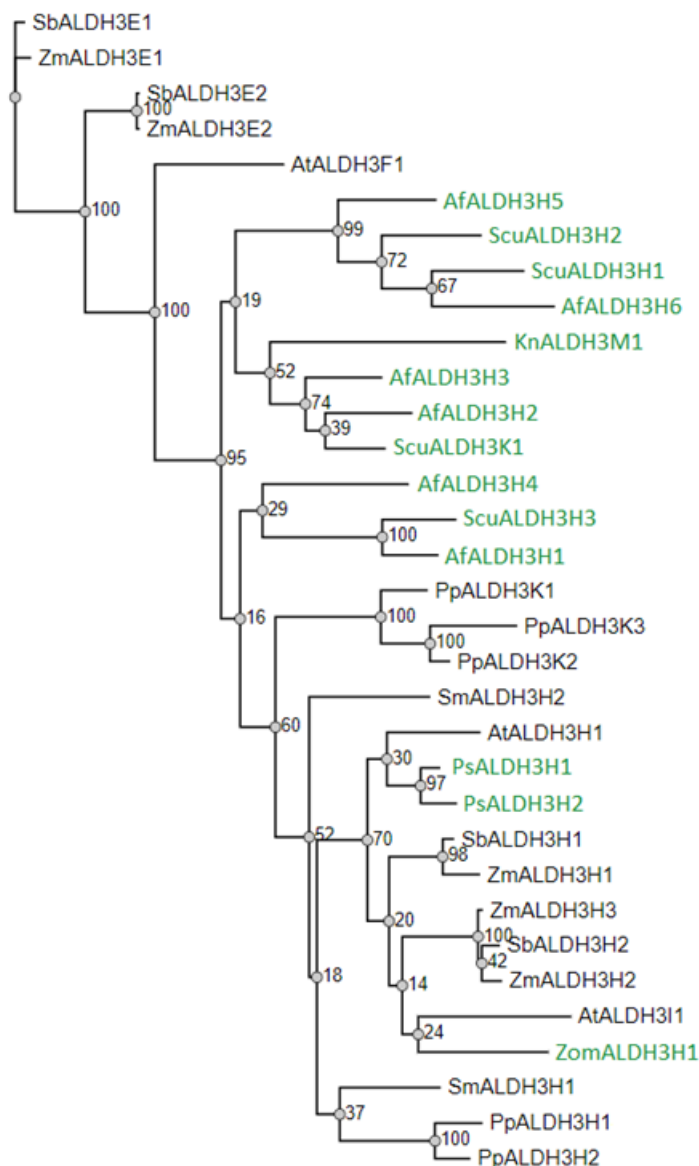




**Obr. 31** Fylogenetický strom rodiny ALDH2 pro vybrané rostlinné zástupce. Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě jsou značeny sekvence anotované v Brocker *et al.*, (2013), zeleně pak ostatní zkoumané sekvence. Pro tento strom bylo vybráno jen určité množství suchozemských rostlin, z důvodu hojnosti genů, jež tato rodina obsahuje. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt – *Amborella trichopoda*, At – *Arabidopsis thaliana*, Cs – *Coccomyxa subellipsoidea*, Cv – *Chlorella variabilis*, ChR – *Chlamydomonas reinhardtii*, Chz – *Chromochloris zofingiensis*, Kn – *Klebsormidium nitens*, Pp – *Physcomitrella patens*, Ps – *Picea sitchensis*, Scu – *Salvinia cucullata*, Sf – *Sphagnum fallax*, Sm – *Selaginella moellendorffii*, So – *Spinacia oleracea*, Vc – *Volvox carteri*, Zm – *Zea mays*.

ALDH3 je další velmi rozšířenou rodinou mezi suchozemskými rostlinami, u zástupců řas se však naopak téměř nevyskytuje. Při fylogenetické analýze bylo možné pozorovat párování jednotlivých rostlinných zástupců. Například kapradiny, mechy a jednoděložné i dvouděložné rostliny tvořily jasné podstromy ve struktuře stromu. Výjimkou byly ojedinělí zástupci z řas, kteří se příliš nepodobali žádné se zavedených podrodin a jejich sekvence byly tedy zařazeny do podrodin ALDH3L či ALDH3M.

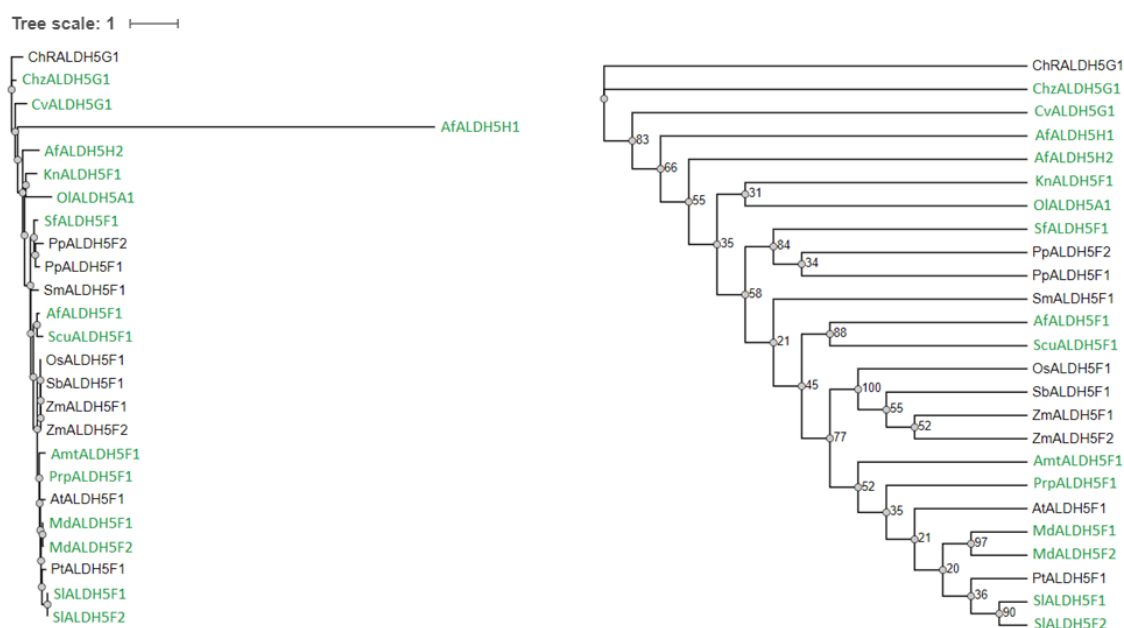
Tree scale: 0.1



**Obr. 32 Fylogenetický strom rodiny ALDH3 pro vybrané rostlinné zástupce.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě byly označeny sekvence analyzované v Brocker *et al.*, (2013), zeleně pak ostatní databázové sekvence. Stejně jako u rodiny ALDH2, bylo pro tento strom užito pouze části zástupců suchozemských rostlin, z důvodu jejich velkého množství. Měřítka odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, At – *Arabidopsis thaliana*, Kn -

*Klebsormidium nitens*, Pp – *Physcomitrella patens*, Ps - *Picea sitchensis*, Sb - *Sorghum bicolor*, Scu - *Salvinia cucullata*, Sm - *Selaginella moellendorffii*, Zm – *Zea mays*, Zom – *Zostera marina*.

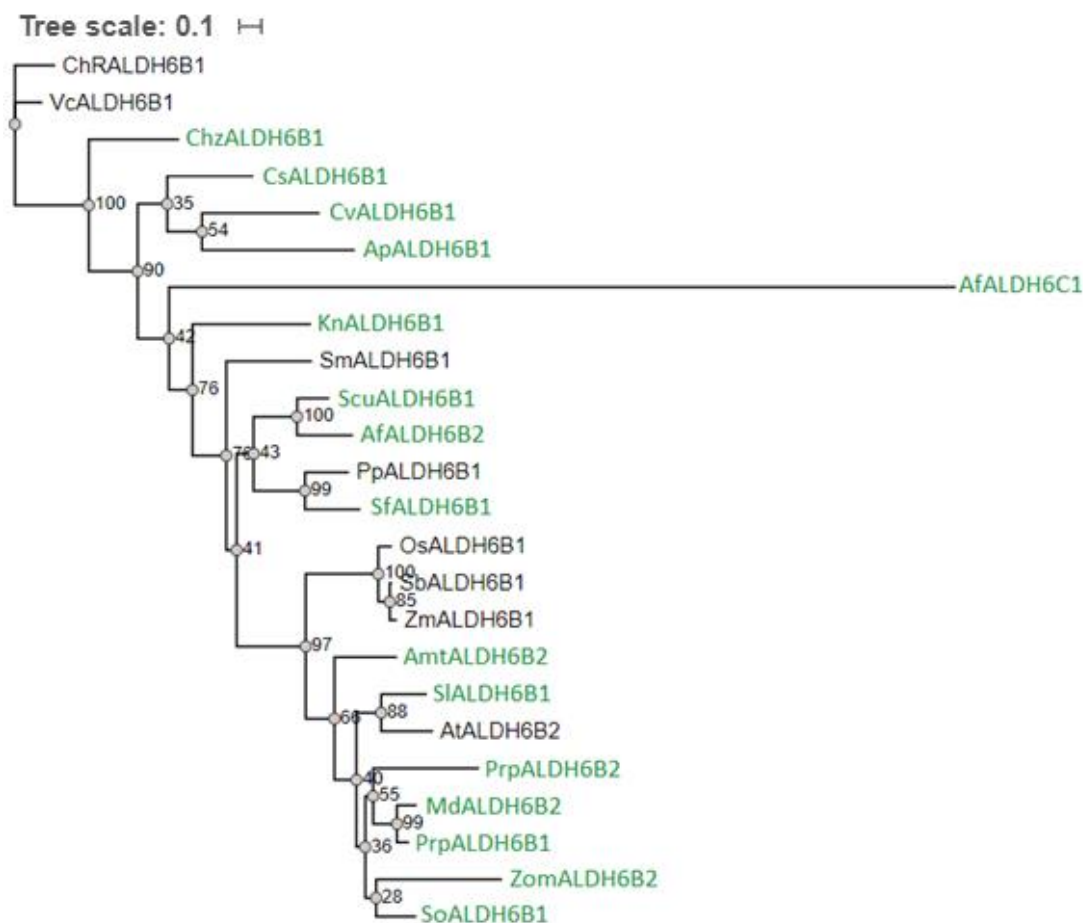
Fylogenetický strom rodiny ALDH5 ukazuje značnou podobnost všech rostlinných zástupců, kdy převažuje řazení do podrodiny ALDH5F, avšak některé řasy byly, z důvodu nižší sekvenční identity, umístěny do podrodiny ALDH5G. Je možné také pozorovat značnou odlišnost u zástupce z kapradiny *Azolla filiculoides*, a jelikož i identita této sekvence k ostatním zástupcům podrodiny 5G byla menší než 60%, byla zařazena do odlišné podrodiny než ostatní geny.



**Obr. 33 Fylogenetický strom pro rodinu ALDH5.** Z důvodu přehlednosti je uvedeno ohodnocení pomocí bootstrappingu ve speciálním dendrogramu. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Pro konstrukci těchto konkrétních stromů byla opět vynechána část rostlinných zástupců, z důvodu redundance. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt - *Amborella trichopoda*, At – *Arabidopsis thaliana*, Cv - *Chlorella variabilis*, ChR - *Chlamydomonas reinhardtii*, Chz - *Chromochloris zofingiensis*, Kn - *Klebsormidium nitens*, Md – *Malus domestica*, Ol - *Ostreococcus lucimarinus*, Os – *Oriza sativa*, Pp – *Physcomitrella patens*, Prp – *Prunus persica*, Pt - *Populus trichocarpa*, Sb – *Sorghum bicolor*, Scu - *Salvinia cucullata*, Sf - *Sphagnum fallax*, Sl – *Solanum lycopersicum*, Sm - *Selaginella moellendorffii*, Zm – *Zea mays*.

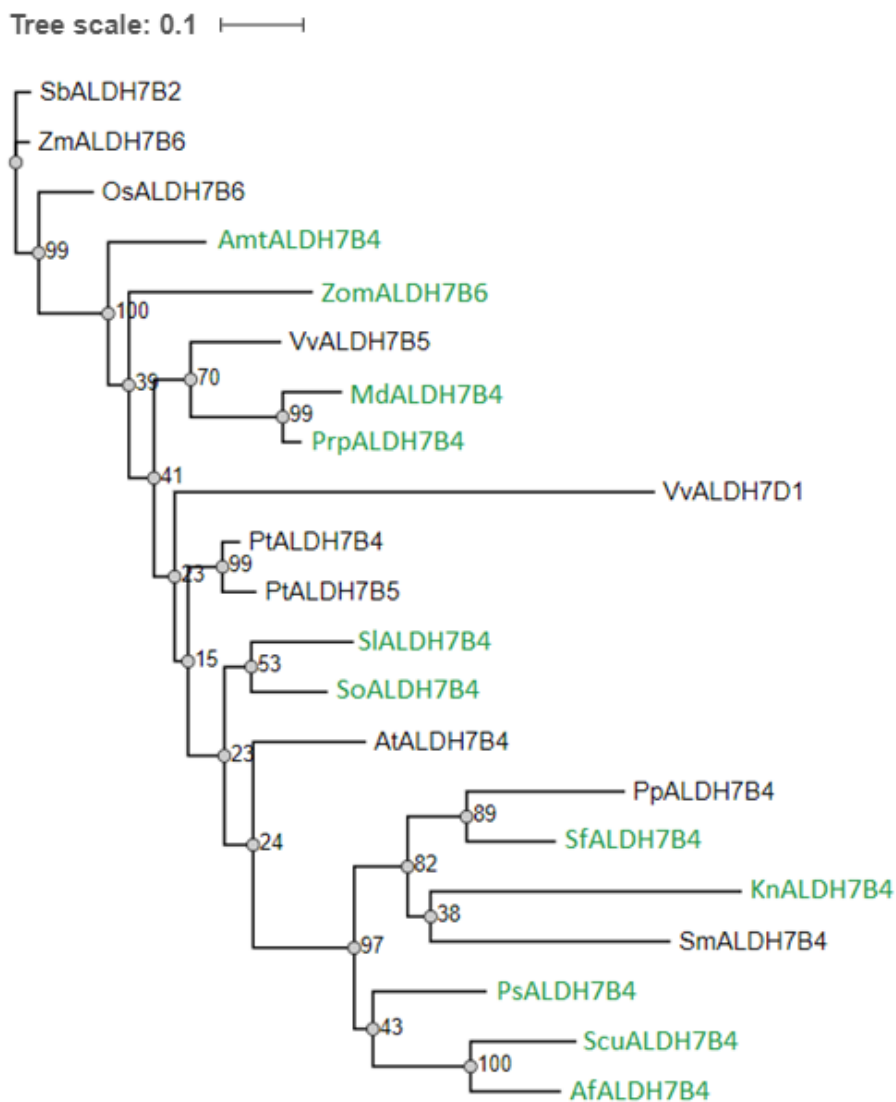
Geny rodiny *ALDH6* jsou poměrně konzervované napříč druhy a většina rostlinných sekvencí z databáze se tedy řadila do podrodiny ALDH6B. Kromě shlukování genů na základě tříd jako jsou jednoděložné a dvouděložné rostliny, můžeme také pozorovat značnou odlišnost genu *ALDH6* z kapradiny *Azolla filiculoides* a jeho vyčlenění oproti ostatním zástupcům. Tato sekvence byla na základě nomenklatury, a její

nízké podobnosti k ostatním zástupcům podrodiny ALDH6B, zařazena do separátní podrodiny ALDH6C. Fylogenetický strom toto rozhodnutí podporuje. (Obr. 34)



**Obr. 34 Fylogenetický strom genů rodiny ALDH6.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě jsou značeny sekvence anotované v Brocker *et al.*, (2013), zeleně pak ostatní zkoumané sekvence. Pro vyšší přehlednost byly vynechány sekvence z určitých rostlinných zástupců. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt - *Amborella trichopoda*, Ap - *Auxenochlorella protothecoides*, At – *Arabidopsis thaliana*, Cv - *Chlorella variabilis*, ChR - *Chlamydomonas reinhardtii*, Cs – *Coccomyxa subellipsoidea*, Chz - *Chromochloris zofingiensis*, Kn - *Klebsormidium nitens*, Md – *Malus domestica*, Os – *Oriza sativa*, Pp – *Physcomitrella patens*, Prp – *Prunus persica*, Sb – *Sorghum bicolor*, Scv - *Salvinia cucullata*, Sf - *Sphagnum fallax*, Sl – *Solanum lycopersicum*, Sm - *Selaginella moellendorffii*, So - *Spinacia oleracea*, Vc – *Volvox carteri*, Zm – *Zea mays*, Zom – *Zostera marina*.

Rodina ALDH7 je velmi konzervovaná a většina zkoumaných genů se řadila mezi zástupce podrodiny ALDH7B. Jedinou významně lišící se sekvencí byla jedna ze dvou ALDH7 z vinné révy, které byla zařazena do podrodiny ALDH7D v Brocker *et al.*, (2013). Její vyčlenění je možné pozorovat také na vzniklém fylogenetickém stromě (Obr. 35).



**Obr. 35** **Fylogenetický strom pro rodinu ALDH7.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt – *Amborella trichopoda*, At – *Arabidopsis thaliana*, Kn – *Klebsormidium nitens*, Md – *Malus domestica*, Os – *Oriza sativa*, Pp – *Physcomitrella patens*, Ps - *Picea sitchensis*, Prp – *Prunus persica*, Pt - *Populus trichocarpa*, Sb – *Sorghum bicolor*, Scu - *Salvinia cucullata*, Sf - *Sphagnum fallax*, Sl – *Solanum lycopersicum*, Sm - *Selaginella moellendorffii*, So - *Spinacia oleracea*, Vv – *Vitis vinifera*, Zm – *Zea mays*, Zom – *Zostera marina*.

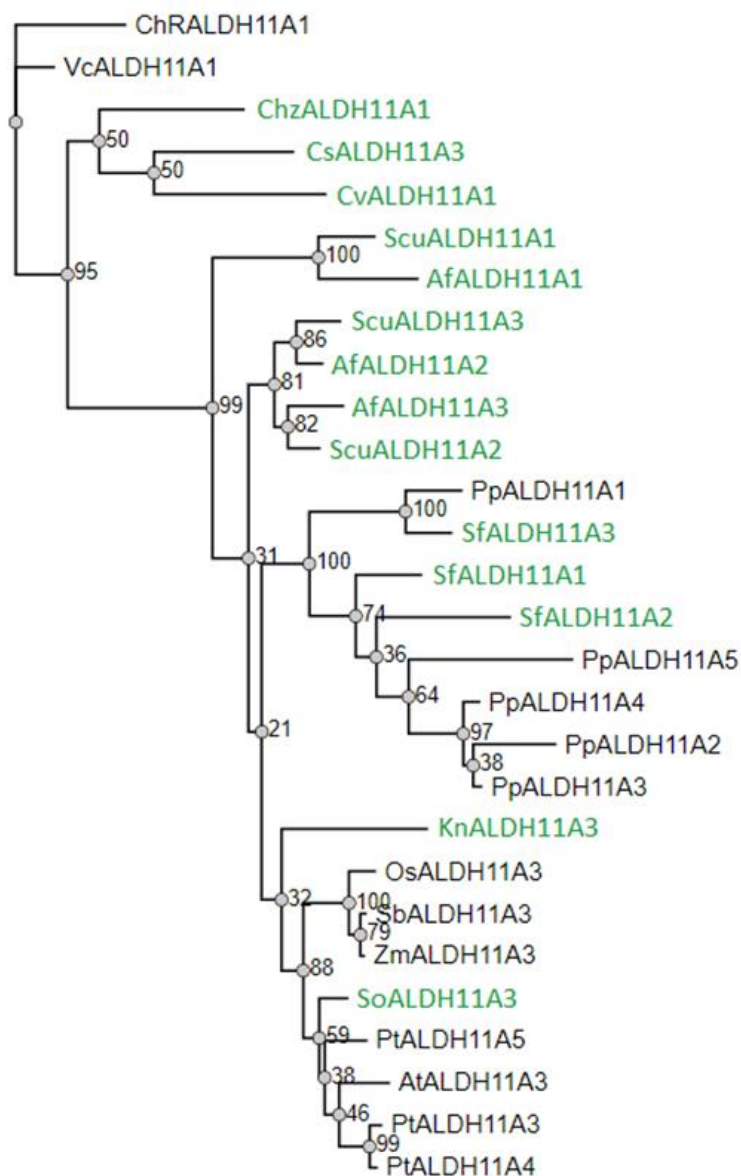
Sekvence rodiny ALDH10 jsou taktéž velmi konzervované a téměř všichni její rostlinní zástupci se řadili do podrodiny ALDH10A. Výjimkou je sekvence z řasy *Volvox carteri*, jež byla v Brocker *et al.*, (2013) zařazena do podrodiny ALDH10C a jeden gen obsažený ve vinné révě (*Vitis vinifera*), jež byl z důvodu odlišností zařazen do podrodiny ALDH10B.



**Obr. 36** Fylogenetický strom pro geny rodiny **ALDH10**. Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Pro vyšší přehlednost byly vynechány sekvence z určitých rostlinných zástupců. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Ap - *Auxenochlorella protothecoides*, At – *Arabidopsis thaliana*, Cs – *Coccomyxa subellipsoidea*, ChR - *Chlamydomonas reinhardtii*, Chz - *Chromochloris zofingiensis*, Kn – *Klebsormidium nitens*, Os – *Oriza sativa*, Pp – *Physcomitrella patens*, Ps - *Picea sitchensis*, Pt - *Populus trichocarpa*, Sb – *Sorghum bicolor*, Scu - *Salvinia cucullata*, Sf - *Sphagnum fallax*, Sl – *Solanum lycopersicum*, Sm - *Selaginella moellendorffii*, So - *Spinacia oleracea*, Vc – *Volvox carteri*, Vv – *Vitis vinifera*, Zm – *Zea mays*, Zom – *Zostera marina*.

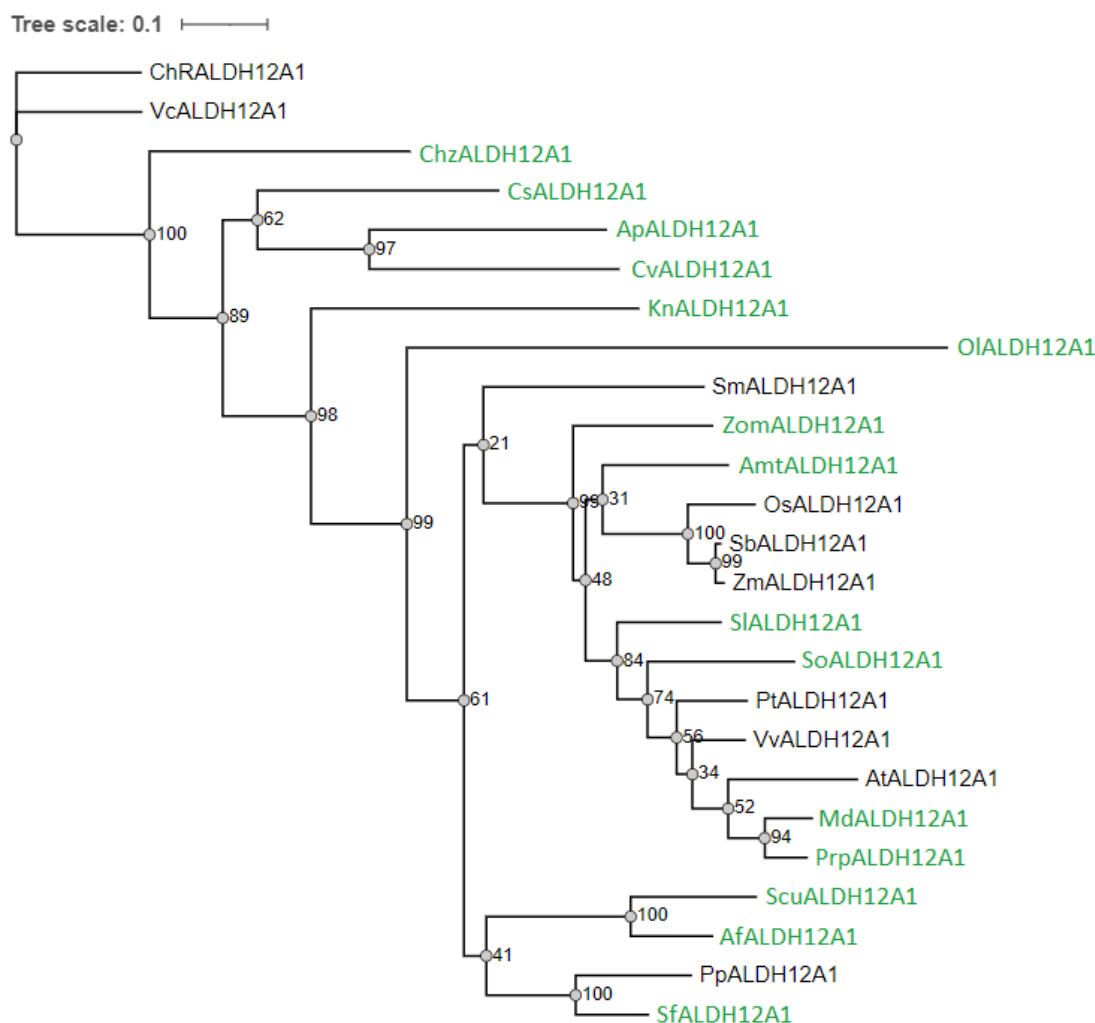
Většina genů *ALDH11* z databázových rostlin náležela na základě sekvenční podobnosti do podrodiny ALDH11A. Z fylogenetického stromu tedy žádný gen příliš nevystupuje a můžeme vidět klustrování jednotlivých zástupců kapradin, mechů či monokotů a dikotů.

Tree scale: 0.1



**Obr. 37 Fylogenetický strom rostlinných genů rodiny ALDH11.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Pro vyšší přehlednost byly vynechány sekvence z určitých rostlinných zástupců. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, At – *Arabidopsis thaliana*, Cs – *Coccomyxa subellipsoidea*, Cv – *Chlorella variabilis*, ChR – *Chlamydomonas reinhardtii*, Chz – *Chromochloris zofingiensis*, Kn – *Klebsormidium nitens*, Os – *Oriza sativa*, Pp – *Physcomitrella patens*, Pt – *Populus trichocarpa*, Sb – *Sorghum bicolor*, Scu – *Salvinia cucullata*, Sf – *Sphagnum fallax*, So – *Spinacia oleracea*, Vc – *Volvox carteri*, Zm – *Zea mays*.

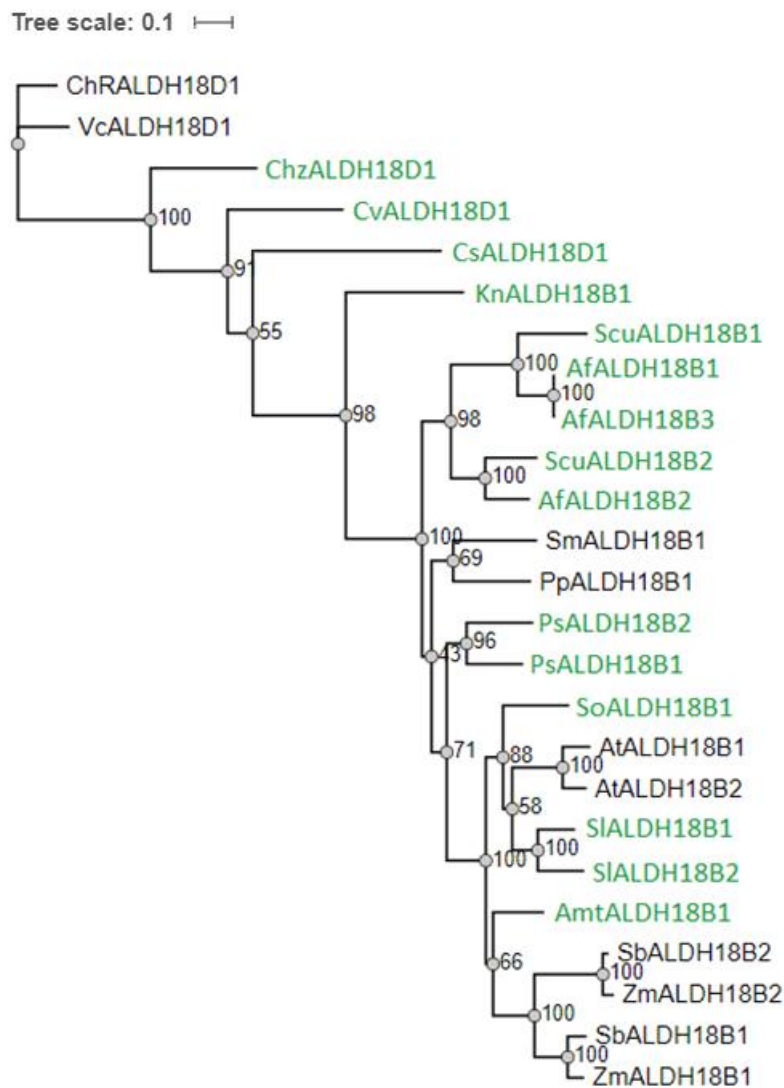
Všechny sekvence rodiny ALDH12 jsou značně konzervované a fylogenetický strom této rodiny je tedy poměrně rovnoměrně rozložený. Lze opět pozorovat shlukování jednotlivých rostlinných tříd a oddělení uvnitř stromu.



**Obr. 38 Fylogenetický strom pro rodinu ALDH12.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt – *Amborella trichopoda*, Ap – *Auxenochlorella protothecoides*, At – *Arabidopsis thaliana*, Cs – *Coccomyxa subellipsoidea*, Cv – *Chlorella variabilis*, ChR – *Chlamydomonas reinhardtii*, Chz – *Chromochloris zofingiensis*, Kn – *Klebsormidium nitens*, Md – *Malus domestica*, Ol – *Ostreococcus lucimarinus*, Os – *Oriza sativa*, Pp – *Physcomitrella patens*, Prp – *Prunus persica*, Pt – *Populus trichocarpa*, Sb – *Sorghum bicolor*, Scu – *Salvinia cucullata*, Sf – *Sphagnum fallax*, Sl – *Solanum lycopersicum*, Sm – *Selaginella moellendorffii*, So – *Spinacia oleracea*, Vc – *Volvox carteri*, Vv – *Vitis vinifera*, Zm – *Zea mays*, Zom – *Zostera marina*.

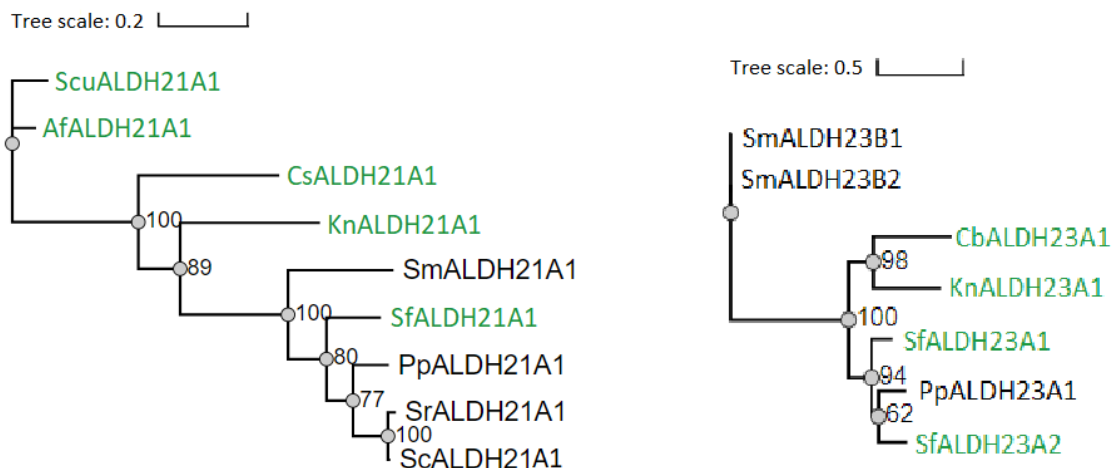


Rodinu ALDH18 databázových sekvencí můžeme rozdělit do dvou podrodin ALDH18B a ALDH18D. Většina řas byla zařazena do podrodina ALDH18D z důvodu vysoké odlišnosti. To můžeme pozorovat i na výsledném fylogenetickém stromě, kde se zástupci podrodiny ALDH18B shlukují, a naopak u sekvencí z řas dochází k vychýlení vzhledem k ostatním genům.



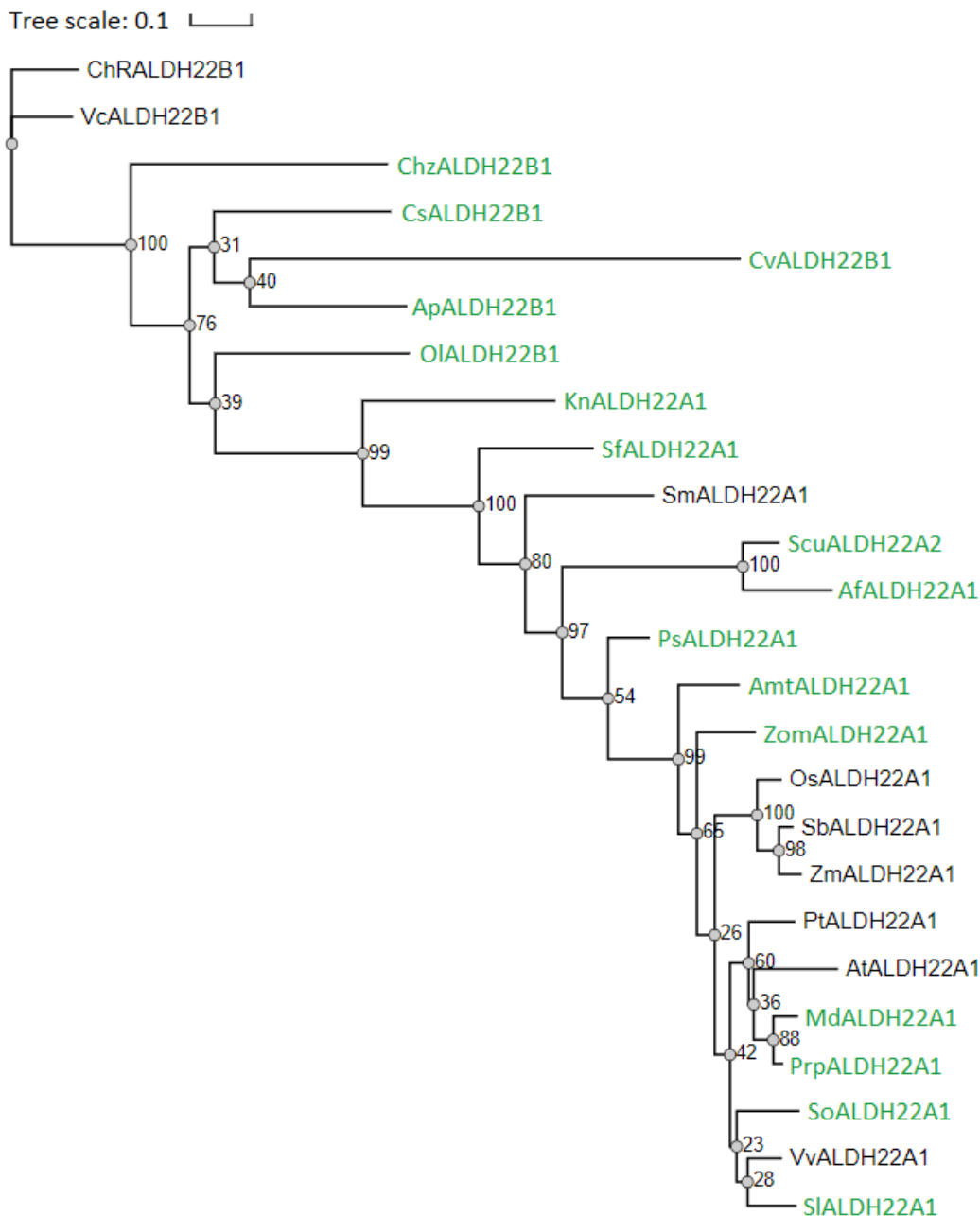
**Obr. 39 Fylogenetický strom pro rodinu ALDH18.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Pro vyšší přehlednost byly vynechány sekvence z určitých rostlinných zástupců. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt – *Amborella trichopoda*, At – *Arabidopsis thaliana*, Cs – *Coccomyxa subellipsoidea*, Cv – *Chlorella variabilis*, ChR – *Chlamydomonas reinhardtii*, Chz – *Chromochloris zofingiensis*, Kn – *Klebsormidium nitens*, Pp – *Physcomitrella patens*, Ps – *Picea sitchensis*, Sb – *Sorghum bicolor*, Scu – *Salvinia cucullata*, Sl – *Solanum lycopersicum*, Sm – *Selaginella moellendorffii*, So – *Spinacia oleracea*, Vc – *Volvox carteri*, Zm – *Zea mays*.

Zástupci rodiny ALDH21 se všichni řadily do jedné podrodiny a fylogenetický strom těchto genů neobsahuje přílišné odchylky. Lze pozorovat pouze mírné vyčlenění sekvencí z kapradin. U stromu genů rodiny ALDH23, můžeme naopak vidět jasné oddělení podrodiny ALDH23B z plavuně *Selaginella moellendorffii* od nižších zástupců. (Obr. 40)



**Obr. 40 Fylogenetické stromy pro rodiny ALDH21 a ALDH23.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Cb – *Chara braunii*, Cs – *Coccomyxa subellipsoidea*, Kn – *Klebsormidium nitens*, Pp – *Physcomitrella patens*, Sc - *Syntrichia caninervis*, Scu - *Salvinia cucullata*, Sf - *Sphagnum fallax*, Sm - *Selaginella moellendorffii*, Sr - *Syntrichia ruralis*.

V rodině ALDH22 se většina genů suchozemských rostlin řadí do podrodiny ALDH22A. Naopak sekvence z řas byly, kvůli své odlišnosti, zařazeny do nové podrodiny ALDH22B. Tyto rozdíly lze také sledovat pomocí fylogenetického stromu, jehož struktura odpovídá zvolené nomenklatuře. (Obr. 41)



**Obr. 41 Fylogenetický strom genů rodiny ALDH22.** Ohodnocení větví bylo stanoveno metodou „bootstrapping“. Černě značené geny byly anotovány v Brocker *et al.*, (2013), ostatní jsou značeny zeleně. Měřítko odpovídá evoluční odlišnosti sekvencí (udáváno v jednotkách počtu substitucí na část sekvence). Použité zkratky: Af – *Azolla filiculoides*, Amt – *Amborella trichopoda*, Ap - *Auxenochlorella protothecoides*, At – *Arabidopsis thaliana*, Cs – *Coccomyxa subellipsoidea*, Cv - *Chlorella variabilis*, ChR - *Chlamydomonas reinhardtii*, Chz - *Chromochloris zofingiensis*, Kn – *Klebsormidium nitens*, Md – *Malus domestica*, Ol - *Ostreococcus lucimarinus*, Os – *Oriza sativa*, Ps - *Picea sitchensis*, Prp – *Prunus persica*, Pt - *Populus trichocarpa*, Sb – *Sorghum bicolor*, Scu - *Salvinia cucullata*, Sf - *Sphagnum fallax*, Sl – *Solanum lycopersicum*, Sm - *Selaginella moellendorffii*, So - *Spinacia oleracea*, Vc – *Volvox carteri*, Vv – *Vitis vinifera*, Zm – *Zea mays*, Zom – *Zostera marina*.

## 5. Diskuse

Ve zkoumaných rostlinných zástupcích bylo identifikováno mnoho genů náležejících do rodiny ALDH2. Obecně je jednoduché sledovat trend, že docházelo k velkému rozmachu této rodiny s přechodem rostlin na souš, jelikož zkoumané suchozemské rostliny obsahovaly zpravidla velké množství zástupců této rodiny, přičemž u nižších rostlin jakými jsou řasy, se vyskytovaly maximálně dvě sekvence, které náležely mezi ALDH2. U suchozemských rostlin byly z velké části nalezeny pouze sekvence genů řadících se do podrodin ALDH2B, která obsahuje mitochondriální ALDH, metabolizující acetaldehydy, a ALDH2C, která je podrodinou cytosolických enzymů. Výjimku tvořila kapradina *Azolla filiculoides*, jež kromě těchto dvou nejpočetnějších podrodin obsahovala také sekvenci s nízkou identitou k ostatním, tvořící tak novou podrodinu, oproti ustanovené nomenklatuře v Brocker *et al.*, (2013), ALDH2F. Tato podrodina vznikla nejspíše již u řas, na základě diverzifikace sekvencí, kvůli nutnosti adaptace na určitý substrát či metabolickou dráhu. Tato potřeba nejspíše vymizela u většiny suchozemských rostlin a s ní i tato podrodina. V řasách se zpravidla vyskytovaly pouze sekvence z podrodiny ALDH2B nebo ALDH2E, specifikované v Brocker *et al.*, (2013) u zástupce *Chlamydomonas reinhardtii*. Některé, jako *Chromochloris zoofingiensis* a *Klebsormidium nitens*, však obsahovaly geny s podobností nižší než 60% k ostatním zástupcům těchto podrodin a byly také zařazeny do podrodiny ALDH2F. Genomy řas *Chara braunii* a *Ostreococcus lucimarinus* neobsahovaly žádné zástupce rodiny ALDH2.

Rodina ALDH3 je velmi diverzifikovanou rodinou, která obsahuje šest známých podrodin (Brocker *et al.*, 2013). Je také evolučně velmi zachovalá a u suchozemských rostlin se vyskytuje v hojném zastoupení. Všechny zkoumané rostliny obsahovaly sekvence z rodiny ALDH3 náležející do známých rostlinných podrodin, popsanych v Brocker *et al.*, (2013). U řas však z velké části tato rodina chybí, byla identifikována pouze u tří zástupců tohoto oddělení, a to *Chara braunii*, *Ostreococcus lucimarinus* (u obou chybí rodina ALDH2) a *Klebsormidium nitens*. Žádná z těchto sekvencí nejevila vysokou identitu se známými rodinami a byly rozřazeny do nových podrodin ALDH3L a ALDH3M. Sekvence z *Ostreococcus lucimarinus* byla navíc ve Stiti *et al.*, (2021) zařazena mezi ALDH3A, přestože tato podrodina obsahuje převážně živočišné geny. Na základě tohoto pozorování, můžeme vyvodit, že teorie, která říká, že k diverzifikaci a expanzi i této rodiny, došlo až při adaptaci vodních rostlin na suchozemské podmínky (Brocker *et al.*, 2013), bude nejspíše pravdivá.

Jelikož hrají geny rodiny ALDH5 důležitou roli v procesu metabolismu GABA, není překvapivé, že alespoň jednu sekvenci této rodiny obsahovala každá zkoumaná rostlina. Výjimku tvořil pouze nahosemenný smrk (*Picea sitchensis*), kde ALDH5 chybí. Tato rodina je poměrně konzervovaná, a tak se většina genů suchozemských rostlin řadí mezi ALDH5F. Genom kapradiny *Azolla filiculoides*, obsahoval tři sekvence této rodiny. Společně se zástupcem podrodiny ALDH5F však obsahoval také další sekvence, s nízkou podobností, které byly zařazeny do nové podrodiny ALDH5H. I řasy obsahovaly převážně zástupce podrodiny ALDH5F. Avšak vyskytly se také sekvence s vysokou podobností vůči nové ALDH podrodině anotované v Brocker *et al.*, (2013) u zástupce *Chlamydomonas reinhardtii*, a to ALDH5G. Ta byla identifikována u řas *Chlorella variabilis* a *Chromochloris zofingiensis*. Na základě nomenklatury ze Stiti *et al.*, (2021) byla ALDH5 z řasy *Ostreococcus lucimarinus* zařazena mezi geny ALDH5A, i když se opět jedná o podrodinu specifickou převážně pro živočišnou říši.

Většina zkoumaných rostlinných zástupců obsahovala také geny rodiny ALDH6, které všechny sdílely poměrně vysokou identitu (větší než 60%) a náležely tedy také do stejné podrodiny, ALDH6B. Výjimku zde tvořila opět kapradina *Azolla filiculoides*, kde kromě zástupce podrodiny ALDH6B, byla identifikována také sekvence s nízkou identitou vůči všem ostatním zkoumaným sekvencím a byla tedy zařazena jako zástupce nové podrodiny ALDH6C. Geny *ALDH6* nebyly obsaženy v řase *Ostreococcus lucimarinus* a smrku *Picea sitchensis*.

Jakožto sekvence vysoce konzervované rodiny, se geny *ALDH7* ve všech zkoumaných rostlinách řadily do podrodiny ALDH7B, a kromě v předchozí studii zkoumané *Vitis vinifera*, žádná rostlina neobsahovala geny podrodiny ALDH7D. Bylo také pozorováno, že většina řas, s výjimkou *Klebsormidium nitens*, zástupce této rodiny neobsahuje, což bylo stanoveno také v předchozí studii Brocker *et al.*, (2013), kde rodina chyběla v genomu řasy *Chlamydomonas reinhardtii* a spekulovalo se, o náhodné deleci genů, či evolučním tlaku, který by zapříčinil vymizení této rodiny u řas (Brocker *et al.*, 2013).

Zástupci rodiny ALDH10 se vyznačují vysokou homologií a ve všech zkoumaných rostlinách, kromě řasy *Chlorella variabilis*, kde geny této rodiny chyběly, byly tyto sekvence zařazeny mezi podrodinu ALDH10A. Ačkoli v Brocker *et al.*, (2013) byly identifikovány také další podrodiny jako ALDH10B, nalezena u vinné révy (*Vitis vinifera*), či ALDH10C u váleče (*Volvox carteri*), žádný zde zkoumaný rostlinný zástupce

nejevil podobnost k těmto podrodinám, či dostatečně nízkou podobnost k sekvencím ALDH10A, aby bylo možné geny rozřadit jinak.

Každá suchozemská rostlina obsahovala zástupce rodiny ALDH11. Všechny sekvence sdílely velmi vysokou identitu (vyšší než 70%) a řadily se tedy jednoznačně do podrodiny ALDH11A. Genomy řas na tom byly podobně, avšak *Auxenochlorella protothecoides* a *Chara braunii*, se vyznačovaly nepřítomností genů této rodiny. Podrodina ALDH11B, známa z Brocker *et al.*, (2013), kdy byly její geny identifikovány v genomu vinné révy (*Vitis vinifera*), nebyla nalezena u žádné další zkoumané rostliny.

Silně konzervovanou rostlinnou rodinou ALDH je ALDH12, která se zpravidla vyskytuje v každém rostlinném zástupci a všechny sekvence zkoumaných rostlin se na základě vysoké identity řadily pouze do jedné podrodiny ALDH12A. Musí tedy existovat silný evoluční tlak pro zachování funkčnosti této rodiny. Proto je překvapivé, že tato rodina chyběla v řase *Chara braunii* a smrku *Picea sitchensis*. U zmíněné řasy se může jednat o náhodnou delecí genu, či nepřítomnost z důvodu špatné sekvenace genomu. Naopak u smrku to může být něco příznačného pro nahosemenné rostliny, nebo opět se jednat o náhodnou delecí, jelikož více nahosemenných rostlin jsem v této práci nezkoumal.

Sekvence rodiny ALDH18 jsou taktéž poměrně konzervované a vyskytovaly se téměř v každé zkoumané rostlině, s výjimkou řasy *Chara braunii*. Na základě identity sekvencí, lze většinu rostlinných zástupců této rodiny zařadit do její nejpočetnější podrodiny ALDH18B. Výjimku tvoří některé řasy, a to *Chlorella variabilis*, *Chromochloris zofingiensis*, *Auxenochlorella protothecoides* a *Coccomyxa subellipsoidea*, které, společně s *Chlamydomonas reinhardtii*, která byla analyzována v Brocker *et al.*, (2013), se řadí do nové podrodiny ALDH18D, specifické pro oddělení řas.

Rodina ALDH21 byla nalezena pouze u některých řas, mechů a kapradin, kdy všechny její sekvence byly řazeny do podrodiny ALDH21A. Tyto výsledky souhlasí s tvrzením, že se tato rodina vyskytuje převážně u nižších rostlin. Podobně se hovoří o rodině ALDH23, ta byla však zastoupena pouze u zkoumaného mechu *Sphagnum fallax*, podobně jako v *Physcomitrella patens*, jehož ALDH byly dříve identifikovány v Brocker *et al.*, (2013), a řas *Chara braunii* a *Klebsormidium nitens*. Všechny byly opět řazeny do jediné podrodiny ALDH23A.

Rodina ALDH22, o které se v minulosti myslelo, že je specifická pro vyšší rostliny byla identifikována téměř v každé zkoumané rostlině. Na základě identity sekvencí byla

tato rodina rozdělena na dvě podrodiny ALDH22A, mezi jejíž zástupce se řadí převážně sekvence s vyšších rostlin, a ALDH22B, která je specifická pro zástupce z řas a byla identifikována již v Brocker *et al.*, (2013) u řasy *Chlamydomonas reinhardtii*.

Na základě těchto zjištění pak byla vytvořena databáze pro naši webovou aplikaci, která je s její pomocí schopná přiřadit sekvence ALDH do rodin a podrodin. Pro realizaci webového rozhraní byla vybrána kombinace jazyků PHP a JavaScript, jelikož jazyk PHP byl původně vytvořen speciálně pro programování webových stránek, je k dispozici zdarma a bez licenčních omezení na většinu nástrojů a usnadňuje mnohé interakce, které byly pro funkčnost webu PAFF podstatné, jako třeba práce s MySQL databází. JavaScript doplňuje funkčnost PHP na straně uživatele (v prohlížeči) a umožňuje stránce chovat se dynamicky bez nutnosti se znovu načítat (AJAX). Alternativou k programovacímu jazyku PHP, je licenčně výrazně striktnější platforma .NET, případně podobně limitovaná Java. Výhodou PHP oproti .NET a Javě je však nižší náročnost na zdroje, snadné zvládnutí jazyka, široké rozšíření na řadě menších, středních i velmi velkých projektů, plná podpora objektově orientovaného programování a také existence řady užitečných knihoven. Dále je výhodou rozsáhlá podpora v podobě dokumentace i komunity, a tedy i hledání řešení určitých problémů při implementaci, je často snadnější.

## 6. ZÁVĚR

Vytvořil jsem webové rozhraní s názvem PAFF (Plant ALDH Family Finder), které je schopné rozpoznat sekvence náležející do nadrodiny ALDH a určit jejich zařazení do rodin, na základě prohledávání vlastní databáze známých sekvencí. Tato databáze byla sestavena ze sekvencí ALDH dvaceti sedmi genomů rostlin a tří živočišných genomů. Přičemž zastoupení rostlin pokrývá jak jednoděložné, tak dvouděložné zástupce vyšších rostlin a databáze obsahuje také kapradiny, mechy i řasy.

Na základě identifikace rostlinných ALDH uložených v databázi je možné pozorovat expanzi nadrodiny ALDH u suchozemských rostlin oproti nižším rostlinám jako jsou například řasy. Mimo to byly tyto sekvence rozřazeny do rodin a podrodin, přičemž byly nalezeny také nové podrodiny, z důvodu nízké homologie některých genů ALDH. Ty byly nazvány ALDH2F, ALDH3L, ALDH3M, ALDH5H a ALDH6C. Správnost zvolené nomenklatury byla ověřena pomocí fylogenetické analýzy.



## 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. (1990) Basic local alignment search tool. *Journal of Molecular Biology*. s. 403–410.

Akaboshi S, Hogema BM, Novelletto A, et al., (2003) Mutational spectrum of the succinate semialdehyde dehydrogenase (ALDH5A1) gene and functional analysis of 27 novel disease-causing mutations in patients with SSADH deficiency. *Hum Mutat* **22**:442-50

Arikit S, Yoshihashi T, Wanchana S, Uyen TT, Huong NT, Wongpornchai S, Vanavichit A (2011) Deficiency in the amino aldehyde dehydrogenase encoded by GmAMADH2, the homologue of rice Os2AP, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). *Plant Biotechnol J* **9**:75–87

Baumgartner MR, Rabier D, Nassogne MC, et al., (2005) Delta1-pyrroline-5-carboxylate synthase deficiency: neurodegeneration, cataracts and connective tissue manifestations combined with hyperammonaemia and reduced ornithine, citrulline, arginine and proline. *Eur J Pediatr* **164**:31-6

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

Bradbury LM, Gillies SA, Brushett DJ, Waters DL, Henry RJ (2008) Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. *Plant Mol Biol* **68**:439–449

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.

Buchfink, Xie and Huson (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. **12** (1): 59–60.

Burnham, K. P.; Anderson, D. R. (2004), Multimodel inference: understanding AIC and BIC in Model Selection, *Sociological Methods & Research*, **33**: 261–304,

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540-552.

Codd EF (1970) A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*. **13**(6): 377-387

Cronk QC (2001) Plant evolution and development in a post-genomic context. *Nat Rev Genet* **2**:607–619

- Davis I., Yang Y., Wherritt D., Liu A., (2018) Reassignment of the human aldehyde dehydrogenase ALDH8A1 (ALDH12) to the kynurenine pathway in tryptophan catabolism. *Journal of Biological Chemistry* **293**(25), 9594-9603
- 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.
- Desper R, Gascuel O. (2002) Fast and Accurate Phylogeny Reconstruction Algorithms Based on the Minimum-Evolution Principle. *Journal of Computational Biology*, **9**:687–705
- Duester G. (2000) Families of retinoid dehydrogenases regulating vitamin A function: production of visual pigment and retinoic acid, *Eur. J. Biochem.* **267**, 4315–4324.
- Edgar RC (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Research*. **32** (5): 1792–97.
- Edgar RC (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity". *BMC Bioinformatics*. **5** (1): 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
- Farres J, Julia P, Pares X. (1988) Aldehyde oxidation in human placenta. Purification and properties of L-pyrroline-5-carboxylate dehydrogenase. *Biochem J* **256**:461-7
- Flanagan, David (2006). *JavaScript: The Definitive Guide: The Definitive Guide*. 5th ed., O'Reilly Media, Inc.. Gravenstein Highway North, Sebastopol, 1032 stran.
- Forte-McRobbie CM, Pietruszko R. (1986) Purification and characterization of human liver 'high Km' aldehyde dehydrogenase and its identification as glutamic gamma-semialdehyde dehydrogenase. *J Biol Chem* **261**:2154-63
- 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
- Geraghty MT, Vaughn D, Nicholson AJ, et al., (1998) Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. *Hum Mol Genet* **7**:1411-5
- Goloboff PA, Catalano SA. (2016) {TNT} version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics*, **32**:221-238
- Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood, *Syst. Biol.* **52**(5):696–704
- Henikoff S, Henikoff JG. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A.* **89**(22):10915-9.
- Higgins DG, Sharp PM. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*. **73**(1):237-44.

Historie PHP: <https://www.php.net/manual/en/history.php.php> (24.3.2021)

Hu CA, Lin WW, Valle D. (1996) Cloning, characterization, and expression of cDNAs encoding human delta 1-pyrroline-5-carboxylate dehydrogenase. *J Biol Chem* **271**:9795-800

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

Chen Y., Orlicky D.J., Matsumoto A., Singh S., Thompson D.C., Vasiliou V. (2011) Aldehyde dehydrogenase 1B1 (ALDH1B1) is a potential biomarker for human colon cancer, *Biochem. Biophys. Res. Commun.* **405**:173–179.

Chern MK, Pietruszko R. (1995) Human aldehyde dehydrogenase E3 isozyme is a betaine aldehyde dehydrogenase. *Biochem Biophys Res Commun* **213**:56-8

Jackson B., Brocker C., Thompson D.C., Black W., Vasiliou K., Nebert D.W., Vasiliou V. (2011) Update on the aldehyde dehydrogenase gene (ALDH) superfamily, *Hum. Genomics* **5**, 283–303.

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

JSON: <https://www.json.org/json-en.html> (24.3.2021)

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.

Kent, W. James (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research.* **12** (4): 656–664.

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

Knihovna jQuery: <https://jquery.com/> (24.3.2021)

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.

Končítíková R, Vigouroux A, Kopečná M, Šebela M, Moréra S, Kopečný D (2019) Kinetic and structural analysis of human ALDH9A1. *Biosci. Rep.* **39** (4), BSR20190558.

Kopečná M, Vigouroux A, Vilím J, Končítíková R, Briozzo P, Hájková E, Jašková L, von Schwartzberg K, Šebela M, Moréra S, Kopečný D. (2017) The ALDH21 gene found in lower plants and some vascular plants codes for a NADP<sup>+</sup>-dependent succinic semialdehyde dehydrogenase. *Plant J.* **92**(2):229-243.

Korasick D. A., Končítíková R., Kopečná M., et al., (2019) Structural and Biochemical Characterization of Aldehyde Dehydrogenase 12, the Last Enzyme of Proline Catabolism in Plants, *Journal of Molecular Biology*, **431**(3), 576-592

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

- Kurys G, Ambroziak W, Pietruszko R. (1989) Human aldehyde dehydrogenase. Purification and characterization of a third isozyme with low  $K_m$  for gamma-aminobutyraldehyde. *J Biol Chem* **264**:4715-21
- Lavenier, D. (2009). PLAST: parallel local alignment search tool for database comparison. *BMC Bioinformatics*. **10**: 329.
- Le, Si Quang; Gascuel, Olivier (2008) An Improved General Amino Acid Replacement Matrix, *Molecular Biology and Evolution*, **25** (7): 1307–1320
- Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, Gascuel O (2019) NGPhylogeny.fr: new generation phylogenetic services for non-specialists, *Nucleic Acids Research*, **47**(W1): W260–W265,
- Lerdorf R., Tatroe K., MacIntyre P. (2006) *Programming PHP*. 2nd ed., O'Reilly Media, Inc., Gravenstein Highway North, Sebastopol, 542 stran
- Lin SW, Chen JC, Hsu LC, et al., (1996) Human gamma-aminobutyraldehyde dehydrogenase (ALDH9): cDNA sequence, genomic organization, polymorphism, chromosomal localization, and tissue expression. *Genomics* **34**:376-80
- Lin M, Napoli JL. (2000) cDNA cloning and expression of a human aldehyde dehydrogenase (ALDH) active with 9-cis-retinal and identification of a rat ortholog, ALDH12. *J Biol Chem* **275**:40106-12
- Lipman, DJ; Pearson, WR (1985). Rapid and sensitive protein similarity searches. *Science*. **227** (4693): 1435–41.
- MacKerell AD Jr, Blatter EE, Pietruszko R. (1986) Human aldehyde dehydrogenase: kinetic identification of the isozyme for which biogenic aldehydes and acetaldehyde compete. *Alcohol Clin Exp Res* **10**:266 -70
- Madden T, Boratyn G, (2017) QuickBLASTP: Faster Protein Alignments. *Proceedings of NIH Research Festival*.
- Marchitti SA, Deitrich RA, Vasiliou V. (2007) Neurotoxicity and metabolism of the catecholamine-derived 3,4-dihydroxyphenylacetaldehyde and 3,4-dihydroxyphenylglycolaldehyde: the role of aldehyde dehydrogenase. *Pharmacol Rev*, 59125-50
- 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
- Marchitti S.A., Brocker C., Orlicky D.J., Vasiliou V., (2010) Molecular characterization, expression analysis, and role of ALDH3B1 in the cellular protection against oxidative stress, *Free Radic. Biol. Med.* **49**:1432–1443.
- Mills, P. B., Struys, E., Jakobs, C., Plecko, B., Baxter, P., Baumgartner, M., Willemsen, M. A., Omran, H., Tacke, U., Uhlenberg, B., Weschke, B. and Clayton, P. T. (2006) Mutations in antiquitin in individuals with pyridoxine-dependent seizures. *Nat. Med.* **12**, 307–309
- Missihoun TD, Kirch HH, Bartels D (2012) T-DNA insertion mutants reveal complex expression patterns of the aldehyde dehydrogenase 3H1 locus in *Arabidopsis thaliana*. *J Exp Bot* **63**:3887–3898

Mount, D. W. (2004) *Bioinformatics: Sequence and Genome Analysis*. 2nd. vyd. [s.l.]: Cold Spring Harbor Press.

Murphy JM, Mace PD, Eyers PA. (2017) Live and let die: insights into pseudoenzyme mechanisms from structure. *Curr Opin Struct Biol*. **47**:95–104.

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

Needleman, Saul B. & Wunsch, Christian D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*. **48** (3): 443–53.

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

Oehmen, C., Nieplocha, J. (2006) ScalaBLAST: A Scalable Implementation of BLAST for High-Performance Data-Intensive Bioinformatics Analysis. *IEEE Transactions on Parallel and Distributed Systems*. **17** (8): 740

Oguchi K, Tanaka N, Komatsu S, Akao S (2004) Methylmalonatesemialdehyde dehydrogenase is induced in auxin-stimulated and zinc-stimulated root formation in rice. *Plant Cell Rep* **22**:848–858

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

Pappa A, Estey T, Manzer R, Brown D, Vasiliou V, (2003) Human aldehyde dehydrogenase 3A1 (ALDH3A1): biochemical characterization and immunohistochemical localization in the cornea. *Biochem J* **376**:615–23

PhyML models: <http://www.atgc-montpellier.fr/models/index.php?model=lg> (24.3.2021)

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](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=ProgSelectionGuide) (24.3.2021)

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

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

Seung-Hye Hong, Ho-Phuong-Thuy Ngo, Hyun-Koo Nam, Kyoung-Rok Kim, Lin-Woo Kang, Deok-Kun Oh (2016) Alternative Biotransformation of Retinal to Retinoic Acid or Retinol by an Aldehyde Dehydrogenase from *Bacillus cereus*. *Applied and Environmental Microbiology*, **82**(13), 3940-3946

- Sheldon R., Moes G. (2005) *Beginning MySQL*. Wiley publishing, Inc., Indianapolis, Indiana 840 stran.
- Shin JH, Kim SR, An G (2009) Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol* **149**:905–915
- Shmueli O, Horn-Saban S, Chalifa-Caspi V, et al., (2003) GeneNote: whole genome expression profiles in normal human tissues. *C R Biol* **326**:1067-72
- Sophos NA, Vasiliou V (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact* **143–144**:5–22
- Stagos D., Chen Y., Brocker C., Donald E., Jackson B.C., Orlicky D.J., Thompson D.C., Vasiliou V. (2010) Aldehyde dehydrogenase 1B1: molecular cloning and characterization of a novel mitochondrial acetaldehyde-metabolizing enzyme, *Drug Metab. Dispos.* **38**, 1679–1687.
- 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
- Stiti N, Giarola V, Bartels D, (2021) From algae to vascular plants: The multistep evolutionary trajectory of the ALDH superfamily towards functional promiscuity and the emergence of structural characteristics. *Environmental and Experimental Botany*, **185**:104376
- Strausberg RL, Feingold EA, Grouse LH, et al., (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA* **99**:16899-03
- Strickland K.C., Krupenko N.I., Dubard M.E., Hu C.J., Tsybovsky Y., Krupenko S.A., (2011) Enzymatic properties of ALDH1L2, a mitochondrial 10-formyltetrahydrofolate dehydrogenase, *Chem. Biol. Interact.* **191**:129–136.
- Sunkar R, Bartels D, Kirch HH (2003) Overexpression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. *Plant J Cell Molecular Biol* **35**, 452–464
- Y. Tsybovsky, S.A. Krupenko, (2011) Conserved catalytic residues of the ALDH1L1 aldehyde dehydrogenase domain control binding and discharging of the coenzyme, *J. Biol. Chem.* **286**:23357–23367.
- Ullman, Chris; Dykes, Lucinda (2007): *Beginning Ajax*. 1st ed., Wiley Publishing, Inc., Indianapolis, Indiana, 498 stran.
- 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
- Vasiliou V, Bairoch A, Tipton KF, Nebert DW (1999) Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics* **9**, 421–434
- Vasiliou, V., Pappa, A. and Estey, T. (2004), 'Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism', *Drug Metabolism Reviews* Vol. **36**, pp. 279 – 299.

Vasiliou, V., Thompson, D. C., Smith, C., Fujita, M., & Chen, Y. (2013). Aldehyde dehydrogenases: from eye crystallins to metabolic disease and cancer stem cells. *Chemico-biological interactions*, **202**(1-3), 2–10.

Vasiliou V, Sandoval M, Backos DS, Jackson BC, Chen Y, Reigan P, Lanaspa MA, Johnson RJ, Koppaka V, Thompson DC. (2013) ALDH16A1 is a novel non-catalytic enzyme that may be involved in the etiology of gout via protein-protein interactions with HPRT1. *Chem Biol Interact.* **202**(1-3):22-31.

Vaz FM, Fouchier SW, Ofman R, et al., (2000) Molecular and biochemical characterization of rat gamma-trimethylaminobutyraldehyde dehydrogenase and evidence for the involvement of human aldehyde dehydrogenase 9 in carnitine biosynthesis. *J Biol Chem* **275**:7390-4

Využití PHP: <https://www.php.net/manual/en/intro-whatcando.php> (24.3.2021)

Wang X, Zhu S, Khan IA, Dasmahapatra AK. (2007) Ethanol attenuates Aldh9 mRNA expression in Japanese medaka (*Oryzias latipes*) embryogenesis. *Comp Biochem Physiol B Biochem Mol Biol* **146**:357-63

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

What is LAMP: <https://whatis.techtarget.com/definition/LAMP-Linux-Apache-MySQL-PHP>

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.

XAMPP: <https://sourceforge.net/projects/xampp/>

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

Yang, H., Zhang, D., Li, H., Dong, L. and Lan, H. (2015) Ectopic overexpression of the aldehyde dehydrogenase ALDH21 from *Syntrichia caninervis* in tobacco confers salt and drought stress tolerance. *Plant Physiol. Biochem.* **95**, 83–91.

Yoon KA, Nakamura Y, Arakawa H. (2004) Identification of ALDH4 as a p53-inducible gene and its protective role in cellular stresses. *J Hum Genet* **49**:134-40

Yoshida 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

Zhang Y, Mao L, Wang H, Brocker C, Yin X, Vasiliou V, Fei Z, Wang X (2012) Genome-wide identification and analysis of grape aldehyde dehydrogenase (ALDH) gene superfamily. *PLoS ONE* **7**:e32153

Zhao K, Chu X (2014). G-BLASTN: accelerating nucleotide alignment by graphics processors. *Bioinformatics.* **30** (10): 1384–91.