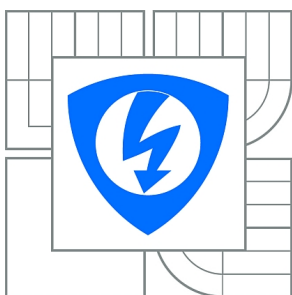




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF BIOMEDICAL ENGINEERING

ANALÝZA MITOCHONDRIÁLNÍCH GENŮ ŽIVOČICHŮ PRO DNA BARCODING

ANALYSIS OF ANIMAL MITOCHONDRIAL GENES FOR DNA BARCODING

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. KLÁRA BRABENCOVÁ

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. DENISA MADĚRÁNKOVÁ

BRNO 2014



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav biomedicínského inženýrství

Diplomová práce

magisterský navazující studijní obor
Biomedicínské inženýrství a bioinformatika

Studentka: Bc. Klára Brabencová

ID: 124991

Ročník: 2

Akademický rok: 2013/2014

NÁZEV TÉMATU:

Analýza mitochondriálních genů živočichů pro DNA barcoding

POKYNY PRO VYPRACOVÁNÍ:

1) Vypracujte literární rešerši na téma mitochondriální genom a DNA barcoding. 2) Z databáze GenBank sestavte dataset mitochondriálních sekvencí pro živočichy. 3) V libovolném programovém prostředí vytvořte funkci pro automatickou extrakci jednotlivých genů z mitochondriálních sekvencí ve formátu GenBank. 4) Pomocí funkce extrahujte z datasetu sekvencí jednotlivé geny pro proteiny. 5) Určete vnitrodruhové a mezidruhové variability pro jednotlivé geny. 6) Na základě odstupu vnitrodruhových a mezidruhových variabilit zhodnoťte, které geny či jejich části jsou vhodné pro DNA barcoding.

DOPORUČENÁ LITERATURA:

[1] SCHEFFLER, Immo E. Mitochondria. 2nd ed., Wiley-Blackwell, 2007, 472 s. ISBN 978-0-470-04073-7.

[2] BLAXTER, Mark. The promise of a DNA taxonomy. Phil. Trans. R. Soc. Lond. B., 2004, vol. 359, pp. 669-679.

Termín zadání: 10.2.2014

Termín odevzdání: 23.5.2014

Vedoucí práce: Ing. Denisa Maděránková

Konzultanti diplomové práce:

prof. Ing. Ivo Provazník, Ph.D.

Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Abstrakt:

Tato diplomová práce obsahuje literární rešerši na téma mitochondriální genom a DNA barcodingu. Praktická část se zabývá sestavením datasetu mitochondriálních sekvencí z databáze GenBank a vytvořením funkce pro extrakci jednotlivých genů, které jsou obsaženy v mitochondriálním genomu. Tato funkce byla vytvořena v programovém prostředí Matlab. DNA barcoding je metoda, která vytvořením čárového kódu života přiřadí každému živočišnému druhu na Zemi jeho specifickou a unikátní značku, podle které by mohl být daný jedinec snadno a rychle identifikován. Neexistuje ucelená práce zkoumající vhodnost jednotlivých mitochondriálních genů. Proto tato práce zkoumá potenciály ostatních mitochondriálních genů a hodnotí jejich účinnost pro DNA barcoding výpočtem jejich vnitrodruhových a mezidruhových variabilit.

Klíčová slova:

Mitochondriální genom, DNA barcoding, identifikace druhů, vnitrodruhová a mezidruhová variabilita

Abstract:

The aim of this work is a literature review on the topic of the mitochondrial genome and DNA barcoding, building a dataset of mitochondrial sequences from GenBank database and creation of a software function for extraction of individual genes that are present in the mitochondrial genome. This function was developed in Matlab. DNA barcoding is a method that uses short DNA sequence of mitochondrial genome for identification of species. There is no comprehensive work examining the appropriateness of different mitochondrial genes. This aim investigates the potential of other mitochondrial genes and evaluate their effectiveness for DNA barcoding and calculation of intra-and interspecific variability.

Keywords:

Mitochondrial genome, DNA barcoding, species identification, intra-and interspecific variability

Bibliografická citace mé práce:

BRABENCOVÁ, K. *Analýza mitochondriálních genů živočichů pro DNA barcoding* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2014. 64s. Vedoucí diplomové práce Ing. Denisa Maděránková.

Prohlášení

Prohlašuji, že svoji diplomovou práci na téma „Analýza mitochondriálních genů živočichů pro DNA barcoding“ jsem vypracovala samostatně pod vedením vedoucího semestrálního projektu a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedeného semestrálního projektu dále prohlašuji, že v souvislosti s vytvořením tohoto projektu jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne 23. května 2014

.....

podpis autora

Poděkování

Děkuji vedoucí diplomové práce Ing. Denise Maděránkové za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mého semestrálního projektu.

V Brně dne 23. května 2014

.....

podpis autora

Obsah

Seznam obrázků	6
Seznam tabulek	8
Úvod	9
1. Mitochondrie a její dědičnost	10
1.1 Evoluce a funkce mitochondrií	10
1.2 Mitochondriální dědičnost	12
1.2.1 Mitochondriální genom	12
1.2.2 Mitochondriální mutace	14
1.3 Taxonomie	16
1.3.1 Molekulární taxonomie	17
2. DNA barcoding	17
2.1 Cílová sekvence	18
2.2 Historie	18
2.3 Výhody a nevýhody DNA barcodingu	19
2.4 Databáze sekvencí	20
2.4.1 CBOL	20
2.4.2 Databáze BOLD	21
2.5 Modely evoluce nukleotidových sekvencí a konstrukce stromu	22

3. Potenciální účinnost mitochondriálních genů pro DNA barcoding	23
3.1 Jiný pohled na DNA barcoding	23
3.2 Použití formátu genbank	24
3.3 Výběr vhodného genu	24
4. Dataset mitochondriálních genomů	26
5. Programové řešení	28
5.1 Extrakce jednotlivých genů	28
5.2 Výběr sekvencí z datasetu pro výpočet variability	30
5.3 Výpočet vnitrodruhové a mezidruhové variability	31
6. Výsledky	34
6.1 Dendrogramy pro jednotlivé zkoumané regiony	40
6.2 Analýza celého mitochondriálního genomu	54
Závěr	56
Literatura	58
Seznam zkratk	62
Seznam příloh	64

Seznam obrázků

Obr. 1: Obecně přijímaná teorie původu mitochondrie.....	10
Obr. 2: Stavba mitochondrie	11
Obr. 3: Přeměna ADP na ATP	12
Obr. 4: Lidský mitochondriální genom	13
Obr. 5: Matroklinní dědičnost	15
Obr. 6: Strom života	16
Obr. 7: Systém DNA barcodingu	21
Obr. 8: Všechny geny, jejich názvy, pozice a sekvence.....	30
Obr. 9: Blokové schéma pro výpočet vnitrodruhové variability	32
Obr. 10: Blokové schéma pro výpočet mezidruhové variability.....	33
Obr. 11: Průměrná vnitrodruhová variabilita pro 13 genů, barcod a celou sekvenci výpočtem p-distance	36
Obr. 12: Průměrná mezidruhová variabilita pro 13 genů, barcod a celou sekvenci výpočtem p- distance.....	37
Obr. 13: Průměrná vnitrodruhová vzdálenost pro 13 genů, barcod a celou sekvenci výpočtem podle Kimura modelu.....	38
Obr. 14: Průměrná mezidruhová vzdálenost pro 13 genů, barcod a celou sekvenci výpočtem podle Kimura modelu.....	38
Obr. 15: Dendrogram mitochondriálního genu NAD1	40
Obr. 16: Dendrogram genu NAD2	41
Obr. 17: Dendrogram genu COX1	42
Obr. 18: Dendrogram genu COX2	43
Obr. 19: Dendrogram genu ATP8	44
Obr. 20: Dendrogram genu ATP6	45
Obr. 21: Dendrogram genu COX3	46
Obr. 22: Dendrogram genu NAD3.....	47
Obr. 23: Dendrogram genu NAD4.....	48
Obr. 24: Dendrogram genu NAD4L	49
Obr. 25: Dendrogram genu NAD5.....	50
Obr. 26: Dendrogram genu NAD6.....	51

Obr. 27: Dendrogram genu COB	52
Obr. 28: Dendrogram genu BARCODE	53
Obr. 29: Dendrogram pro celé mitochondriální genomy	54

Seznam tabulek

Tab. 1: Odlišnosti v kódování aminokyselin u mtDNA	14
Tab. 2: Popis formátu GenBank	26
Tab. 3: Stažené sekvence z NCBI	27
Tab. 4: Mezidruhová variability pro různé počty sekvencí.....	31
Tab. 5: Ukázka hodnot vnitrodruhové variability	32
Tab. 6: Mezidruhové a vnitrodruhové variability	35
Tab. 7: Správně určené dendogramy pro jednotlivé regiony	53

Úvod

Tato práce je zaměřená na metodu pro identifikaci živočišných druhů, a to metodu DNA barcodingu. Tato metoda je poměrně mladá, ale má velký potenciál. Zatím byly druhy identifikovány na základě morfologické stavby těla, tato metoda je ovšem první, která začíná živočišné druhy systematicky uspořádat na základě molekulární predispozice. Projevuje se v mnoha odvětvích, jako je evoluční biologie, kvalita životního prostředí, ochrana živočichů, taxonomie a dokonce i v identifikaci potravin. Na rozdíl od klasického morfologického porovnávání znaků je DNA barcoding rychlejší a levnější.

Vhodný úsek sekvence pro tuto metodu se nachází na mitochondriálním genomu. A to proto, že cílová sekvence by měla podléhat mutacím tak rychle, aby byla odlišná od sekvencí blízké příbuzných druhů a zároveň tak pomalu, aby se minimalizovaly rozdíly mezi jednotlivci uvnitř druhu. Zároveň je mitochondrií v buňce nespočet, tudíž i zastoupení mtDNA je v buňce větší než jaderné a získání mtDNA je snazší než získání jaderné DNA. V teoretické části této práci je rozebrána mitochondrie a hlavně věci týkající se dědičnosti a mitochondriálního genomu. Je zde zmínka o původu a vlastnostech dané organely, o mutacích, které zde probíhají, a dále se zde dozvíme něco o taxonomii. V samotném teoretickém rozboru DNA barcodingu, je řečeno, jaké jsou na tuto metodu kladeny požadavky a výhody či nevýhody oproti morfologickému popisu.

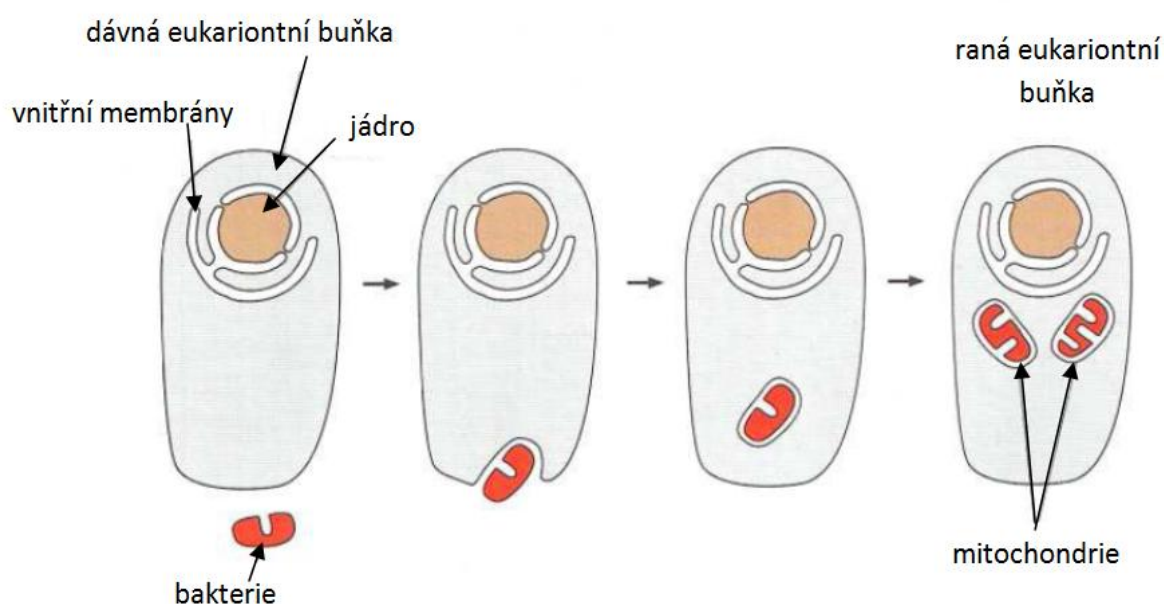
Praktická část se zabývá sestavením datasetu mitochondriálních sekvencí pro živočichy z databáze GenBank. Kompletní mitochondriální sekvence jsem stahovala na stránkách NCBI jako formát genbank, který je zde blíže popsán. V programovém prostředí Matlab, byla vytvořena funkce sloužící pro extrakci jednotlivých genů z datasetu sekvencí ve formátu GenBank a dále skript pro výpočet vnitrodruhových a mezidruhových variabilit, které jsou počítány dvěma způsoby, a to pomocí Kimura modelu nebo výpočtem jednoduchého p-distance.

Poslední důležitou částí je diskuze na téma potenciální účinnosti mitochondriálních genů. Zde jsou zmiňovány nevýhody formátu genbank, několik nezodpovězených otázek týkajících se DNA barcodingu a nároky, které jsou kladeny na čárový kód a samozřejmě výsledky, které zobrazují vhodnost a účinnost jednotlivých genů.

1. Mitochondrie a její dědičnost

1.1 Evoluce a funkce mitochondrií

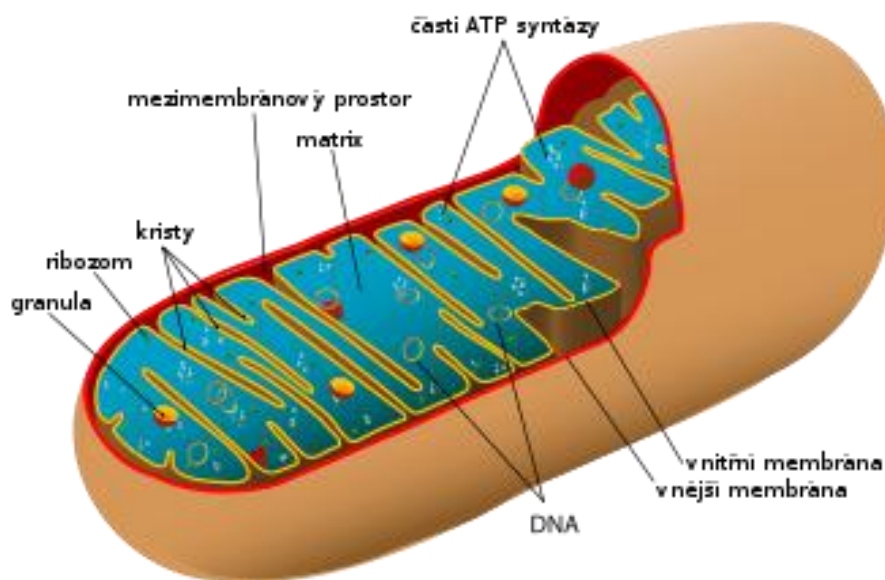
Mnoho výzkumů prokázalo, že v eukaryotických buňkách není jediným nositelem genetické informace jádro, ale kromě jádra obsahují malé množství genetické informace i mitochondrie a plastidy. Nejpravděpodobnější původ mitochondrií vysvětluje teorie endosymbiózy. Mitochondrie se pravděpodobně vyvinuly ze samotně žijících aerobních bakterií, které vstoupily do jiných anaerobních buněk. Díky přínosu enzymů dýchacího řetězce bakterie, mohla původně anaerobní buňka začít využívat kyslík k získání energie. Příčinou endocytózy byla nejspíš vzrůstající koncentrace kyslíku v atmosféře a lepší podmínky pro evoluci. V průběhu dalšího vývoje se mitochondrie určitým způsobem transformovala v semiautonómni organelu a nezbytnou součástí buněčného aparátu eukaryotické buňky. Dokonce se část mitochondriálního genomu přesunula do buněčného jádra, proto dnes obsahují jen zlomek své původní genetické informace, čímž se na eukaryotické buňce staly závislé a samostatného života neschopné.[3]



Obr. 1: Obecně přijímaná teorie původu mitochondrie
(převzato z [3])

Slovo mitochondrie pochází z řeckých slov *mitos* a *chondros*, což znamená vláknitá zrníčka, protože tak vypadají pod světelným mikroskopem. Jsou přítomny prakticky ve všech eukaryotických buňkách. Je jeden až několik μm dlouhá a v buňce se jich vyskytuje několik

stovek až tisíc. Každá mitochondrie se skládá ze dvou membrán. Vnější membrána má podobné vlastnosti, jako u ostatních membránových organel. Tvarově připomíná jednoduchý ovoid a není zvrásněna. Asi 40 % tvoří lipidy. Vnitřní mitochondriální membrána je zvrásněná a to buď do podoby jednoduchých výběžků (krist), nebo trubiček, které prochází skrz mitochondriální cytoplasmu (tubulární mitochondrie). Zvrásnění zvětšuje několikanásobně plochu membrány. Celková hmotnost lipidů je zde pouze 20 %. Zbytek tvoří transportní proteiny a proteiny dýchacího řetězce. Tyto membrány vymezují vnitřní prostor mitochondrie zvaný matrix a vnější mezimembránový prostor. V matrixu se nachází molekula kruhové mtDNA a proteosyntetický aparát. Dále je zde přítomna široká škála enzymů nezbytných při získávání energie oxidací glukosy.

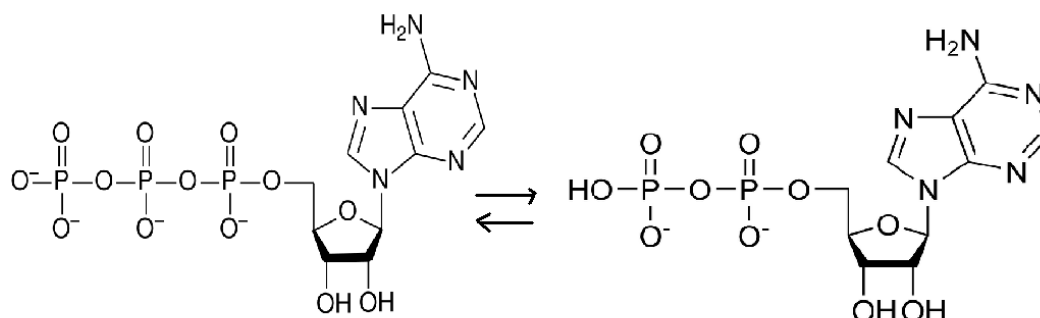


Obr. 2: Stavba mitochondrie

(převzato z <http://cs.wikipedia.org/wiki/Mitochondrie>)

Nejdůležitější mitochondriální funkcí je dodávat energii ve formě ATP. Tuto funkci realizuje mitochondrie systémem biologických oxidací, což jsou tři na sebe navazující pochody: Krebsův cyklus, oxidace vodíku v dýchacím řetězci a oxidativní fosforylace. V Krebsově cyklu je z organických látek uvolňován vodík, který je v dýchacím řetězci oxidován na vodu. Při této reakci je energie, získaná z přenosu elektronů, akumulována do makroergních fosfátových vazeb. Tento proces se označuje jako oxidativní fosforylace. Při něm dochází k přeměně adenosintrifosfátu (ADP) na energii bohatý adenosintrifosfát (ATP). Molekula ATP je univerzálního dárce energie pro všechny buněčné pochody. Přitom ATP se při odevzdání

energie mění na ADP, který vstupuje znovu do oxidativní fosforylace, kde je znovu přeměněn na ATP.



Obr. 3: Přeměna ADP na ATP

1.2 Mitochondriální dědičnost

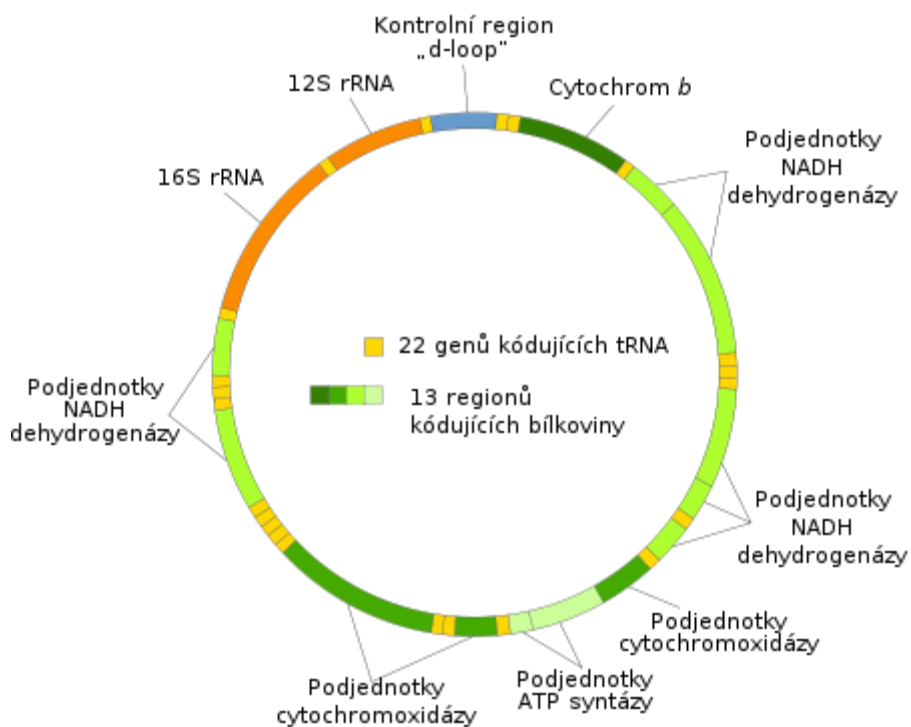
Jak už bylo řečeno, mitochondrie jsou semiautonorní organely eukaryotních buněk. Původ mají pravděpodobně v symbióze s prokaryotními buňkami (archebakteriemi), které ztratily během evoluce schopnost samostatné existence a staly se naopak nezbytnou součástí eukaryotních buněk. Mitochondriální genom (mtDNA) je nositelem mimojaderné dědičnosti a je prokaryotního typu. Je cirkulární, není vázaná s bílkoviny. Je uložena v kruhovém chromozomu. Genetický kód není vždy totožný s jaderným kódem. Mitochondriální DNA kóduje ribosomální RNA a transferovou RNA využívanou při transkripci a translaci v mitochondriích. Mitochondriální DNA se dědí materálně (po matce), protože mitochondrie v zárodku pocházejí většinou pouze z vajíčka (při oplození vajíčka jsou otcovské mitochondrie ze spermie zničeny). V ojedinělých případech se v cytoplasmě vyskytují mitochondrie mateřské i otcovské.[4]

1.2.1 Mitochondriální genom

Genetická informace je v mitochondriích uložena stejně jako v buněčném jádře ve formě DNA. Je zde však uložena volně v matrix nebo připojená k vnitřní membráně. Mitochondriální DNA je tvořena kruhovou molekulou (cirkulární tvar) a skládá se pouze z deoxyribonukleové kyseliny bez histonů. Díky těmto vlastnostem nemá žádný konec.

MtDNA je různě dlouhá u různých živočišných druhů. Průměrně je její velikost 16-19 kb. V porovnání s rostlinami a bakteriemi je to malé množství, protože velikost u rostlin je asi 150-2500 kb a u bakterií ještě větší. Lidská mtDNA obsahuje 16 569 bp, které tvoří 37 genů. Z toho 22 genů kóduje strukturu 22 druhů tRNA, ve dvou genech je uložena ribosomální

RNA a dalších 13 genů kóduje proteiny účastníci se oxidační fosforylace. Tyto polypeptidy tvoří 7 podjednotek NADH-CoQ z reduktázového komplexu, 3 podjednotky cytochromu c tvořící oxidoreduktázy, 2 podjednotky enzymu ATPazy a jednu podjednotku cytochromu b tvořící CoQ reduktázu. Protože neobsahuje žádné introny a je úsporně uspořádaná, je využíván veškerý obsah její genetické informace. [8]



Obr. 4: Lidský mitochondriální genom
(<http://cs.wikipedia.org/wiki/Mitochondrie>)

Mitochondriální DNA je, jak již víme, dvoušroubovice a skládá se ze dvou cirkulárních řetězců. Jsou to: H – řetězec neboli těžký řetězec, kde H je zkratka z angličtiny od slova heavy a druhým řetězcem je lehký L – řetězec (light). Těžký H - řetězec nese genetickou informaci pro obě rRNA, dále 14 molekul tRNA a 12 polypeptidů. Počátek tohoto řetězce se označuje jako *ori* H nebo také jako D loop. Je umístěn mezi úsekem kódujícím malou rRNA a úsekem kódujícím cytochrom b a obsahuje promotory jak pro H – řetězec, tak i pro L - řetězec. L – řetězec slouží jako matrice pro 8 molekul tRNA a zbývající polypeptid. Jeho počátek se označuje jako *ori* L. [6].

Všechny mRNA mitochondrií začínají kodonem ATG, který kóduje aminokyselinu methionin a končí stop-kodonem. Další zajímavostí u mRNA je také určitá odlišnost při kódování aminokyselin. Například kodon UGA slouží u klasické jaderné mRNA jako stop-kodon, ale u

mitochondriální mRNA je to kodon pro aminokyselinu tryptofan. Naproti tomu kodony AGA a AGG, sloužící normálně jako kodon pro arginin, u lidské mitochondriální mRNA vystupují jako stop-kodony.[7]

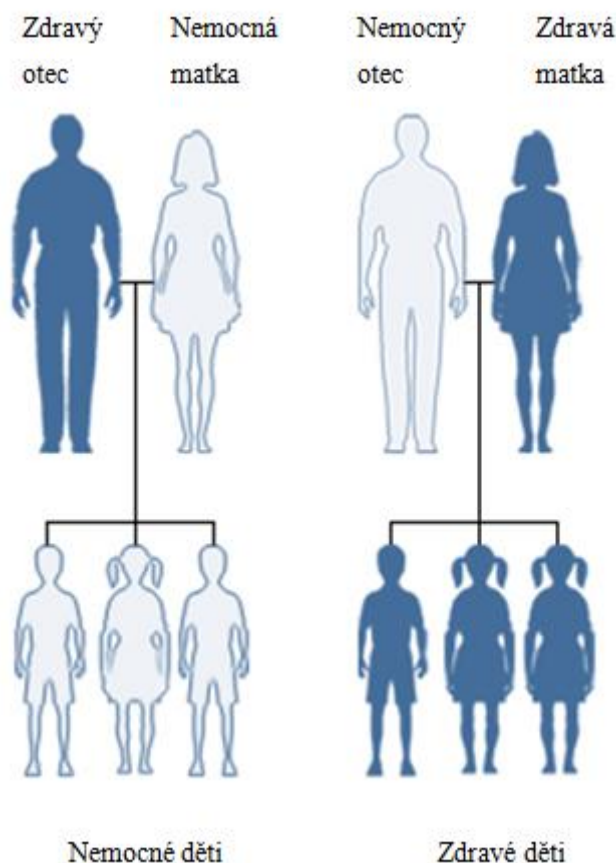
Tab. 1: Odlišnosti v kódování aminokyselin u mtDNA

Kodón	Univerzální kód	mtDNA u obratlovců	mtDNA u octomilky	mtDNA u kvasinky
UGA	Stop	Tryptofan	Tryptofan	Tryptofan
AUA	Izoleucin	Metionin	Metionin	Metionin
AGA	Arginin	Stop	Serin	Arginin

1.2.2 Mitochondriální mutace

V genetické informaci mitochondrií dochází velmi často k takzvaným neutrálním mutacím. Tyto změny v genetické informaci nemají žádný vliv na tvorbu a funkci proteinů a nezpůsobují žádné fenotypové projevy. Tyto mutace probíhají pravidelně, proto jsou analyzovány za pomoci různých metod a výsledky těchto analýz je možné použít ke stanovení evoluce daného druhu. Ze vzorků odebraných jedinců můžeme porovnávat výskyt určitých mutací v mtDNA a na základě rozložení mutací určit příbuzenskou závislost a také určit přibližnou dobu, kdy se tyto dvě linie jedinců od sebe oddělily v evolučním vývoji. Všeobecně platí, čím méně mutací mezi sebou dva jedinci mají, tím jsou si vývojově blíže.[4]

Přes zdánlivou nepodstatnost mohou být mutace v mtDNA zdrojem různých závažných dědičných chorob. Mluvíme zde o materální dědičnosti (nemendlovská dědičnost), protože všechny mitochondrie v zygotě jsou od matky. Proto, jestliže je matka přenašečkou mutací v mtDNA, pak tuto mutaci předá všem svým potomkům, zatímco otec žádnou mutaci svému potomkovi předat nemůže. Další zvláštností je, že při dělení buněk nedochází k rovnoměrnému rozdělení mitochondrií do dceřiných buněk, proto mohou získat zcela různý počet normálních a mutací zatížených mitochondrií. Homoplazmie je stav, kdy buňka dostane pouze normální nebo pouze mutované mitochondrie. Naopak při heteroplazmii obdrží různý počet normálních a mutovaných mitochondrií. V případě, že jedinec obdrží malý počet mutovaných mitochondrií, se nefunkčnost mitochondrií a následná choroba nemusí projevit. Obecně platí, že čím větší je podíl mutantních mitochondrií proti mitochondriím se zdravou mtDNA, tím jsou znaky choroby závažnější.[5]



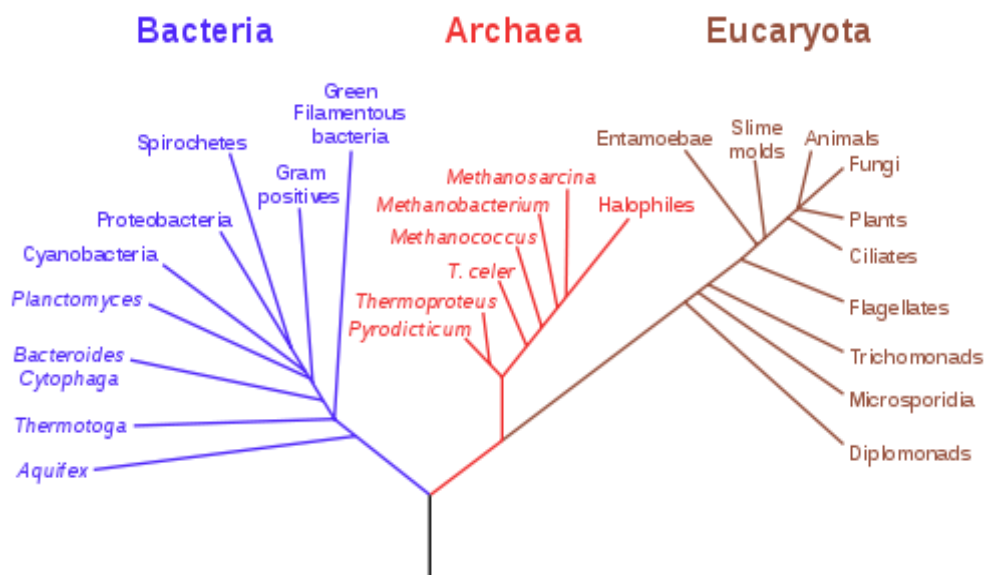
Obr. 5: Matroklinní dědičnost

(http://www.wikiskripta.eu/index.php/Matroklinní_dědičnost)

Genové mutace probíhají na úrovni vlákna DNA. Jsou to změny, které mění pořadí nukleotidů oproti normální sekvenci a jsou v rozsahu jednoho nukleotidu, proto jsou označovány jako bodové mutace. Rozlišujeme několik typů genových mutací. Za prvé je to adice nebo inserce, což je zařazení jednoho nebo více nadbytečných nukleotidů. Při této genové mutaci může dojít k posunu čtecího rámce a následně k syntetizování zcela odlišného polypeptidu. Další bodovou mutací je delece, což je přesný opak inserce. Jde o ztrátu jednoho nebo více nukleotidů původní sekvence. Účinek je podobný jako u adice, akorát dochází ke zkracování polypeptidového řetězce. Poslední genovou mutací je substituce. Substituce je náhrada či záměna báze původní sekvence bází jinou. Záměna báze purinové za bázi purinové nebo pyrimidinové za pyrimidinovou bázi, pak je tato substituce nazývána transice. Záměna purinové báze za pyrimidinovou a naopak se nazývá transverze.[9]

1.3 Taxonomie

Taxonomie nebo systematika se snaží katalogizovat diverzitu organismů a uspořádat je do hierarchicky uspořádaných skupin-příbuzné druhy do rodů, příbuzné rody do čeledí atd. Klasifikace, identifikace a názvosloví jsou tři odlišné, avšak příbuzné pododdělení taxonomie. Názvosloví je přiřazování jmen taxonomickým skupinám dle mezinárodních pravidel. Klasifikace je proces vytváření a charakterizace taxonomických skupin (taxonů) na základě podobnosti organismů nebo jejich vztahů. Konečně identifikace je proces určování a zařazování nových kmenů do jedné z již existujících a pojmenovaných skupin. Klasifikační schéma dovoluje rychlou charakterizaci nových kmenů na základě podobnosti s již pojmenovaným taxonem. Identifikační postup umožňuje potvrdit totožnost sledovaného kmene. A nakonec správné názvosloví umožňuje nejen vědeckou komunikaci, ale také má praktický význam. Umožňuje jednoznačné označení organismů. Domény jsou rozděleny na kmeny, kmeny na třídy, ty na řády, čeledi, rody, skupiny a druhy. Taxonomické úrovně jsou charakterizovány odlišnou příponou v názvu. Základní jednotka klasifikačního systému je druh. Druh je definován jako monospecifická a geneticky jednotná skupina jednotlivých organismů, které vykazují vyšší stupeň celkové podobnosti s ohledem na více nezávislých charakteristik, které jsou diagnostikovány na základě rozdílných fenotypových vlastností (Rosselló-Mora and Amann, 2001). Druhy mohou být rozděleny do dvou nebo více poddruhů, proto je poddruh nejnižší taxonomická jednotka s určitým postavením v názvosloví.



Obr. 6: Strom života

(<http://evolucezivota.wz.cz/>)

1.3.1 Molekulární taxonomie

Molekulární taxonomie sleduje molekulární znaky, které jsou uloženy v sekvencích informačních molekul, jako je DNA, RNA a proteiny. Primární struktura DNA je právě tou úrovní, na které vznikají evoluční novinky ve formě mutací. Právě DNA se předává z generace na generaci podle určitých pravidel, proto v sobě nese informaci o historii, identitu jedince a jeho příbuznost s ostatními jedinci v populaci, druhovou příbuznost a příbuzenský vztah tohoto druhu s jinými. Mnoho molekulárních znaků nezávisí na prostředí. Další výhodou molekulárních znaků je jejich obrovské množství a použitelnost na všech úrovních taxonomie. Molekulární znaky se dají jednoznačně popsat, protože nabývají několika diskrétních znaků a můžeme je považovat za molekulární hodiny, které nám mohou poskytnout odhady stáří.[10]

Organizmy se třídí podle jejich příbuzenského vztahu, tedy fylogeneze. Fylogenetika, která se zabývá rekonstrukcí fylogeneze, se snaží vystopovat pořadí větvení taxonů a zajímá se také o vývoje vlastností v rámci linií. Numerická taxonomie vyvinula matematické postupy, jak vypočítat celkovou podobnost nebo naopak odlišnost (distanci) mezi taxony. Přítomnost velkého množství polymorfismu na úrovni DNA (2 lidé ve 3 milionech bází) nejen v nekódujících, ale i v kódujících oblastech genomu vysvětluje teorie neutrální evoluce, což znamená, že většina znaků neovlivňuje fenotyp, jinak by byly přírodním výběrem eliminovány nebo naopak. To znamená, že molekulární znaky jsou často selekčně neutrální, nejsou ovlivňovány přírodním výběrem.[10]

2. DNA barcoding

Základní myšlenkou této metody je, že každý biologický druh obsahuje v genomu krátkou sekvenci, která je pro něj naprosto unikátní, jako například otisk prstu. Vytvoření čárového kódu života je snahou přiřadit každému živočišnému druhu na Zemi jeho specifickou a unikátní značku, podle které by mohl být daný jedinec snadno a rychle identifikován. Název „DNA barcoding“ má svůj původ v čárových kódech (bar codes) na výrobcích podle kterých se dá zboží naprosto přesně identifikovat. Čárový kód se skládá z 10 číslic na 11 pozicích, což znamená 10^{11} možných kombinací. Genetický kód se skládá pouze ze 4 písmen na jednu pozici, ale délka kódu je podstatně větší, proto máme k dispozici mnohem více kombinací. [12]

2.1 Cílová sekvence

Čtení celého genomu je pro velké množství dat časově náročné a prozatím to zvládne jen odborník. Proto Dr. Herbert položil základy použití genu COI, což je krátký úsek mitochondriálního DNA. Požadoval, aby daný úsek byl dost krátký pro rychlou analýzu, ale zároveň dost dlouhý na to, aby mohl spolehlivě identifikovat jednotlivé druhy, a musí to být stejná část stejného genu pro všechny živočichy. Výběr vhodného kódu v DNA není jednoduchý, protože některá místa v genomu nepodléhají téměř žádným změnám a jsou více konzervativní a jiná zase mají až příliš vysokou mutační rychlost, než aby se dala použít. [12]

Takováto sekvence by měla podléhat mutacím tak rychle, aby byla odlišná od sekvencí blízkce příbuzných druhů a zároveň tak pomalu, aby se minimalizovaly rozdíly mezi jednotlivci uvnitř druhu. Proto byla vybrána mitochondriální DNA před jadernou. Zároveň je mitochondrií v buňce nespočet, tudíž i zastoupení mtDNA je v buňce větší než jaderné a proto získání mtDNA je snazší než získání jaderné DNA. Mitochondriální DNA, jak už bylo řečeno, kóduje celkem 13 genů a neobsahuje introny. Jako nejvhodnější se zatím ukázal být gen COI, který navrhl pan Herbert pro DNA barcoding. Je to gen pro podjednotku I cytochrom c oxidázy, který je 648 bp dlouhý a protein, který kóduje, tvoří součást respiračního řetězce, kde katalyzuje redukci kyslíku na vodu. Tento gen má dvě výhody a to, že primery pro 5' konec sekvence COI jsou velmi kvalitní a spolehlivé, a proto umožní sekvenaci u většiny živočišných kmenů a za druhé se zdá, že tento gen má největší fylogenetický signál než jakýkoliv jiný.

2.2 Historie

V současnosti je popsána zhruba desetina všech biologických druhů na Zemi. Klasické taxonomické metody založené na morfologických a fyziologických znacích, mají svá omezení. Genotypová a fenotypová vnitrodruhová variabilita ve znacích používaných pro určování druhů může vést k chybným zařazením. Jelikož morfologické znaky jsou často efektivní pouze v určitém vývojovém stádiu nebo pohlaví, může dojít k chybnému nebo vůbec žádnému zařazení jedinců. [19]

Jako první o DNA barcodingu publikoval Dr. Paul Herbert v roce 2003, který pro posouzení vnitrodruhové variability navrhl hranici 2% a divergence mezi druhovými sekvencemi musí být 10x vyšší než uvnitř druhu. To znamená, že mezidruhová variabilita by měla přesáhnout 2-3% divergence, nicméně tato hranice nebyla ověřena pro všechny biologické druhy a může být zavádějící. [18]

Bylo navrženo i použití kratších úseků COI, o délce 100-200 bp pro identifikaci starých a špatně zachovalých muzejních exemplářů nebo pro částečně natrávených kousků

potravy v žaludku zvířat a výsledky vykazují, že ačkoliv volba konkrétní pozice krátké sekvence hraje významnou roli, lze říci i to, že analýza pomocí těchto krátkých úseků poskytuje srovnatelné výsledky jako analýza pomocí celé sekvence genu. [13]

2.3 Výhody a nevýhody DNA barcodingu

Komplikacím, které nastávají při klasickém morfologickém popisu, se můžeme pomocí DNA barcodingu vyhnout. Morfologický popis velmi podobných druhů a jejich odlišení od sebe je často tak obtížné, že biologa zaměstnává studium těchto příbuzných druhů téměř celý život. DNA barcoding je naopak rychlá, praktická a jednotná metoda mapování biodiverzity na Zemi. Tuto metodu je možné použít i pro identifikaci vzorků, které jinou metodou rozpoznat nedokážeme, jako např. máme-li v rukou nevylíhnuté vajíčko, popřípadě část těla jedince nebo zbytek potravy pro určení potravního řetězce. Metodou analýzy mtDNA se navíc můžeme něco dozvědět i o historii a předcích daného jedince a druhu. Na základě této metody bylo objeveno spousty skrytých druhů mezi zástupci jedinců považovaných za jeden druh.

Ovšem i tato metoda má své omezení a nevýhody. Tato metoda je používána na základě odlišné vnitrodruhové a mezidruhové variability, ale hranice nejsou přesně stanoveny a mohou se lišit. Některé studie říkají, že DNA barcoding není schopen odlišit druhy, které se od sebe oddělili téměř nedávno.[15]

Navíc neexistuje, žádný univerzální kód pro všechny říše organismů. Sekvence COI se ukázala, jako dobře využitelná v živočišné říši, ale u rostlinné je to složitější. Dokáže rozlišit některé druhy řas, ale u většiny suchozemských rostlin se ukazuje jako nepoužitelná pro nízkou variabilitu v sekvenci způsobenou pomalou mutační rychlostí. Změny v mitochondriální DNA často nejsou stejné jako v jaderné, a proto bude docházet velmi často k mylným zařazením jedinců do druhů. Došlo i k zpochybnění Herbertovi studie prohlášením, že existují dva druhy s identickými DNA barcody. Nakonec se ukázalo, že tyto kódy patří genu ND6, který nebyl navržen pro tuto metodu. [14]

Dnes se většina lidí přiklání k názoru, že DNA barcoding je třeba používat ve spolupráci s další taxonomickou metodou, aby byly všechny problémy, pochybnosti a chyby odstraněny a nedocházelo ke špatné identifikaci a klasifikaci druhu. Tato metoda je pouze takovým prvním sítím a můžeme ji považovat např. za nápovědu, že se objevil nový druh, ale nemůže tento druh popsat ani vymezit. [16]

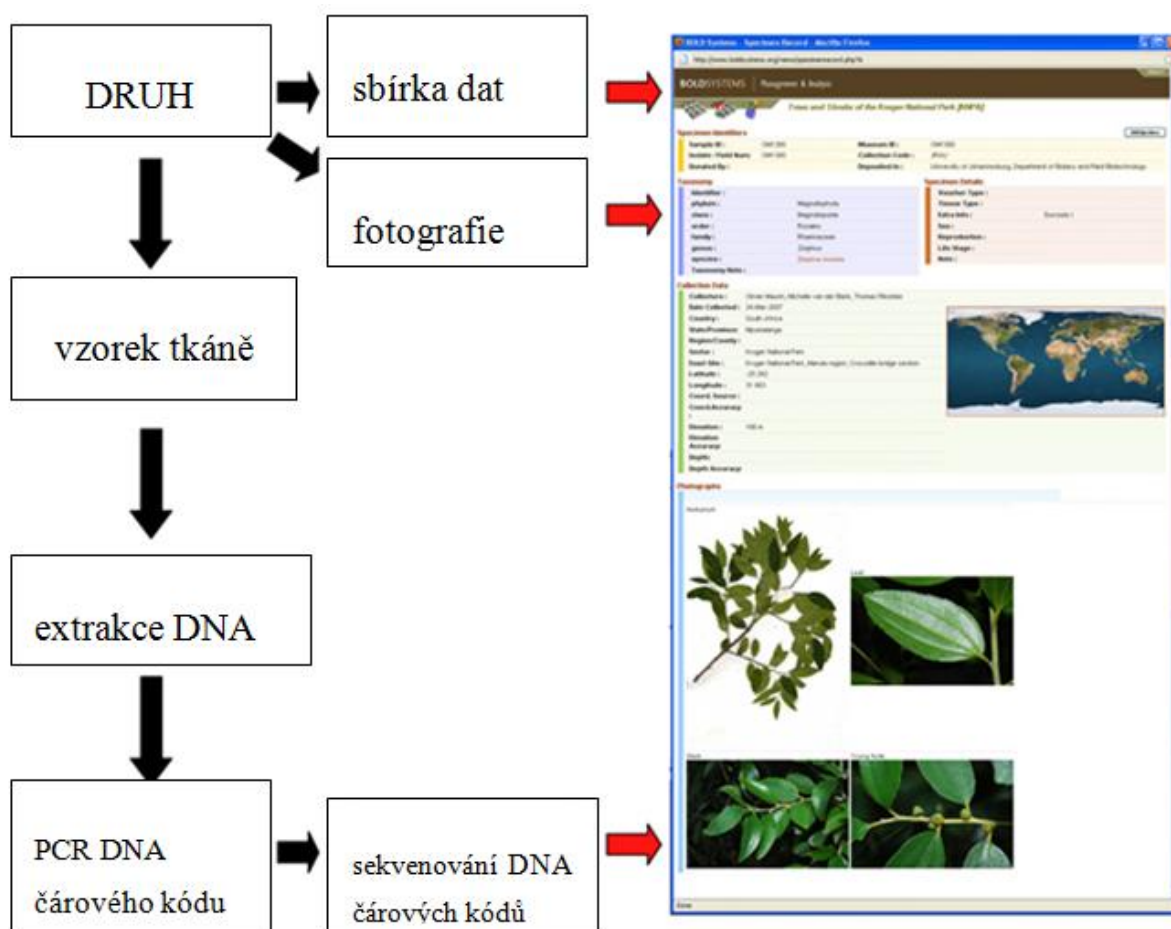
2.4 Databáze sekvencí

Aby mohl být druh správně identifikován a zařazen do druhu, rodu nebo řádu, musíme mít k dispozici komplexní a obsáhlou databázi sekvencí DNA. Nejrozsáhlejší a veřejně přístupná je například evropská EMBL, která je organizována Evropskou molekulárně biologickou laboratoří. Je to evropská primární nukleotidová databáze se sídlem v Anglii. Další databází je například americká GenBank při Národním centru pro biotechnologické informace (NCBI) nebo japonská DDBJ (The DNA Data Bank of Japan). Pro označení sekvence jako „barcode“ musí obsahovat řadu povinných polí a splňovat několik požadavků, které schválilo mezinárodní sdružení CBOL. Požadované prvky dat musí:

- obsahovat odkaz na vzorek, ze kterého byla získána s odkazem na dostupná data ve veřejné databázi o jedinci
- obsahovat odkaz na data o druhu v jednom ze zdrojů schváleného skupinami CBOL a NCBI
- obsahovat znak země, ze kterého sekvence pochází, podle pravidel GenBank
- pocházet z oblasti genomu, která byla CBOlem schválena jako barcode
- obsahovat alespoň 500 jednoznačně určených nukleotidů z obousměrného sekvencování v oblasti barcode. Pokud GenBank zažádá o uznání sekvence za čárový kód u kratší sekvence, následné směrnice odkazují na uznání CBOL.
- obsahovat ne více jak 1% nejednoznačných pozic nukleotidů na celou sekvenci
- obsahovat jméno oblasti použitého genu
- být spojeny s trasovacím souborem v NCBI Trace Archive nebo Ensembl Trace Server
- obsahovat sekvence všech použitých primerů použitých v PCR

2.4.1 CBOL

The Consortium for the Barcode of Life (CBOL) je mezinárodní sdružení, které se věnuje vývoji DNA čárových kódů jako globálnímu standardu pro identifikaci biologických druhů. Posláním CBOL je podpořit průzkum a vývoj DNA čárových kódů. CBOL podporuje rychlé sestavení vysoce kvalitních záznamů DNA čárových kódů ve veřejné knihovně DNA sekvencí, vývoj nových nástrojů a postupů, které učiní DNA barcoding levnější, rychlejší a více dostupnou metodu pro taxonomii a taxonomické výzkumné organizace ve všech regionech a zemích, a použití DNA barcodingu ve prospěch vědy a společnosti. Zahrnuje experty z oblastí genomiky, taxonomie, elektroniky a informačních technologií. Toto sdružení bylo založeno v květnu 2004 v USA. Pro jeho podporu pořádá řadu školení, seminářů a konferencí. [20]



Obr. 7: Systém DNA barcodingu
 (<http://worlduin.homestead.com/BARCODES.html>)

2.4.2 Databáze BOLD

Barcode of Life Data System (BOLD) je databáze sloužící k uložení, zveřejnění a analýze všech barcode sekvencí. Je to databáze sekvencí volně dostupná pro každého, tím tvoří základ pro studii DNA barcodingu. Probíhá zde správa, analýza a kontrola kvality barcodů. Tato databáze měla původně sloužit pro jeden konkrétní projekt DNA barcodingu, ale postupně byl čím dál více využíván i pro další účely a projekty, až se z něj stala mezinárodně uznávaná databáze. Databáze BOLD je velmi úzce spojena s dalšími veřejnými databázemi, jako jsou GenBank, EMBL, DDBJ. Mezi těmito databázemi dochází k výměně a převodům dat. První funkcí BOLD je identifikace neznámého druhu. Vložením sekvence do databáze, která musí být minimálně 300bp dlouhá, porovnáváme tuto sekvenci s existujícími záznamy. Další výhodou této databáze jsou nástroje, jako je např. multiple alignment, distanční metoda konstrukce stromu, nástroje pro BLAST atd. Při vkládání sekvencí musí proběhnout několik testů a poté je sekvenci přiřazena kvalitativní hodnota podle kritérií BOLD. Uložené záznamy musí obsahovat jak informace o sekvenci tak informace o vzorku. [16]

2.5 Modely evoluce nukleotidových sekvencí a konstrukce stromu

Evoluce organismů může být popsána kvantitativními modely. Modely umožňují zjistit podobnost dvou sekvencí, míru mutací vzniklých vývojem, a další. Nejjednodušší způsob modelování je zarovnat sekvence a spočítat rozdílná místa nebo podíl rozdílných míst k celkové délce sekvencí. Počet bodových mutací mezi dvěma sekvencemi vztažený na délku sekvencí je označován jako tzv. proporcionální vzdálenost (distance) sekvencí p-distance. Mezi předkem a potomky je evoluční vzdálenost d daná časem t . Kimurův model je druhý nejjednodušší evoluční model. Kromě pravděpodobnosti substituce jednoho nukleotidu za jiný hodnotíme i pravděpodobnost, že se změní nukleotid v rámci své chemické skupiny (puriny, pyrimidiny). Je založen na předpokladu, že transice (A-G, C-T) se objevují častěji než transverze (A-C, G-T). Je stanovena frekvence α pro transice a frekvence β pro transverze. Jedná se o dvouparametrický model (parametrem je α a β). Bývá označován jako K2P model. Je zanedbán možný rozdílný počet pyrimidinů a purinů v sekvenci. Celkový počet míst, která se změní, je součet mezi počtem nukleotidů, které se mění v důsledku transice a počtem nukleotidů, které se mění v důsledku transverze. Evoluční vzdálenost je:

$$d = -\frac{1}{2}\ln(1 - 2P - Q) - \frac{1}{4}\ln(1 - 2Q)$$

Evoluční vzdálenosti jsou nezbytným parametrem pro konstrukci fylogenetických stromů. Jejich hodnoty mohou být vypočteny na základě modelů (např. Jukesův-Cantorův model, Kimurův model). Výběr modelu samozřejmě určuje charakteristiku výsledného stromu. Z vícenásobného zarovnání můžeme jednoduše vypočítat evoluční vzdálenost mezi každými dvěma sekvencemi. Po té lze zkonstruovat matici s prvky, vzdálenostmi mezi sekvencemi druhů. Přesný výpočet této vzdálenosti vyžaduje nejprve vhodné zarovnání porovnávaných sekvencí – tzv. aligning. Jedná se o výpočetně velmi obtížnou úlohu. [29]

Fylogenetický strom je grafické znázornění příbuzenských vztahů mezi různými taxonomickými jednotkami. Příbuzenské vztahy se zde posuzují na základě morfologické či genetické podobnosti. Místo taxonomických jednotek mohou v některých stromech vystupovat přímo jednotlivé biologické druhy nebo i jednotlivé geny. Každý vrchol představuje určitou taxonomickou jednotku a hrana mezi dvěma vrcholy značí vztah mezi taxonomickými jednotkami, které tyto vrcholy reprezentují. V závislosti na typu fylogenetického stromu může délka hrany udávat dobu vývoje nebo míru podobnosti mezi příslušnými taxonomickými jednotkami. Nejčastěji používanou vzdálenostní metodou pro konstrukci fylogenetického stromu je spojování sousedů NJ (neighbor-joining). Metoda spojování sousedů NJ (neighbor-joining) provádí shlukování výběrem sousedních sekvencí, jejichž součet délek větví vůči ostatním sekvencím je nejmenší. [29]

3. Potenciální účinnost mitochondriálních genů pro DNA barcoding

Správný výběr genetického lokusu je základem účinnosti DNA barcodingu. Aktuální DNA barcoding zahrnuje použití 5' poloviny mitochondriálního genu pro cytochrom c oxidázu 1 (COI) pro identifikaci a vymezení druhů. Nicméně není znám žádný přesvědčivý a prioritní důvod pro zaměření se výlučně na tento region a bylo prokázáno, že pro některé skupiny zvířat analýza na základě této části genu funguje špatně. Byly vypočítány vnitrodruhové a mezidruhové variability vybraných organismů pro všechny proteiny kódující mitochondriální geny. Ideální sekvence pro DNA barcodingu by měla být dostatečně dlouhá, správně identifikovat a zařadit druhy a mezidruhová variabilita by měla být minimálně 10 krát větší než hodnota vnitrodruhové variability.

3.1 Jiný pohled na DNA barcoding

Identifikace druhů probíhá na základě výpočtu vzdáleností mezi sekvencemi, jsou proto stanovena kritéria tzv. odlišnosti, to znamená, jak daleko od sebe mohou být vzdáleny dvě sekvence, aby došlo ke správné identifikaci. Tyto kritéria pro vnitrodruhovou variabilitu jsou 3% pro hmyz a 2% pro ptáky a savce. [19]

Pokud analyzovaná sekvence odpovídá tomuto kritériu podobnosti, pak je vzorek identifikován, když ovšem bude odlišnost vyšší jak 3 %, zůstává vzorek jako neznámý a musí být vyšetřen taxonomy. Tato teorie tedy znamená, že DNA čárové kódy nejsou určeny pro studium evoluce taxonů a neposkytují o nich informace, spíše je to prostředek pro „ano“ a „ne“ identifikaci na základě předurčených skupin, o kterých již byly informace shromážděny.

Protože vznik nových druhů je dynamický a kontinuální proces, taxony, které nebudou odpovídat divergenčním pravidlům 3% nebo 2%, se tak stanou normou, ale ne výjimkou. Bylo by tedy potřeba navrhnout další ošetření, aby se tomuto problému dalo předejít, ale bylo by to nepraktické z hlediska univerzálnosti, použitelnosti a účinnosti, jak slibují vědci. Další otázkou, která může nastat je, že pokud se dva jedinci budou rozcházet v podobnosti mitochondriálního genomu, ale ne v jaderné DNA, patří do stejného druhu? Tuto otázku opět nemůžeme zvažovat, protože DNA barcoding nepoužívá jaderné údaje. Jako demonstrující příklad špatného určení druhů bych zmínila nesoulad při zkoumání rozdílu u afrických ryb z Viktoriina jezera, kde bylo u 9 ze 14 druhů zjištěno <1% mtDNA divergence, a to navzdory obrovské morfologické a ekologické rozdílnosti. [16]

3.2 Použití formátu genbank

Je všeobecně známé, že formát genbank může obsahovat mylné nebo nepřesné sekvence a proto by výsledky mohly být negativně ovlivněny třemi faktory, které však mohou nastat i u databází DNA čárových kódů. Za prvé, jakákoliv databáze DNA čárových kódů je podobná databázi Genbank v tom, že mnoho vědců do ní přispívá sekvencemi jako je tomu v případě NCBI, to znamená, že výskyt mylných sekvencí by měl být stejný jako v databázi DNA čárových kódů. Je nereálné se domnívat, že většina přídatků bude kontrolována taxonomickými odborníky.

Za druhé, čárové kódy DNA jsou generovány i na základě tkání v krych sbírkách, ačkoliv tyto sbírky též mohou obsahovat mylné vzorky. Takže opět jako v Genbank mohou tyto databáze obsahovat špatně určené sekvence.

A samozřejmě za třetí, každou strukturu stažené sekvence jsem pečlivě zvlášť prohlédla a sekvence, které se zdály být mylně předloženy nebo se jejich struktura lišila od daných norem, odstranila.

I přes tyto nedostatky jsem přesvědčena, že data z databáze Genbank v současné době poskytují nejlepší testovací vzorky, protože pouze tyto sekvence mají podobné profily jako databáze DNA čárových kódů.

3.3 Výběr vhodného genu

DNA barcoding by měl umožnit rychle, spolehlivě, automatizovaně a levně identifikovat druhy uživatelem s žádnými taxonomickými zkušenostmi. Identifikace obvykle probíhá porovnáním neznámých sekvencí se známými druhy DNA čárových kódů přes jejich vzdálenosti. Jednou z otázek, které jsou položeny v účinnosti DNA barcodingu, je výběr vhodného čárového kódu. Mezidruhová variabilita by měla být v této oblasti jasně větší než vnitrodruhová variabilita. Konkrétně 10 krát větší. Musí tvořit tzv. „barcoding gap“ neboli mezeru čárového kódu. [22]

Jak už bylo řečeno, mitochondriální genom se skládá ze 13 genů, které v této práci analyzujeme. Jsou to cytochromoxidázy podjednotek 1, 2 a 3, NADH dehydrogenázy podjednotek 1, 2, 3, 4, 4L, 5 a 6, ATPázy podjednotek 6 a 8 a nakonec cytochrom b.

Doposud nejpoužívanější DNA čárový kód je COI, nebo aspoň jeho část. Ačkoliv COI se již dlouho používá pro identifikaci v DNA barcodingu, do počátku neexistuje žádný závažný a prioritní důvod zaměřit se jen na tento specifický gen. Ve skutečnosti původce

DNA barcodingu nedal srovnat užitečnost a jedinečnost COI s jinými mitochondriálními geny.

Nicméně se nedá opomenout fakt, že COI fragment, který byl původně vybrán, má tu výhodu, že je lemovaný dvěma vysoce univerzálními primery, a proto je vhodný pro automatizaci při sběru dat DNA čárových kódů z nejrůznějších organismů. Jako důkaz jeho úspěšnosti je například dosažení 100% výsledku při studiu 260 druhů severoamerických ptáků. Ovšem existuje i důkaz o jeho selhání, při určování 449 druhů Diptera, kde dosáhl méně jak 70 % správně určených druhů, díky překrývání vnitrodruhové a mezidruhové variability. Je stále nutné hledat alternativní a spolehlivější DNA čárový kód, aby se zabránilo exkluzivní závislosti na COI. Proto jsou v této práci prozkoumány alternativní mitochondriální oblasti a je porovnána účinnost univerzálního COI barcodu s ostatními mitochondriálními geny kódující proteiny. [22]

Vzhledem k rostoucí dostupnosti kompletních mitochondriálních genomů, není výběr vhodného genu omezen dostupností univerzálních primerů. V této práci bych chtěla představit alespoň náznak hodnocení účinnosti všech 13 genů.

4. Dataset mitochondriálních genomů

Databáze je organizovaná sbírka informací. Veškerá data získaná sekvenováním a výsledky mnoha analýz a databází je přístupných veřejně na internetu. V USA je jednou z nejhojněji navštěvovaných adres NCBI (National Center for Biotechnology Information). Existují dva hlavní formáty sekvencí Fasta a GenBank. K řešení této práce byly sekvence celých mitochondriálních genomů stahovány ve formátu GenBank. Tento formát obsahuje velmi podrobnou informaci o sekvenci.

Tab. 2: Popis formátu GenBank

Locus	Základní vlastnosti sekvence (název, délka, typ, datum vložení)
Definition	Popis objektu
Accession	Databázové přístupové číslo
Version	Verze dané sekvence
Keywords	Pod kterými klíčovými slovy ji lze najít
Source organism	Zařazení v systému
Reference	Článek, kde byla daná sekvence publikována
Features	Podrobný popis jednotlivých genů včetně jejich pozic
Origin	Sekvence

Pro stahování mitochondriálních genomů jsem využila genetickou banku na stránkách NCBI (National Centre for Biotechnology Information). Jsou v ní uloženy informace o sekvencích (proteinových, NK, AK). Byla vytvořena v roce 1988 jako součást National Library of Medicine v NIH (Národní institut zdraví). Hlavním cílem NCBI bylo zřídit veřejné databáze, výzkumy v oblasti výpočetní biologie, rozvíjení softwarových nástrojů pro analýzu sekvencí a šířit biomedicínské informace. Kromě veřejně dostupné databáze obsahují stránky NCBI i spoustu vědeckých článků. Z databáze jsem náhodně vybrala druhy, které patří do kmenů, jako jsou strunatci nebo členovci a obsahují minimálně 3 mitochondriální genomy. U těchto druhů jsem postahovala všechny kompletní mitochondriální genomy, které byly v databázi dostupné. Tyto sekvence jsem stahovala a ukládala ve formátu genbank, který obsahuje mimo celé sekvence i informace o 13 genech. Pod názvem řádku CDS najdeme v tomto formátu název daného genu, pozici, kde se v celé sekvenci nachází a další informace.

V této tabulce najdeme jednotlivé postahované druhy. Dále obsahuje informace, do kterého kmenu, podkmenu popřípadě třídy daný druh patří a v posledním sloupci počet postahovaných sekvencí.

Tab. 3: Stažené sekvence z NCBI

název	kmen	podkmen x třída	počet sekvencí
<i>Bipes tridactylus</i>	Chordata	Reptilia(Plazi)	4
<i>Bos grunniens</i> (jak domácí)	Chordata	Mammalia(Savci)	74
<i>Bos taurus</i> (tur domácí)	Chordata	Mammalia(Savci)	190
<i>Branchiostoma belcheri</i>	Chordata	Cephalochordata(Bezlebeční)	16
<i>Bubalus bubalis</i> (buvol indický)	Chordata	Mammalia(Savci)	5
<i>Camelus bactrianus</i> (velbloud dvouhrbý)	Chordata	Mammalia(Savci)	5
<i>Canis lupus familiaris</i> (pes domácí)	Chordata	Mammalia(Savci)	110
<i>Canis latrans</i> (kojot prérijní)	Chordata	Mammalia(Savci)	4
<i>Carpiodes carpio</i> (kaprovec kaprovitý)	Chordata	Actinopterygii(Paprskoploutví)	4
<i>Drosophila simulans</i> (octomilka)	Arthropoda	Insecta(Hmyz)-třída	4
<i>Elephas maximus</i> (slon indický)	Chordata	Mammalia(Savci)	4
<i>Equus caballus</i> (kůň domácí)	Chordata	Mammalia(Savci)	217
<i>Erpetoichthys calabaricus</i> (Bichirek kalabarský)	Chordata	Actinopterygii(Paprskoploutví)	3
<i>Eumetopias jubatus</i> (lachtan ušatý)	Chordata	Mammalia(Savci)	10
<i>Gallus gallus</i> (kur bankivský)	Chordata	Aves(Ptáci)	59
<i>Gallus sonneratii</i> (kur Sonneratův)	Chordata	Aves(Ptáci)	3
<i>Hippoglossus hippoglossus</i> (platýs obecný)	Chordata	Actinopterygii(Paprskoploutví)	5
<i>Hippoglossus stenolepis</i> (platýs tichomořský)	Chordata	Actinopterygii(Paprskoploutví)	5
<i>Latimeria chalumnae</i> (latimérie podivná)	Chordata	Sarcopterygii(Nozdratí)	4
<i>Mammuthus primigenius</i> (mamut srstnatý)	Chordata	Mammalia(Savci)	10
<i>Mus musculus</i> (myš domácí)	Chordata	Mammalia(Savci)	126
<i>Myxocyprinus asiaticus</i> (kaprovec čínský)	Chordata	Actinopterygii(Paprskoploutví)	3
<i>Oncorhynchus mykiss</i> (pstruh duhový)	Chordata	Actinopterygii(Paprskoploutví)	6
<i>Oryzias latipes</i> (medaka japonská)	Chordata	Actinopterygii(Paprskoploutví)	13
<i>Procyon lotor</i> (mýval severní)	Chordata	Mammalia(Savci)	8
<i>Rattus norvegicus</i> (potkan obecný)	Chordata	Mammalia(Savci)	34
<i>Reinhardtius hippoglossoides</i>	Chordata	Actinopterygii(Paprskoploutví)	5
<i>Reticulitermes flavipes</i>	Arthropoda	Insecta(Hmyz)-třída	4
<i>Sus scrofa</i> (prase domácí)	Chordata	Mammalia(Savci)	77
<i>Taeniopygia guttata</i> (zebrčka pestrá)	Chordata	Aves(Ptáci)	6
<i>Theragra chalcogramma</i>	Chordata	Actinopterygii(Paprskoploutví)	11
<i>Tigriopus californicus</i>	Arthropoda	Crustacea(Korýši)	4

5. Programové řešení

5.1 Extrakce jednotlivých genů

Jedním z úkolů práce bylo vytvořit funkci pro extrakci jednotlivých genů z datasetu sekvencí ve formátu genbank. Tato funkce je vytvořena v programovém prostředí MATLAB R2008a, verze 7.6.0 a využívá Bioinformatics toolbox.

Byla vytvořena grafická uživatelská aplikace, která dovoluje uživateli načtení jím zvoleného souboru s mitochondriální sekvencí ve formátu genbank (*.gb). Načítání probíhá po řádcích. V načteném souboru jsou nalezeny klíčová slova „CDS“, která znamenají odkaz na pozice jednotlivých genů. Tyto pozice jsou uloženy i s názvy genů do proměnné.

Kompletní mitochondriální sekvence není načítána pomocí *genbankread*, protože tato funkce je nespolehlivá a velice často kompletní sekvenci nenačte. Proto podobně jako u vyhledávání genů, i pro celou sekvenci najdeme řádek, který obsahuje slovo „ORIGIN“, které ve formátu genbank značí začátek celého mitochondriálního genomu. Cílem je vygenerovat celou sekvenci, ve které bychom se potom mohli odkazovat na jednotlivé pozice genů.

Následuje ukázka, části souboru ve formátu genbank, který obsahuje pozice genů pod názvem „CDS“.

```
LOCUS      AY605477      16222 bp      DNA      circular VRT 29-AUG-2004

DEFINITION Bipes tridactylus voucher MVZ 236261 mitochondrion, complete
            genome.

ACCESSION  AY605477

            tRNA      1..70
                        /product="tRNA-Phe"

            rRNA      71..1023
                        /product="small subunit ribosomal RNA"

            tRNA      1024..1087
                        /product="tRNA-Val"

            rRNA      1088..2605
```

```

        /product="large subunit ribosomal RNA"

tRNA      2606..2680

        /product="tRNA-Leu"

        /note="codons recognized: UUR"

gene      2683..3638

        /gene="nad1"

CDS       2683..3638

        /gene="nad1"

        /note="TAA stop codon is completed by the addition of
        3' A residues to the mRNA"

        /codon_start=1

        /transl_except=(pos:3637..3638,aa:TERM)

        /transl_table=2

        /product="NADH dehydrogenase subunit 1"

        /protein_id="AAT08542.1"

        /db_xref="GI:47026727"

```

Za klíčovým slovem CDS je uvedena počáteční a koncová pozice nukleotidů v sekvenci, které tvoří kódovaný gen. Funkce pro extrakci jednotlivých genů je stěžejní funkcí pro další zpracování a výpočet mezidruhové a vnitrodruhové variability.

Do této funkce vstupuje jen jeden vstup a to `Filename`, jedná se o označení jednoho konkrétního mitochondriálního genomu druhu, pod kterým je genom uložen v databázi. Po načtení daného mitochondriálního genomu nalezneme v uložených informacích jednotlivé geny. Funkce postupně prochází každý řádek v souboru a porovnává hledaný textový řetězec se řádky proměnné. V momentě, kdy se objeví požadovaný textový řetězec, uloží čísla obsažená v tomto řetězci, což je pozice genu. Geny mají různá označení, které není jednotné, proto musí být ve vyhledávání několik podmínek. Například gen NAD6 je kódován na komplementárním vlákně mtDNA a proto, když v řetězci narazíme na textový řetězec ,komplement', tento gen budeme hledat na opačném vlákně. Objevují se zde i další problémy, jako například ,join' nebo ,<', což v prvním případě znamená, že gen je rozdělen na několika částí (většina mt genů pro většinu organismů je však tvořena jen jedním

úsekem), které leží jinde v sekvenci a je potřeba je spojit a v druhém případě, že pokračování struktury genu musíme hledat na začátku sekvence, protože, jak bylo řečeno, mitochondriální genom je tvořen kruhovou molekulou, nebo že část sekvence nebyla osekvenována.

Rozdílem počáteční a koncové hodnoty pozice dostaneme délku genů. V tuto chvíli jsou již vytvořeny sekvence jednotlivých genů a v posledním kroku funkce se vytváří struktury do proměnné výstup, kde se uloží informace o genech, a to konkrétně název druhu, jeho sekvence jednotlivých genů, názvy genů a počáteční a koncová pozice genů.

Nakonec podle vygenerovaných pozic najdeme v kompletní sekvenci jednotlivé geny, jejich celou strukturu a uložíme je do proměnné. Tato proměnná, která má rozměr struktury 3 x 13, obsahuje 13 genů, jejich názvy, pozice v kompletní sekvenci a jejich sekvence. Příklad, jak vypadá obsah proměnné viz. obr. 8.

1	2	3	4	5	6	7	8	9	10	11	12	13
'ND1'	'ND2'	'COX1'	'COX2'	'ATP8'	'ATP6'	'COX3'	'ND3'	'ND4L'	'ND4'	'ND5'	'ND6'	'CYTB'
[2683;3638]	[3849;4888]	[5255;6790]	[6951;7632]	[7699;7860]	[7851;8530]	[8531;9314]	[9383;9728]	[9797;100...	[10087;11...	[11668;13...	[13537;14...	[14119;15...
"gtgctatca...	"attaaccc...	"gtgtcactc...	"atagcaca...	"atgccaca...	"atgatatta...	"atggccca...	"atagccct...	"atatcaac...	"atgttaaaa...	"atgcttgat...	"gcactgtat...	"atggccca...

Obr. 8: Všechny geny, jejich názvy, pozice a sekvence

Tato funkce pro extrakci jednotlivých genů je stěžejní funkcí pro další zpracování a výpočet mezidruhové a vnitrodruhové variability.

5.2 Výběr sekvencí z datasetu pro výpočet variabilit

Z celého datasetu jsem pro výpočty variabilit vybrala od každého druhu čtyři zástupce, a to ze tří důvodů. Prvním důvodem je, že NCBI u některých druhů ani více kompletních mitochondriálních genomů nenabízela, naopak u 3 druhů byly k dispozici kompletní genomy jen tři, a proto jsem je do dalších výsledků nezahrnovala. Dalším důvodem byla potřeba zmenšit zastoupení sekvencí u některých druhů jako například *Bos taurus*, kde bylo staženo 190 kompletních mitochondriálních genomů a mohlo by dojít k chybné analýze, kde by proti sobě stály 4 sekvence jednoho druhu a 190 sekvencí druhého. Navíc by tyto výpočty byly zbytečně dlouhé. A posledním a hlavním důvodem bylo, že jsem v průběhu výpočtů mezidruhové variability zjistila, že výsledky pro 4 sekvence se od výsledků pro 10 a více sekvencí liší jen v rámci tisícín, ovšem výsledky pro 3 sekvence se už rozcházejí značně. Tento jev je názorně vidět v tabulce č. 5, která obsahuje hodnoty pro mezidruhovou variabilitu genu NAD1.

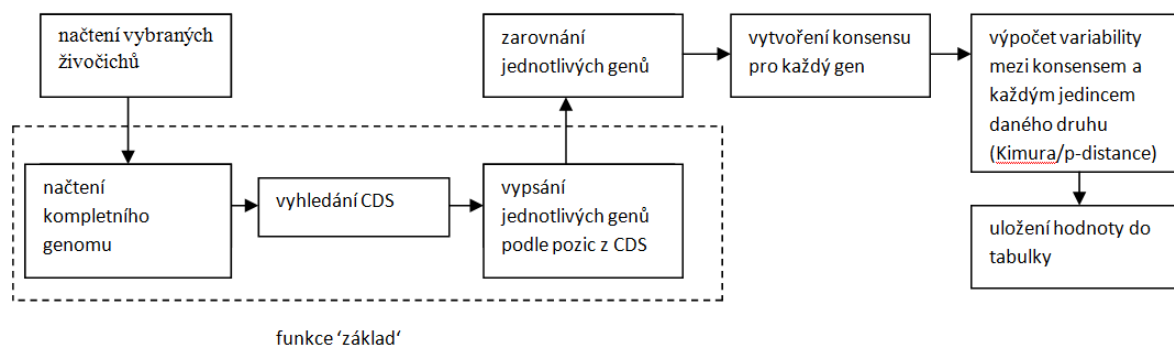
Tab. 4: Mezidruhová variability pro různé počty sekvencí

	3 sekvence	4 sekvence	10 sekvencí
Equus caballus(kůň domácí) Gallus gallus(kur bankivský)	0,228127	0,304027	0,304137
Canis lupus familiaris (pes domácí) Equus caballus(kůň domácí)	0,157248	0,209048	0,208748
Mus musculus(myš domácí) Sus scrofa(prase domácí)	0,242532	0,241641	0,241706
Bos grunniens(jak domácí) Bos taurus(tur domácí)	0,052652	0,0721	0,072356
Equus caballus(kůň domácí) Mammuthus primigenius(mamut srstnatý)	0,196756	0,261755	0,262356

5.3 Výpočet vnitrodruhové a mezidruhové variability

Mezi hlavní funkce pro zjištění podobnosti mezi daty patří výpočet vzdálenosti vybraných sekvencí, výpočet distancí mezi sekvencemi v načteném souboru dat a zobrazení fylogenetických stromů různými metodami. Vstupním souborem jsou soubory s příponou .gb (genbank), obsahující informace o druhu, kompletní mitochondriální genom a pozice jednotlivých genů. Pro načtené soubory dat pak můžeme využít funkce pro výpočet distancí všech sekvencí. Zde volíme buď výpočet vzdálenosti p-distancí nebo p-distance s korekcí dvouparametrickým Kimurovým evolučním modelem. Všechny sekvence jsou nejprve zarovnány pomocí funkce multialign a teprve pak je počítána distance.

Distance neboli vzdálenosti, které program vypočítá, slouží k porovnání podobnosti/rozdílnosti zadaných sekvencí. Jsou to vlastně sumy rozdílů jednotlivých částí sekvence a z toho vyplývá, že čím je číslo vyšší, tím jsou vybrané sekvence rozdílnější. Naopak nula znamená, že se jedná o naprosto identické sekvence. Vypočítané vzdálenosti pak uložíme do Excelu.



Obr. 9: Blokové schéma pro výpočet vnitrodruhové variability

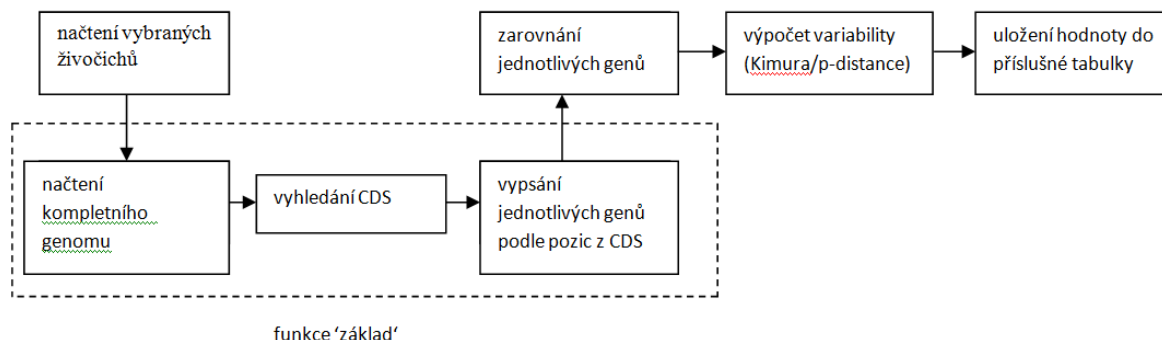
Tento skript s názvem `vnitrodruhova_variabilita.m` slouží k výpočtu vzdáleností v rámci druhu. Po jeho spuštění je potřeba ručně vybrat všechny zástupce jednoho druhu. Takže postupně vybereme čtveřici zástupců pro každý z třiceti druhů. Tato čtveřice byla vždy načtena a pomocí funkce `zaklad.m` rozdělena na jednotlivé geny včetně samotného barcodu, který je jen částí genu COI, poté byly všechny 4 sekvence každého genu vícenásobně zarovnány pomocí funkce `multialign`. Následně pak byl pro každý gen vytvořen konsensus. Konsensuální sekvence shrnuje informaci obsaženou ve všech získaných sekvencích. a v následujících několika cyklech dochází k výpočtu vzdáleností mezi konsensem a jednotlivými sekvencemi v rámci druhu. Výpočet probíhá buď pomocí funkce `seqpdist.m` z Bioinformatics toolboxu, která umožňuje výpočet vzdálenosti pomocí Kimura modelu,

```
d=seqpdist(Seq1,Seq2,'K2P');
```

nebo jednodušší způsob, kde po zarovnání všech sekvencí spočítáme rozdílná místa mezi konsensem a každou sekvencí nebo podíl rozdílných míst k celkové délce sekvencí tzv. proporcionální vzdálenost sekvencí p-distance. Výstupem v obou případech je hodnota vnitrodruhové variability, která je nakonec příkazem `xlswrite` exportována do Microsoft Excel.

Tab. 5: Ukázka hodnot vnitrodruhové variability

název	nad1	nad2	cox1	atp6	cox3	nad3	nad4	nad6	cob
<i>Bipes canaliculatus</i>	0,02242	0,02262	0,02051	0,02757	0,02264	0,02746	0,02804	0,03524	0,03201
<i>Bipes tridactylus</i>	0,0761	0,07788	0,06266	0,09044	0,0743	0,09754	0,07933	0,07056	0,07236
<i>Bos grunniens</i>	0,00105	0,00048	0,00000	0,0022	0,00064	0,00000	0,00018	0,00284	0,00088
<i>Branchiostoma belcheri</i>	0,00556	0,00744	0,00533	0,00878	0,00444	0,00353	0,0046	0,00694	0,00416
<i>Bubalus bubalis</i>	0,01071	0,02778	0,0343	0,02313	0,01216	0,01806	0,01613	0,22013	0,00636
<i>Camelus bactrianus</i>	0,00679	0,00551	0,00372	0,00514	0,00574	0,00643	0,00617	0,04719	0,00789
<i>Canis latrans</i>	0,00131	0,01125	0,00065	0,0011	0,00223	0,00000	0,00127	0,00519	0,00154
<i>Canis lupus familiaris</i>	0,00444	0,00263	0,00291	0,00624	0,00478	0,00288	0,00363	0,04364	0,00219



Obr. 10: Blokové schéma pro výpočet mezidruhové variability

Skript s názvem `mezidruhova_variabilita.m` naopak slouží k výpočtu vzdáleností mezi jednotlivými druhy. Po jeho spuštění je vybrán jeden druh, pro který se budou počítat vzdálenosti mezi ostatními druhy. Takže postupně vybereme čtveřici zástupců pro konkrétní druh a v dalším kroku vybereme čtveřice zástupců všech druhů v datasetu. Výpočet vzdálenosti probíhá v cyklu. V každé nové smyčce je vybrána nová čtveřice zástupců jednoho druhu. Do tohoto cyklu opět vstupuje funkce `zaklad.m`, která každou sekvenci rozdělí na jednotlivé geny. Jediná rozdílnost od výpočtu vnitrodruhové variability je ta, že se zde netvoří konsensus, ale počítáme vzdálenosti mezi každými dvěma sekvencemi. Výpočet probíhá buď opět pomocí funkce `seqdist.m`, která umožňuje výpočet vzdálenosti pomocí Kimura modelu, nebo spočítáme podíl rozdílných míst k celkové délce sekvencí tzv. p-distance. Výstupem v obou případech jsou hodnoty mezidruhové variability pro každý gen a pro každý druh vůči druhu zadanému na začátku, tyto hodnoty jsou nakonec opět příkazem `xlswrite` exportovány do Microsoft Excel do sešitů podle názvu genu. Výstupem celé analýzy je 14 tabulek pro 14 genů, včetně barcodu COI.

Zobrazení fylogenetického stromu z vypočtených distancí provedeme pomocí funkce `seqneighjoin`. Předtím ovšem musíme načíst potřebnou tabulku mezidruhové variability pro jednotlivé geny pomocí příkazu `xlsread`.

6. Výsledky

Tato analýza variability uvnitř a mezi jednotlivými druhy nepotvrzuje potenciál genu COI jako nejvhodnější sekvenci pro DNA barcoding, jelikož vnitrodruhové vzdálenosti některých druhů byly mnohem větší než stanovená hranice, zatím co mezidruhové vzdálenosti mezi některými druhy zase mnohem menší. Například mezidruhová vzdálenost mezi *Bos taurus* a *Bos grunniens* byla menší než 0,7 %. Tento výsledek je pravděpodobný nejspíš díky důsledku hybridizace, ovšem poukazuje na nespolehlivost COI. Tato selhání mitochondriálního DNA na vyšší taxonomické úrovni nejsou úplně neočekávaná, vzhledem k tomu, že pravděpodobnost oddělení těchto převážně savců spojuje relativně krátký časový rámec.

Tato studie naznačuje, že některý z 13 genů může být potencionálně použit jako molekulární diagnostika pro identifikaci druhů. Nicméně s ohledem na kritéria uvedená v normách a pokyny pro výběr podle CBOL, jako například kvalita, univerzálnost a pokrytí sekvence, by měl být první volbou COI čárový kód. Musíme, ale vzít v potaz to, že tyto normy byly napsány na základě samotného čárového kódu COI.

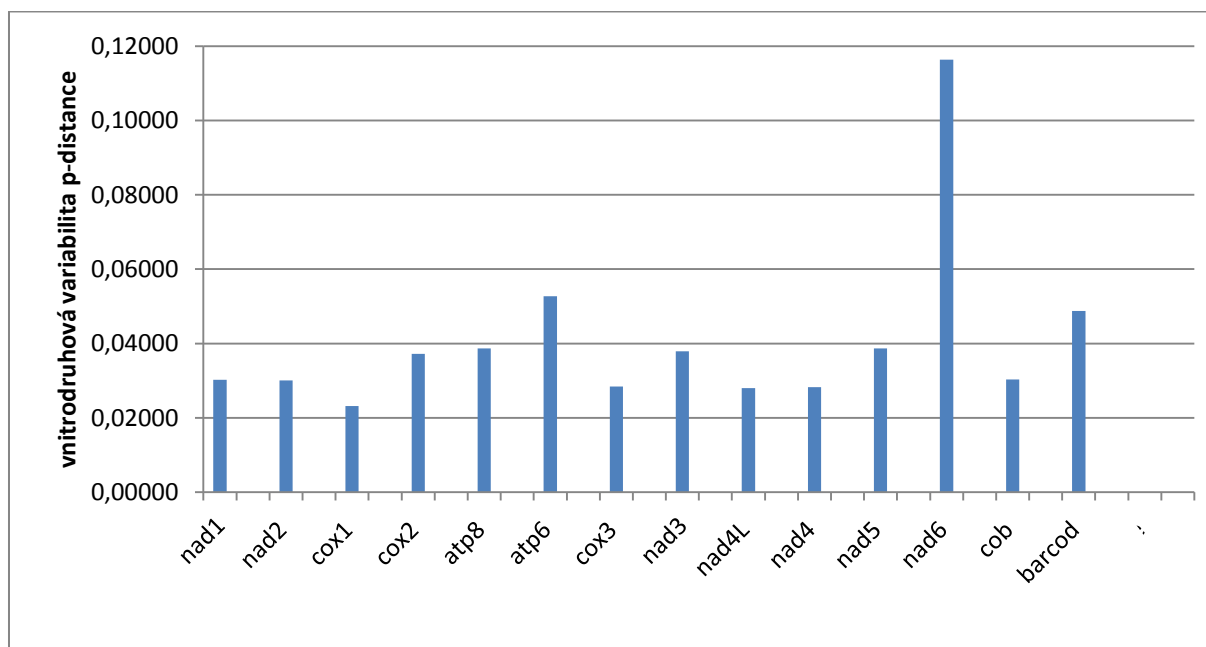
Tab. 6: Mezidruhové a vnitrodruhové variability

		vnitrodruhová variabilita podle Kimury		vnitrodruhová variabilita p- distance		mezidruhová variabilita p- distance		mezidruhová variabilita podle Kimury	
	d (bp)	x	ro	x	ro	x	ro	x	ro
nad1	959	0,018	0,038	0,030	0,069	0,367	0,121	0,411	0,153
nad2	1039	0,021	0,043	0,030	0,064	0,429	0,081	0,490	0,151
cox1	1536	0,017	0,036	0,023	0,050	0,265	0,048	0,292	0,077
cox2	682	0,014	0,026	0,037	0,103	0,316	0,062	0,358	0,111
atp8	162	0,016	0,028	0,039	0,106	0,481	0,118	0,430	0,143
atp6	680	0,030	0,073	0,053	0,124	0,398	0,084	0,438	0,140
cox3	784	0,018	0,039	0,028	0,067	0,301	0,058	0,339	0,100
nad3	346	0,021	0,041	0,038	0,092	0,384	0,079	0,423	0,136
nad4L	297	0,024	0,048	0,028	0,058	0,473	0,159	0,500	0,178
nad4	1373	0,018	0,037	0,028	0,060	0,414	0,096	0,464	0,165
nad5	1788	0,023	0,043	0,039	0,082	0,448	0,147	0,463	0,170
nad6	525	0,033	0,055	0,116	0,154	0,593	0,105	0,597	0,201
cob	1147	0,018	0,036	0,030	0,063	0,374	0,121	0,326	0,108
barcod	648	0,030	0,073	0,049	0,124	0,258	0,045	0,294	0,077
celá sekvence	16523	0,020	0,042					0,411	0,102

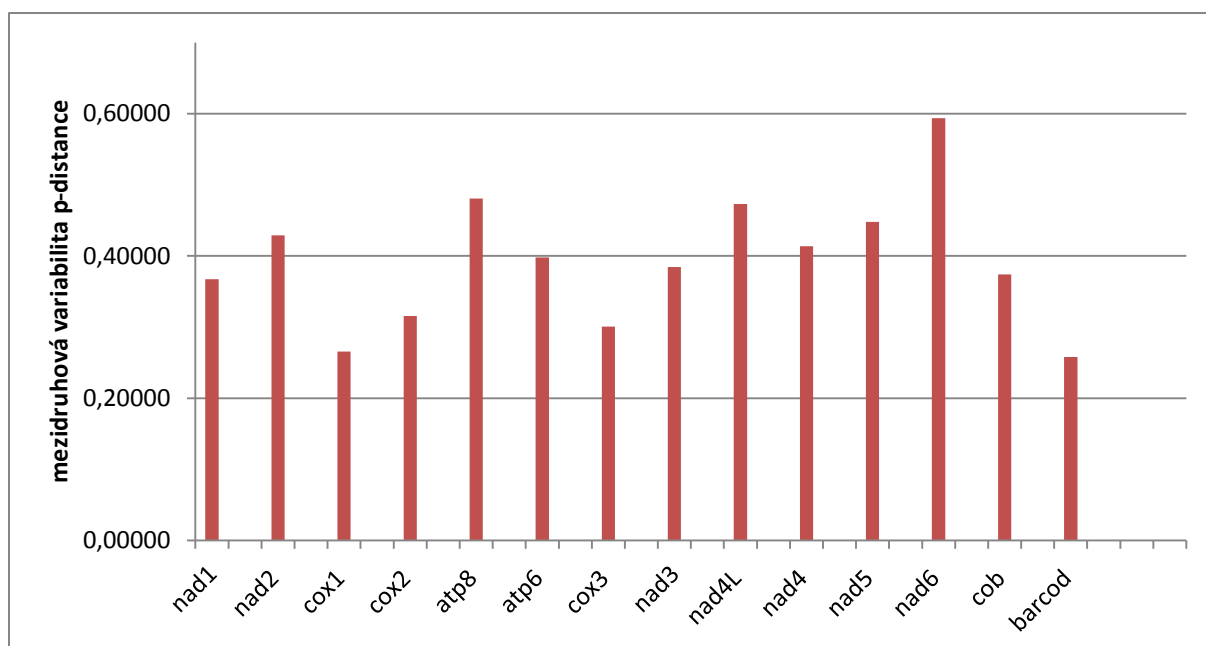
V tabulce č. 6 zkratky znamenají: d – délka sekvence, x – průměr, ro – směrodatná odchylka.

Mnoho sekvenčních dat z této oblasti jsou již k dispozici pro širokou škálu druhů, proto by mohla být COI použita spolu s jinými geny. Výsledky totiž potvrzují fakt, že DNA barcoding stále čelí problému překrývání mezi vnitrodruhovou a mezidruhovou variabilitou. Tam by, ale mohl nastat problém, protože každý gen funguje jako biologická entita s jinou

evoluční rychlostí. DNA barcoding nemusí být omezen na mitochondriální geny. S rostoucí dostupností sekvencí z jaderných genů, je velmi pravděpodobné, že některé jaderné regiony by mohly být také účinné jako DNA čárový kód a měly by být v budoucnu i testovány.

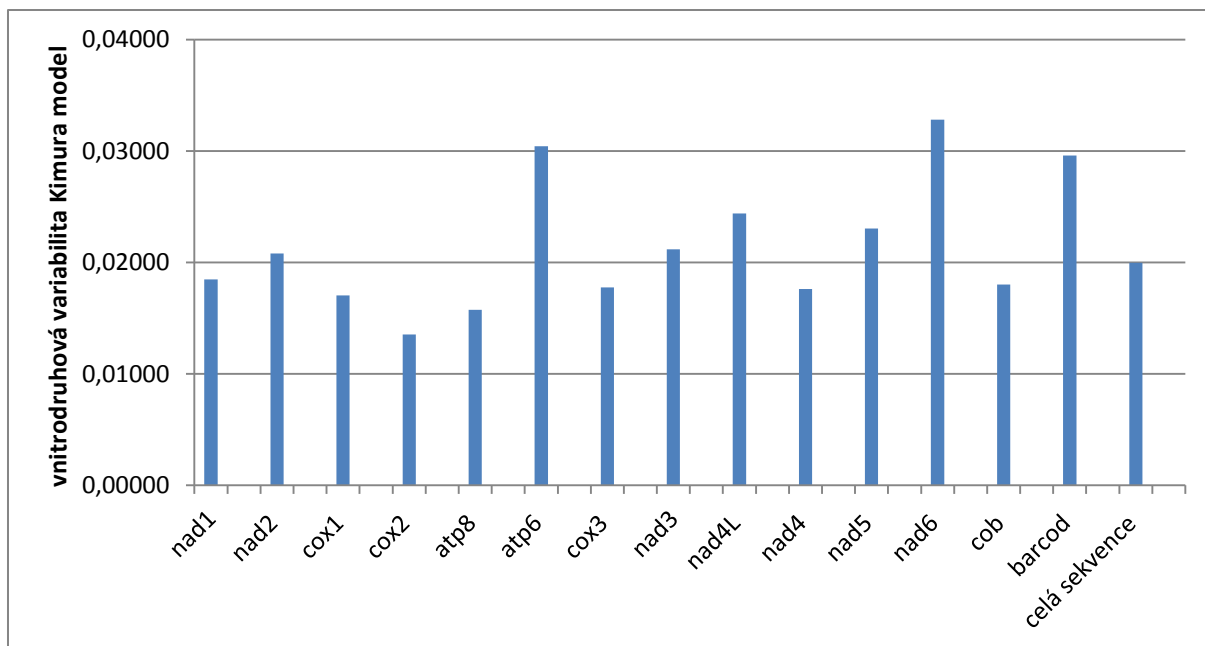


Obr. 11: Průměrná vnitrodruhová variabilita pro 13 genů, barcod a celou sekvenci výpočtem p-distance

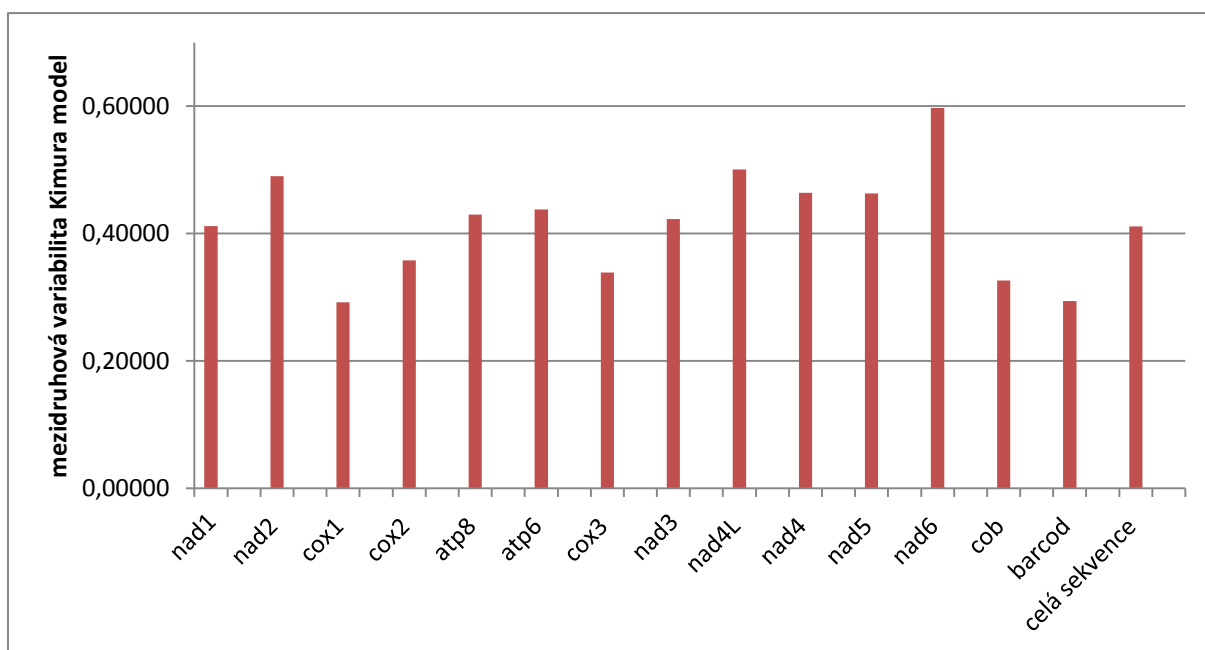


Obr. 12: Průměrná mezidruhová variabilita pro 13 genů, barcod a celou sekvenci výpočtem p-distance

Aby byl DNA barcoding účinný, mezidruhové rozdíly by měly být jasné a výrazné, větší než vnitrodruhové rozdíly. Tato analýza, potenciál nynějšího barcodu, ani nepotvrdila, ani nevyvrátila. V rámci vnitrodruhových variabilit patří do skupin genů s nejnižší hodnotou, ovšem u mezidruhových je tomu také tak a my požadujeme mezidruhovou variabilitu co největší, aby nedocházelo k překrývání těchto hodnot a špatné identifikaci. Vnitrodruhové vzdálenosti některých druhů byly mnohem větší, než je stanovená průměrná vnitrodruhová vzdálenost, zatímco mezidruhové vzdálenosti mezi některými druhy byly mnohem menší, než je stanovená střední mezidruhová vzdálenost. Příklady zahrnují *Bos grunniens*, *Bos taurus*, *Hippoglossus hippoglossus*, *Hippoglossus stenolepis* a dokonce i *Mammuthus primigenius* a *Elephas maximus*, pro které mezidruhové vzdálenosti byly velmi nízké, kolem 3%, ve srovnání s ostatními studiemi DNA barcodingu. Studie ukázala, že mezi těmito podobnými druhy není žádný významný rozdíl mezi intra - a mezidruhovou vzdáleností napříč všemi 14 zkoumanými oblastmi, což je pravděpodobné, nejspíše v důsledku hybridizace.



Obr. 13: Průměrná vnitrodruhová vzdálenost pro 13 genů, barcod a celou sekvenci výpočtem podle Kimura modelu



Obr. 14: Průměrná mezidruhová vzdálenost pro 13 genů, barcod a celou sekvenci výpočtem podle Kimura modelu

Jako nejvhodnější podle grafu č. 2 by se zdál gen NAD6. Ovšem tento gen, je jen kolem 500 bp dlouhý a protože leží na komplementním vlákně a obsahuje mnoho inzercí a

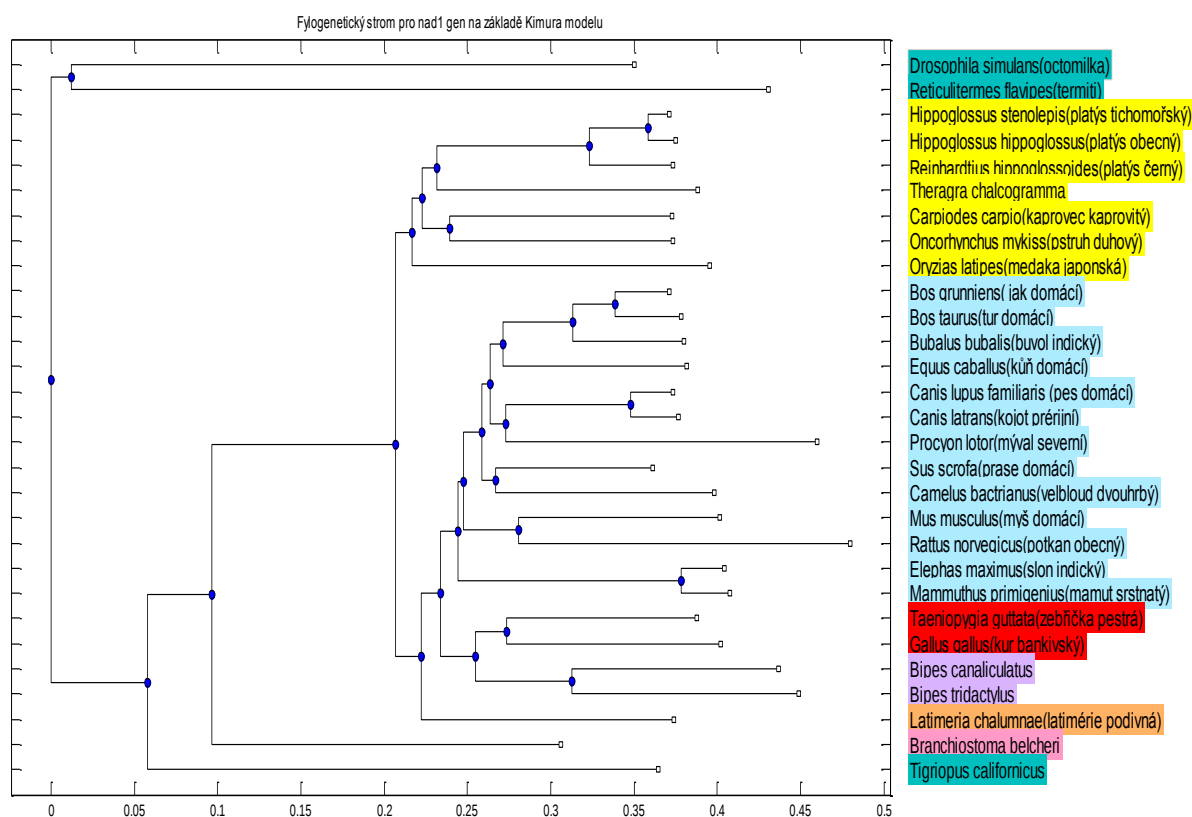
delecí, je nespolehlivý. To můžeme pozorovat i v tabulce č. 6, kde směrodatné odchylky zrovna u tohoto genu jsou určitě nejvyšší, a proto bych ho jako potencionální gen pro DNA barcoding nenominovala.

Další nečekanou věcí, která z výsledku vyplývá, je, že dosud uznávaný barcod a gen COI jako celek, mají nejmenší mezidruhovou variabilitu ze všech studovaných regionů.

Testovaný dataset obsahuje druhy vyhynulé i zástupce z různých zeměpisných oblastí, proto by tyto výsledky měli být objektivní a neovlivňuje je například faktor, jako je zeměpisné zařazení.

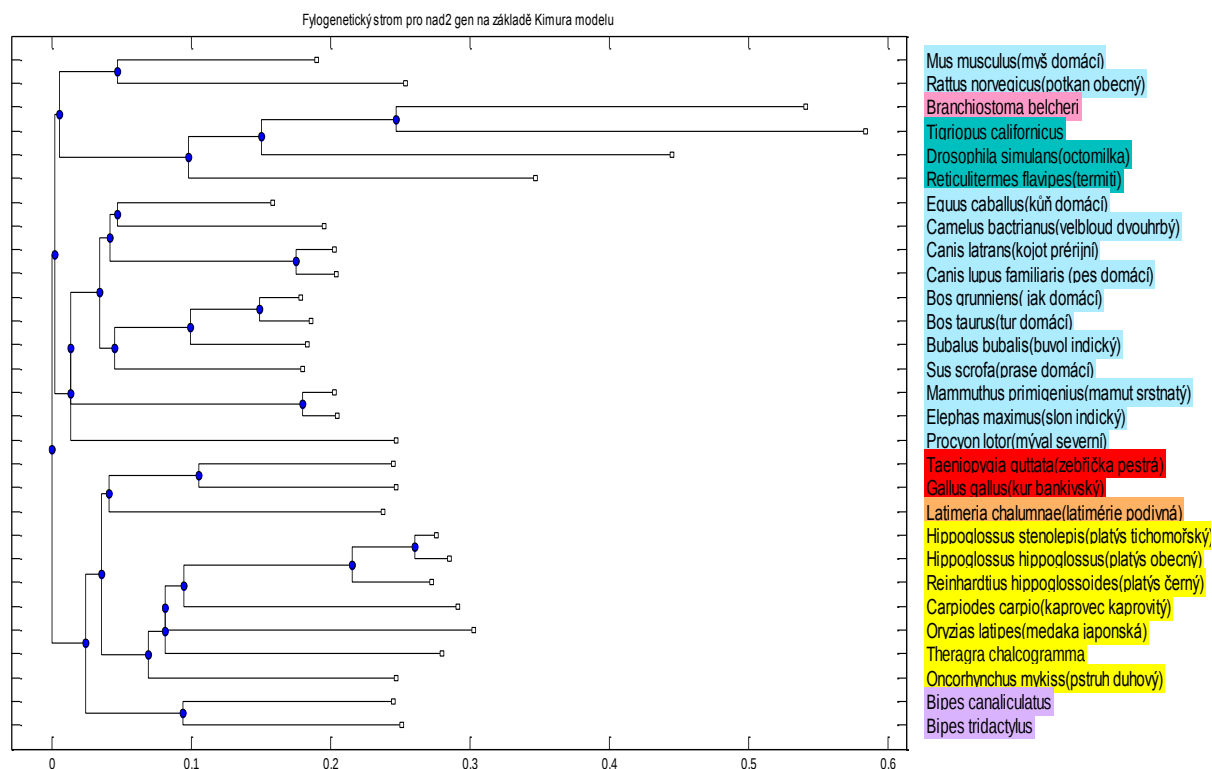
6.1 Dendrogramy pro jednotlivé zkoumané regiony

Dendrogram je druh diagramu používaný ke znázornění jednotlivých kroků shlukové analýzy. Při výpočtu se každý prvek nachází samostatně na svislé ose. Horizontální osa vyjadřuje vzdálenost mezi jednotlivými shluky. Shluky se sjednocují podle nejkratší vzdálenosti.



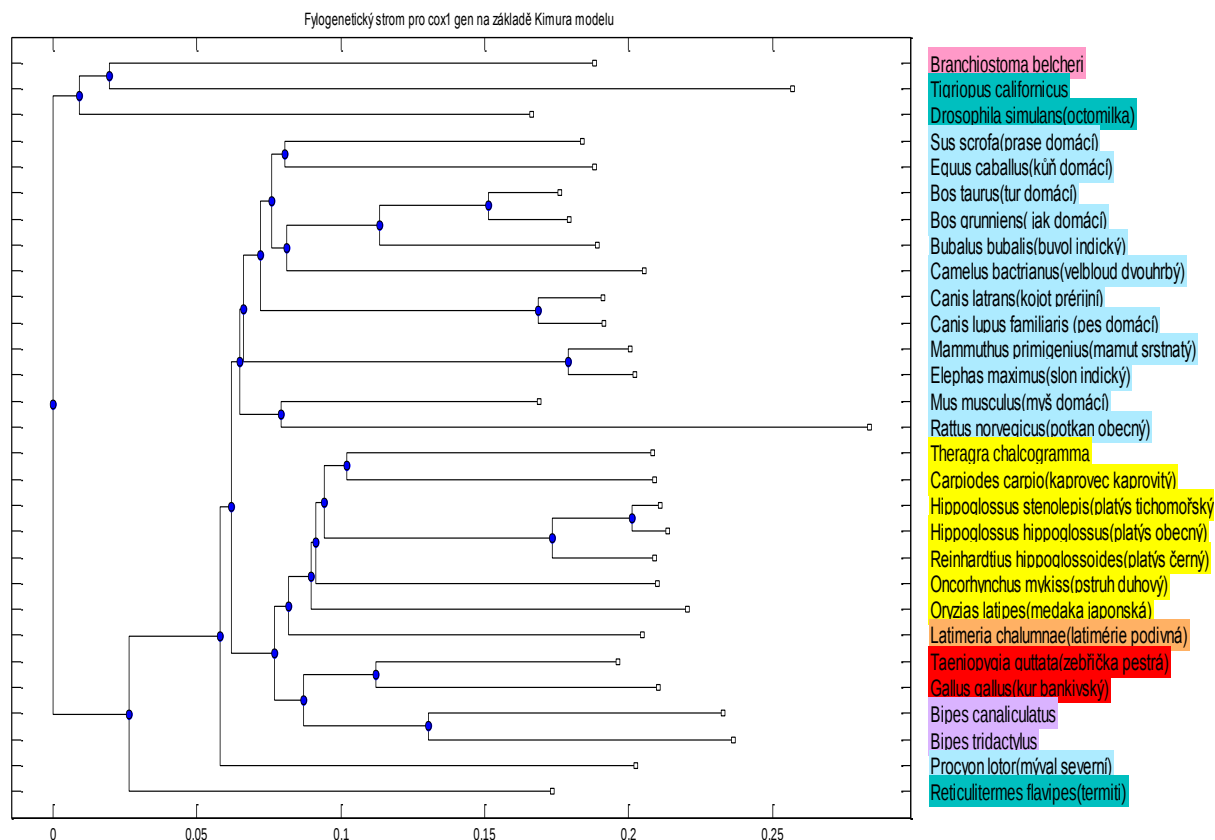
Obr. 15: Dendrogram mitochondriálního genu NAD1

Gen NAD1 kóduje protein pro NADH dehydrogenázu podjednotky 1. Jeho velikost se pohybuje kolem 950 bp. Jeho vnitrodruhová variabilita je v průměru 0,01848 a mezidruhová variabilita 0,41149, což jsou s ohledem na ostatní geny velice dobré výsledky. K dobrým výsledkům lze přičíst i to, že v dendrogramu jsou podobné druhy přiřazeny k sobě. Jsou to ty, které mají stejné rodové jméno a liší se jen jménem druhovým. Zde vidíme takto přiřazené například druhy *Bipes*, *Bos*, *Canis* či *Hippoglossus*. Zařazení každého druhu v dendrogramu proběhlo správně. Savci se nacházejí na stejné vývojové větvi, stejně tak paprskoploutví. Latimérie, o které se říká, že je živoucí fosílií a vědci ji dlouhou dobu měli za vyhynulou, je také zařazena správně, což u většiny genů dělalo potíže. Stejně tak *Drosophila simulans*, *Reticulitermes flavipes* a *Tigriopus californicus*, kteří jsou jediní zástupci kmenu Arthropoda, se nacházejí na stejné vývojové větvi.



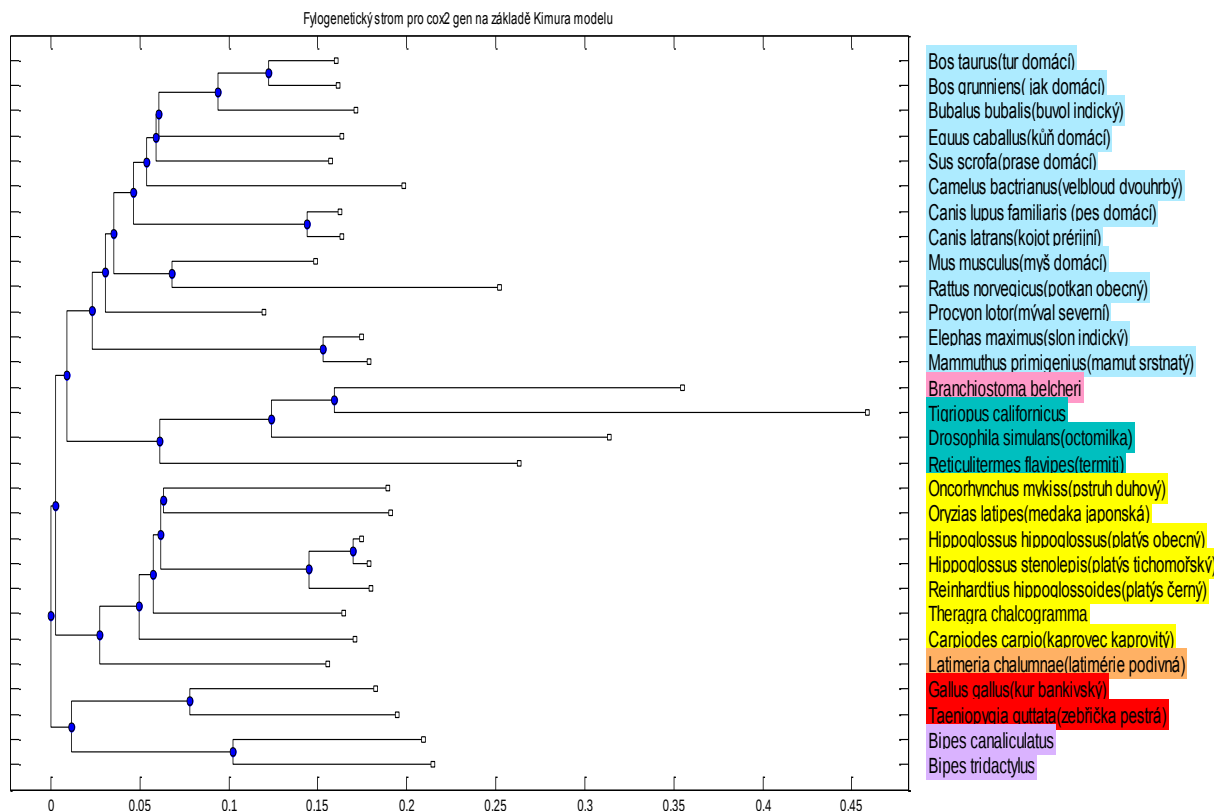
Obr. 16: Dendrogram genu NAD2

Gen NAD2 kódující protein pro NADH dehydrogenázu podjednotky 2, s délkou sekvence kolem 1040 bp, má oproti NAD1, vnitrodruhovou variabilitu vyšší, v průměru 0,02080, a i mezidruhová variabilita se o něco zvýšila, v průměru 0,49. Ovšem, jak je vidět z obr. 12, dendrogram úplně neodpovídá skutečným vztahům. K dobrým výsledkům lze přičíst jen to, že v dendrogramu jsou podobné druhy přiřazeny k sobě. Špatně jsou však zařazeny *Mus musculus* a *Rattus norvegicus*, kteří se nacházejí blíže k zástupcům kmene *Arthropoda*. V rozporu se skutečností je i *Latimeria chalumnae*, která leží na jedné vývojové větvi s ptáky než s paprskoploutvými, jak by se dalo předpokládat.



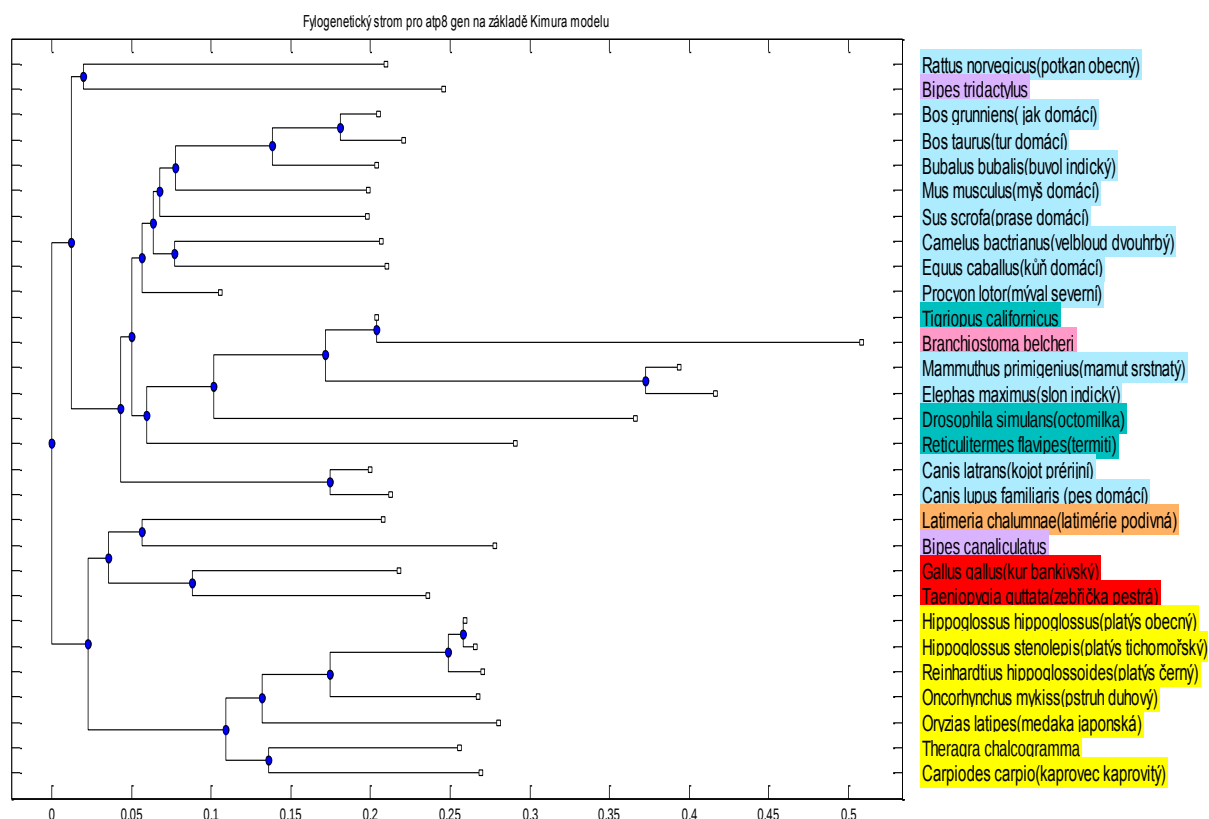
Obr. 17: Dendrogram genu COX1

Dále je tu gen COI kódující cytochromoxidázu c podjednotky 1. Tento gen je asi 1540 bp dlouhý a jeho analýze byla věnována největší pozornost, protože v jeho struktuře se nachází nyní používaný čárový kód pro DNA barcoding. Hodnoty mezidruhové a vnitrodruhové variability tohoto genu byly však zklamáním, protože se ukázal jako gen s nejmenší mezidruhovou variabilitou, a to v průměru kolem 0,292. Vnitrodruhová variabilita se sice pohybuje mezi nižšími hodnotami, asi 0,01704, ale podobně nízkou variabilitu má většina genů, někteří dokonce i nižší. Struktura výsledného dendrogramu pro tento gen je správná, až na zařazení *Procyon lotor* a *Reticulitermes flavipes*, který nebyl přiřazen k zástupcům kmene *Arthropoda*. Ostatní druhy jsou v dendrogramu rozmístěny správně.



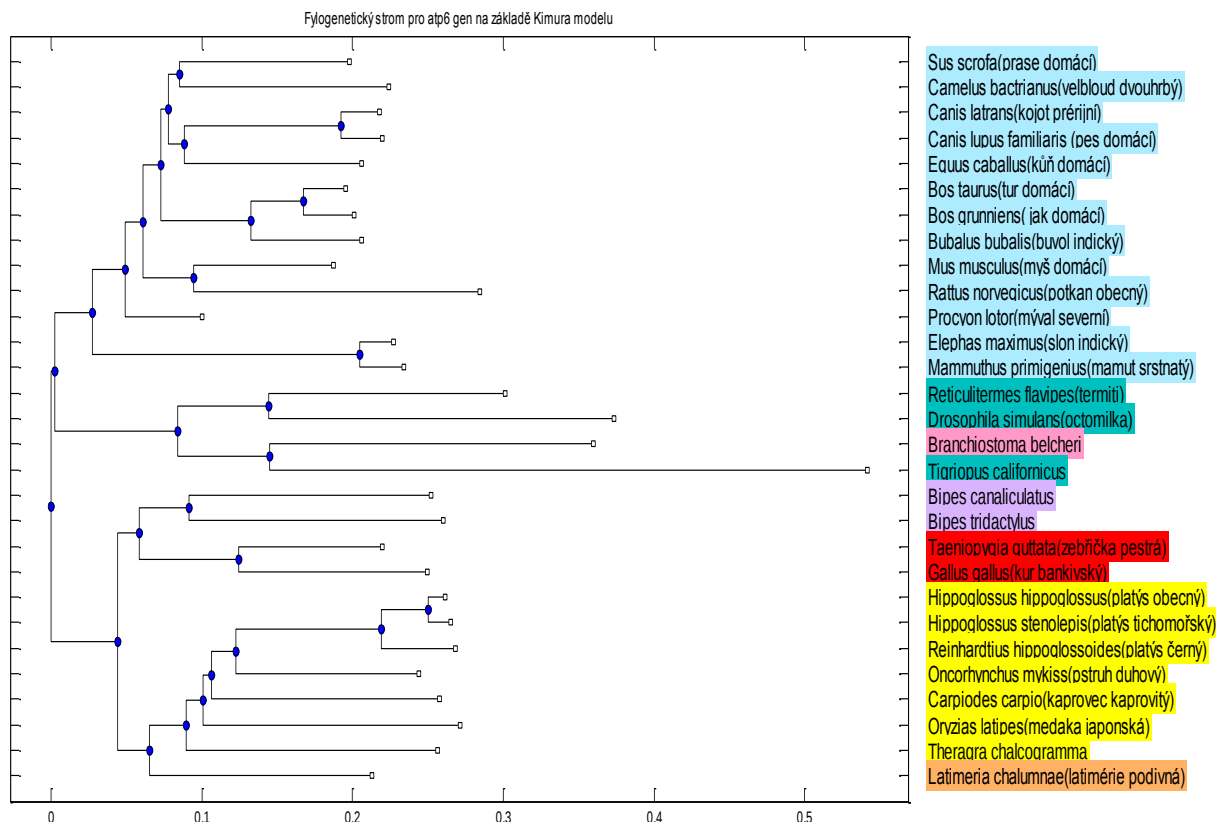
Obr. 18: Dendrogram genu COX2

Gen COX2 kóduje protein pro cytochromoxidázu c podjednotky 2. Jeho velikost se pohybuje kolem 680 bp, což ho řadí mezi kratší geny. S ohledem na ostatní geny je jeho vnitrodruhová variabilita nejmenší, v průměru 0,01354 a mezidruhová variabilita je 0,35786, což je vzhledem k COI mnohem větší hodnota. Až na některé malé výjimky, k sobě dendrogram správně přiřadil jednotlivé druhy. Jediná chyba, která se v dendrogramu nachází, je zařazení *Branchiostoma belcheri*. Tento druh patří do kmene *Chordata* a v dendrogramu byl přiřazen do větve s *Arthropodou*. Tato skutečnost, že se *Branchiostoma* řadí do tohoto kmene, se vyskytuje u všech genů, proto bych to nepovažovala za chybu.



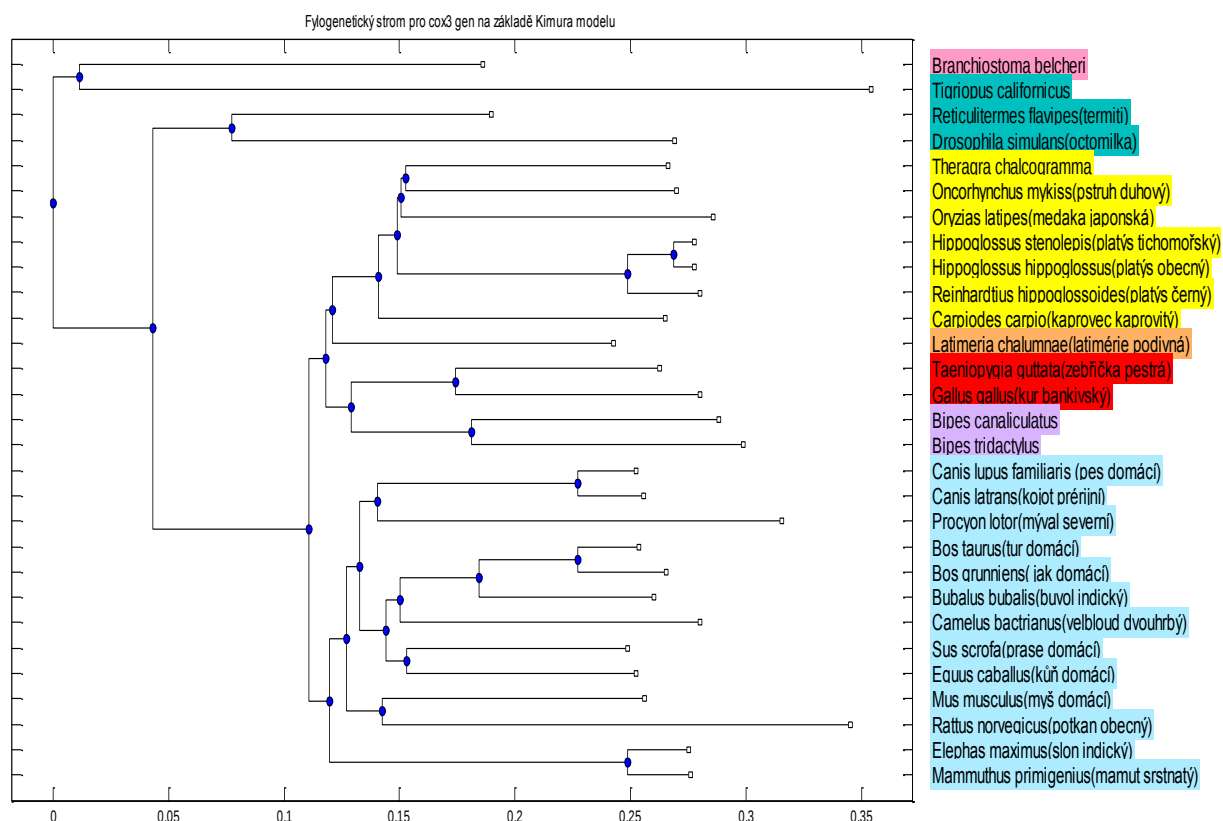
Obr. 19: Dendrogram genu ATP8

ATP8 je gen kódující protein ATP syntázu podjednotky 8. Tento gen je nejkratší, s délkou sekvence kolem 160 bp. Vnitrodruhová variabilita je v průměru 0,0157 a mezidruhová variabilita je v průměru 0,42992. Ovšem, jak je vidět z obr. 15, dendrogram vůbec neodpovídá skutečnostem. V dendrogramu nejsou ani podobné druhy se stejným druhovým jménem přiřazeny k sobě. Špatně jsou zařazení například mamut a slon, kteří se nacházejí na jedné vývojové větvi se zástupci kmene *Arthropoda*. Savci jsou rozděleni do tří oddílů a *Bipes*, což je zástupce plazů, je mylně zařazen mezi savce, jako je pes a potkan. V rozporu se skutečností je zařazena i *Latimeria chalumnae*, která se nachází blíže k plazům než k paprskoploutvým, jak by se dalo předpokládat. Tento gen se zdá nespolehlivý, důvodem je nejspíš jeho krátká sekvence.



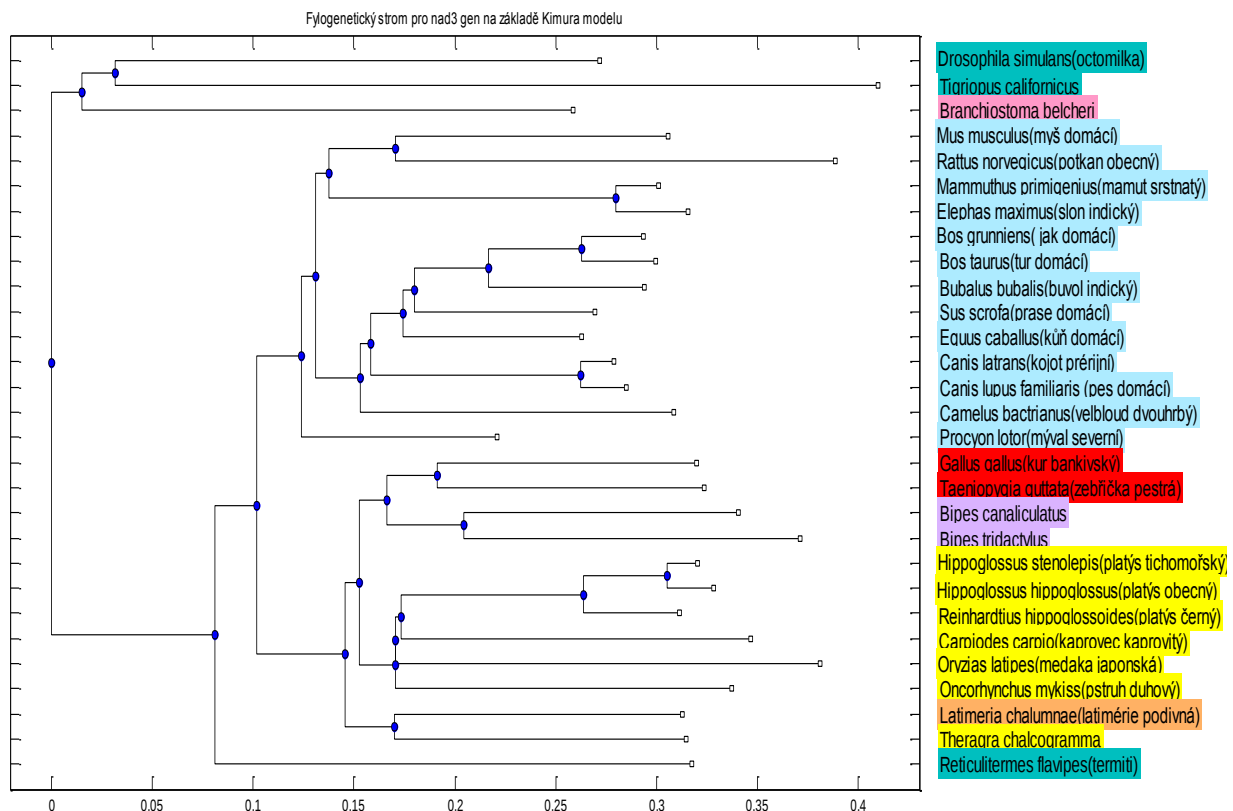
Obr. 20: Dendrogram genu ATP6

Gen ATP6, kódující protein ATP syntázu podjednotky 8, patří také mezi kratší geny s délkou sekvenec kolem 680 bp. Jeho průměrná vnitrodruhová variabilita je druhá největší ze všech genů (0,03042), ale i mezidruhová variabilita patří do vyšších hodnot (0,43803). Dendrogram je velice podobný dendrogramu genu COX2, který, jak už bylo řečeno, je správný. Jediná odchylka je v trochu jiném poskládání savců. Vezmeme-li v úvahu délku sekvenec, výsledky variabilit a výsledný strom, mohl by tento gen být nominován také pro použití DNA barcodingu.



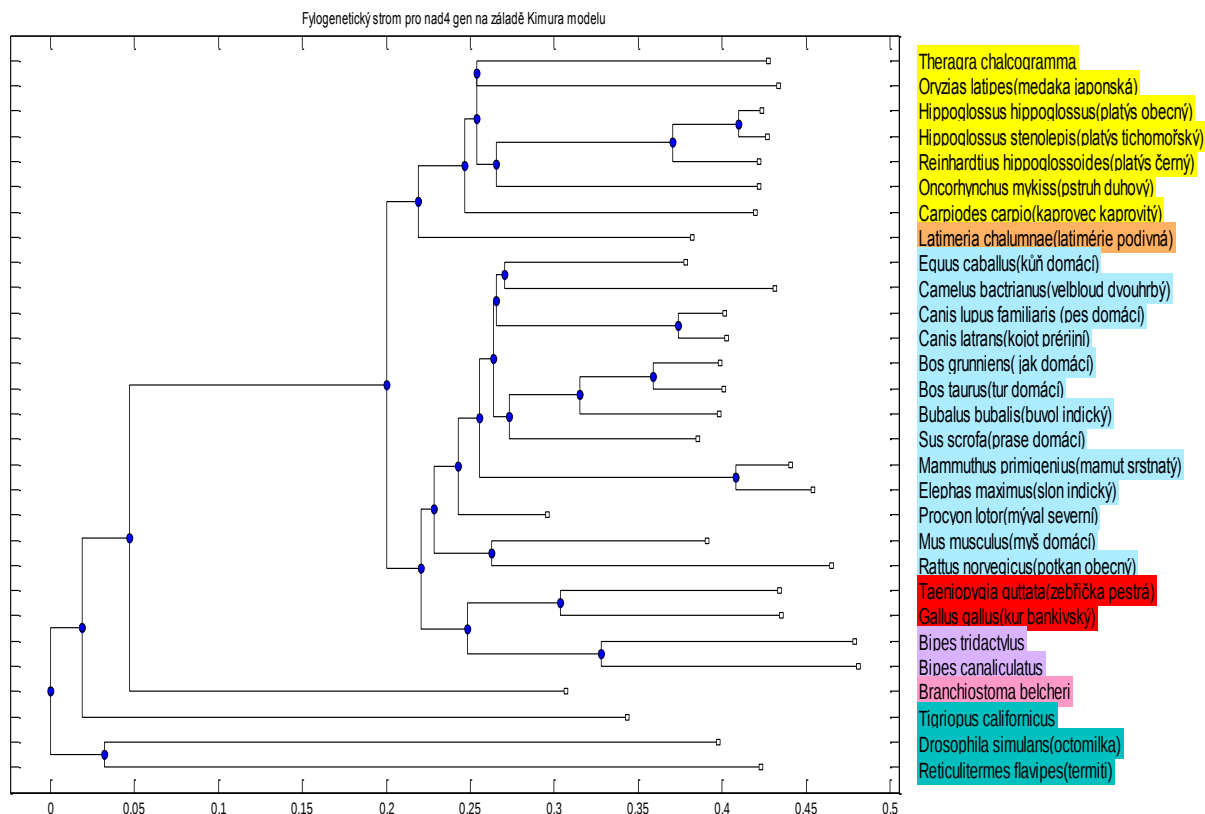
Obr. 21: Dendrogram genu COX3

Další ze zástupců genů, kteří kódují cytochromoxidázu c, je gen COX3. Délka sekvence se pohybuje kolem 784 bp a průměrná vnitrodruhová a mezidruhová variabilita je podobná jako u COX1 a COX2. Je to 0,01775 a 0,33887. Jeho dendrogram je velice podobný genu COX2. Například u savců bych řekla, že je zarovnání více podobné skutečnosti, protože je to jediný strom, který přiřadil *Procyon lotor* do stejného shluku jako *Canis lupus* a *Canis latrans*. Všechny tři druhy jsou zástupci šelem. Jeho další výhodou oproti genu COX2 je jeho délka, která je větší.



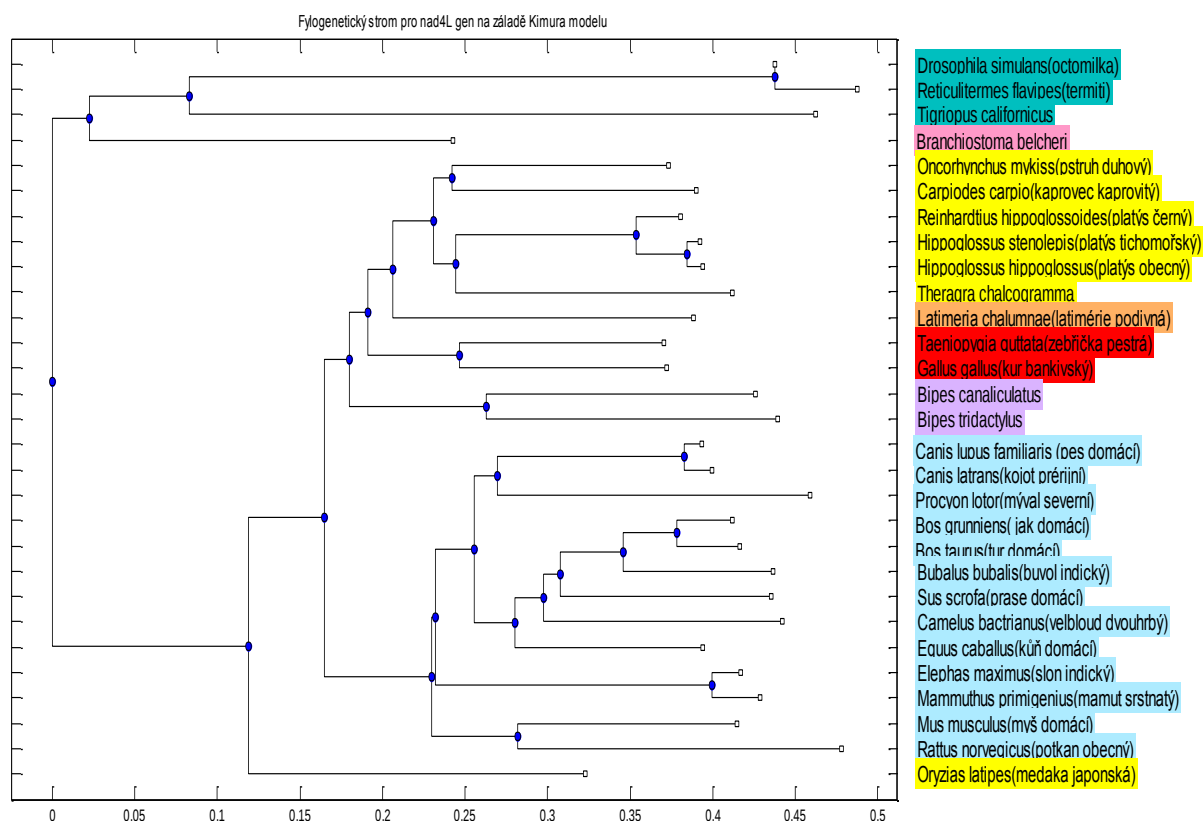
Obr. 22: Dendrogram genu NAD3

Gen NAD3 kódující protein pro NADH dehydrogenázu podjednotky 3 s délkou sekvence kolem 350 bp, má oproti NAD1 a NAD2 nejkratší sekvenci, ale vnitrodruhovou variabilitu vyšší, v průměru 0,02116, a mezidruhovou variabilitu zase o něco nižší, v průměru 0,42287, což je nejhorší možná kombinace. Jak je vidět z obr. 18, dendrogram úplně neodpovídá skutečným. K dobrým výsledkům lze přičíst jen to, že v dendrogramu jsou podobné druhy přiřazeny k sobě. Špatně jsou však zařazeni zástupci hmyzu *Drosophila simulans* a *Reticulitermes flavipes*, kteří se nacházejí jinde. Další chybou je i pozice *Latimeria chalumnae*. Gen patří mezi ty kratší a špatně zařazuje druhy v dendrogramu, proto není vhodný k použití metody DNA barcodingu.



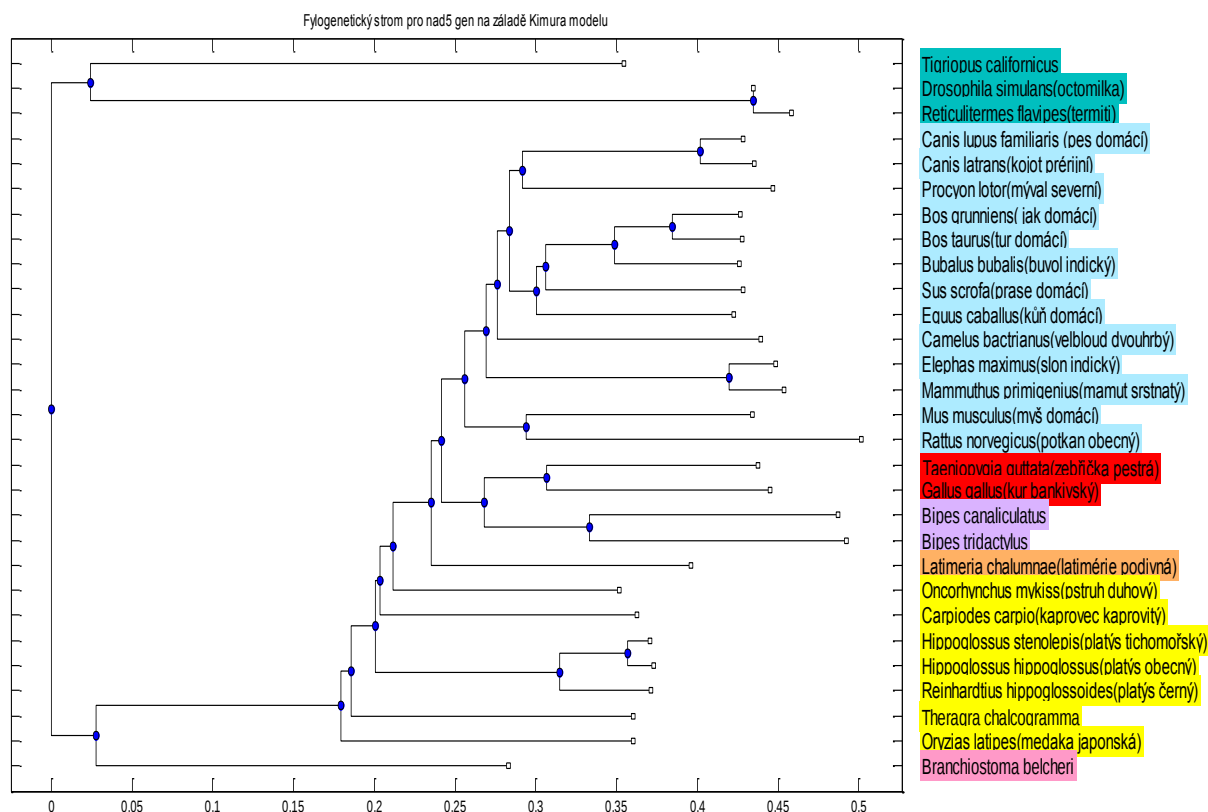
Obr. 23: Dendrogram genu NAD4

NAD4 gen kóduje protein pro NADH dehydrogenázu podjednotky 4 s délkou sekvence kolem 1370 bp, což je třetí nejdelší gen v kompletním genomu. Vnitrodruhová variabilita patří k těm nižším, průměrně 0,01761, a mezidruhová variabilita k těm vyšším, v průměru 0,46417. To je naopak oproti genu NAD3 nejlepší možná kombinace. Gen je dostatečně dlouhý a jeho variability tvoří jednu z největších „barcoding gap“, což je rozdíl mezi variabilitami. Požadujeme ho co největší, aby nedocházelo k překrývání a nesprávné identifikaci. Další dobrou vlastností, která se prokázala ve výsledku, je, že ze všech genů má nejmenší směrodatné odchylky ve všech případech, jak je vidět v tabulce č. 6. Na obr. 19 je vidět dendrogram tohoto genu. Už na první pohled je jasné bezchybné zařazení, podobně jako u ATP6 a COX2, ovšem tento gen je delší, a proto nabízí více možností. Mohl by to být jeden z dalších uchazečů na zástupce DNA barcodingu.



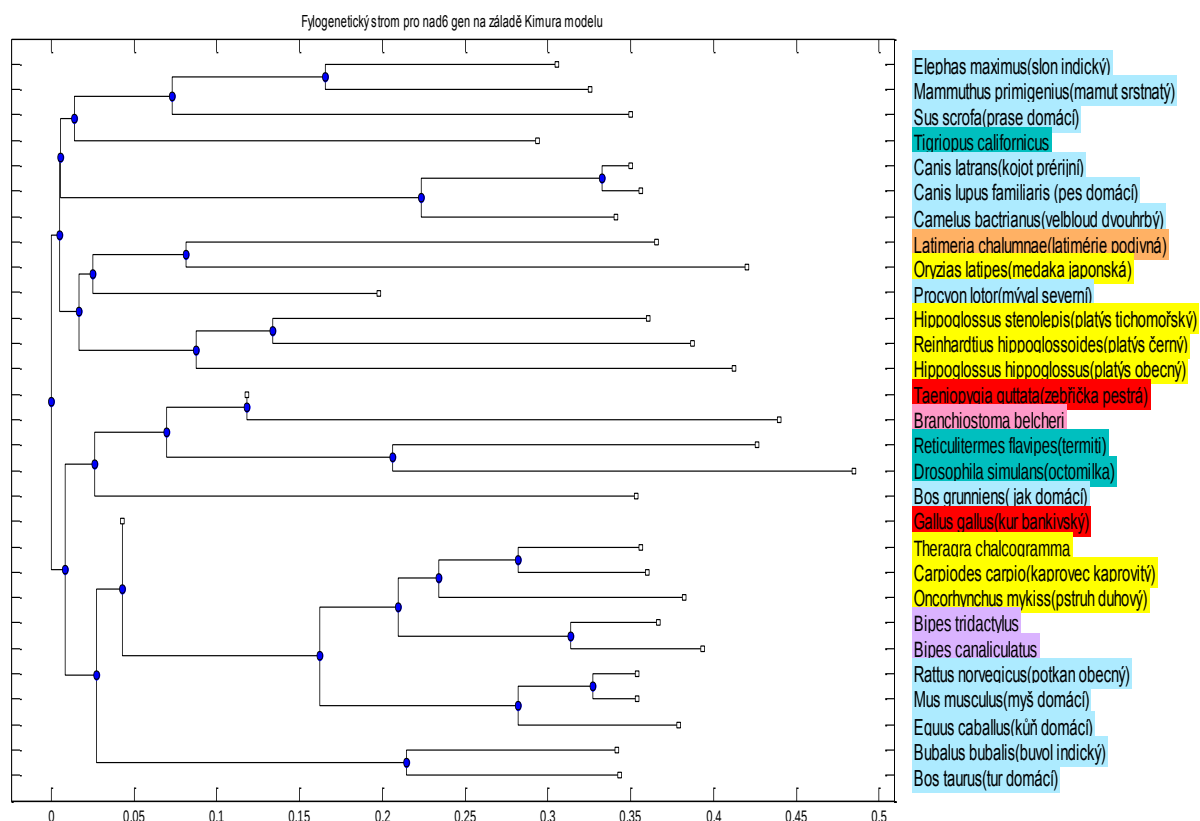
Obr. 24: Dendrogram genu NAD4L

NAD4L gen kóduje protein pro NADH dehydrogenázu podjednotky 4L s délkou sekvence kolem 290 bp, což je druhý nejkratší gen v kompletním genomu. Vnitrodruhová variabilita je průměrně 0,02439 a mezidruhová variabilita je v průměru 0,50049. Kromě genu NAD6 je to největší hodnota mezidruhové variability. Tento gen bych však nedoporučovala brát, jako vhodný pro molekulární taxonomii, a to hlavně z důvodu jeho krátké struktury a jak můžeme vidět i dendrogram neodpovídá řádnému zařazení druhů. Například *Oryzias latipes*, což je zástupce paprskoploutvých, byl mylně zařazen k savcům.



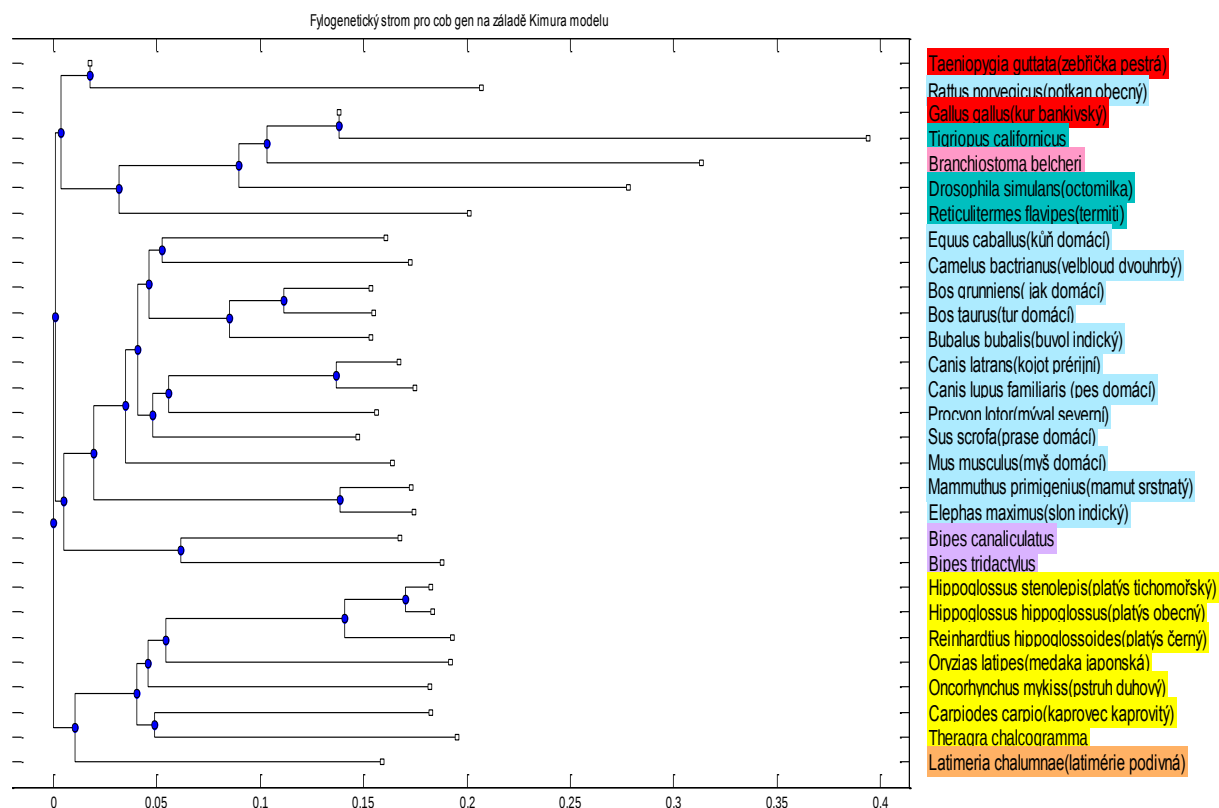
Obr. 25: Dendrogram genu NAD5

Gen NAD5 zabírá v kompletním mitochondriálním genomu největší část, jeho délka se pohybuje okolo 1780 bp. Tento gen kóduje protein NADH dehydrogenázu podjednotky 5 a jeho vnitrodruhová (0,02303) a mezidruhová (0,46294) není nijak zvláštní. Jsou to průměrné hodnoty. V dendrogramu se nepodařilo správně rozdělit druhy, které se všechny nachází na jedné vývojové větvi a už na první pohled je vidět, že tento gen je nespolehlivý pro DNA barcoding.



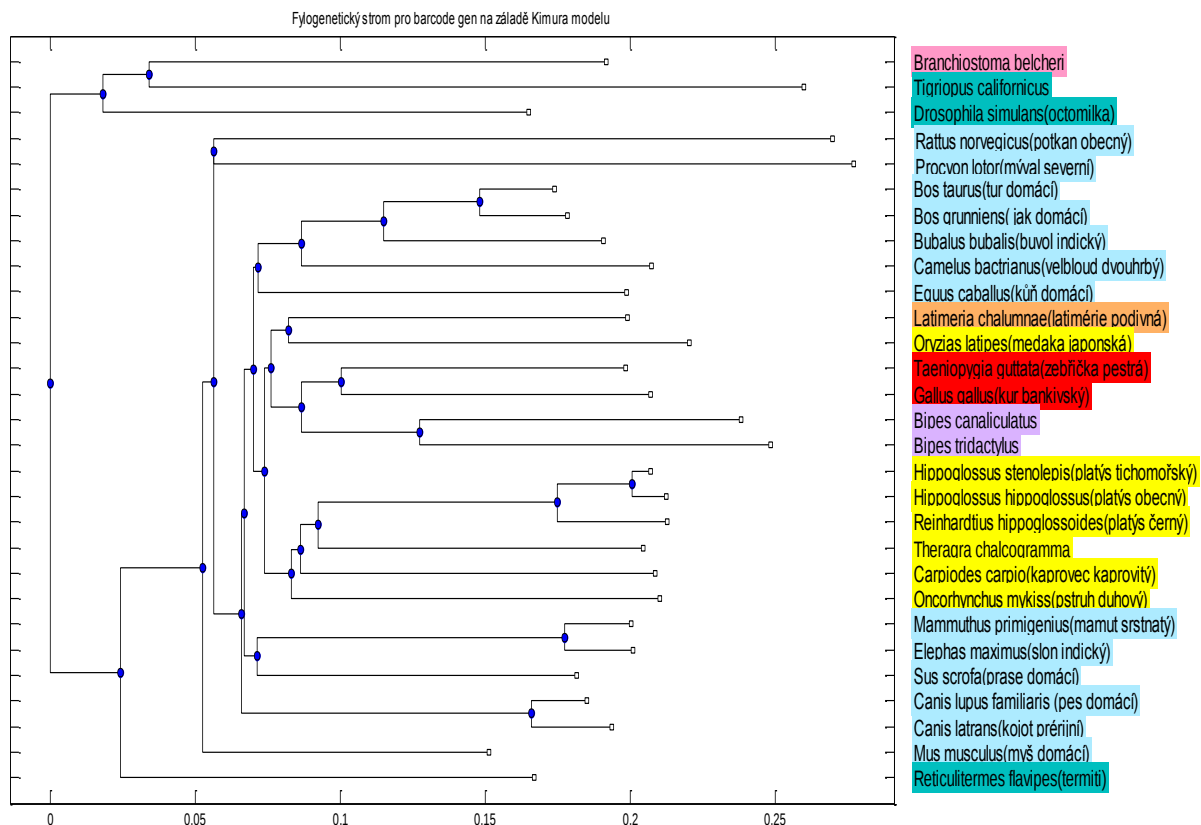
Obr. 26: Dendrogram genu NAD6

Gen NAD6 kóduje protein NADH dehydrogenázu podjednotky 6. Tento gen se v kompletní mitochondriální sekvenci nachází na komplementním vlákně, je kolem 500 bp dlouhý a protože obsahuje mnoho inzercí a delecí, je nespolehlivý. To můžeme pozorovat i v tabulce č. 6, kde směrodatné odchylky u tohoto genu jsou určitě nejvyšší, a proto bych ho jako potenciální gen pro DNA barcoding nedoporučovala. Jeho mezidruhová variabilita je ovšem největší (0,59746), to ale také platí o vnitrodruhové variabilitě (0,03280). Jeho dendrogram netvoří žádné srozumitelné shluky. Dokonce k sobě nejsou přiřazeni ani druhy se stejným druhovým jménem.



Obr. 27: Dendrogram genu COB

Gen COB nebo také CYBT kóduje protein cytochrom b. Jeho velikost je kolem 1147 bp délky sekvenční. Průměrná vnitrodruhová variabilita byla vypočtena na 0,018, což je vzhledem k ostatním genům průměr. Mezidruhová variabilita je však po genu COX1 a barcodu nejnižší (0,32603). Délka sekvenční je dostačující, splňuje podmínku 10 krát větší mezidruhové variability, ovšem dendrogram nevyšel podle očekávání. Mylně jsou zařazeni ptáci, kteří se nacházejí mezi zástupci kmene *Arthropoda*, stejně jako *Rattus norvegicus*.



Obr. 28: Dendrogram genu BARCODE

Nakonec zmiňovaný barcod, což je část genu COX1. Jeho délka je 648 bp. Jeho vnitrodruhová variabilita patřila k těm vyšším, kolem 0,02959 a naopak mezidruhová variabilita (0,29416) byla po celém genu COX1 jedna z nejmenších. Nesplňuje tedy podmínku, že mezidruhová vzdálenost by měla být 10 krát vyšší než vnitrodruhová, což bylo velkým překvapením. Jak je vidět na obr. 24, jeho dendrogram patří k těm horším. Jeho struktura, až na pár výjimek, neodpovídá skutečnosti. Nejvíce zasaženi jsou savci, kteří jsou rozděleni do 3 skupin, kmen *Arthropoda* neleží také na jedné větvi.

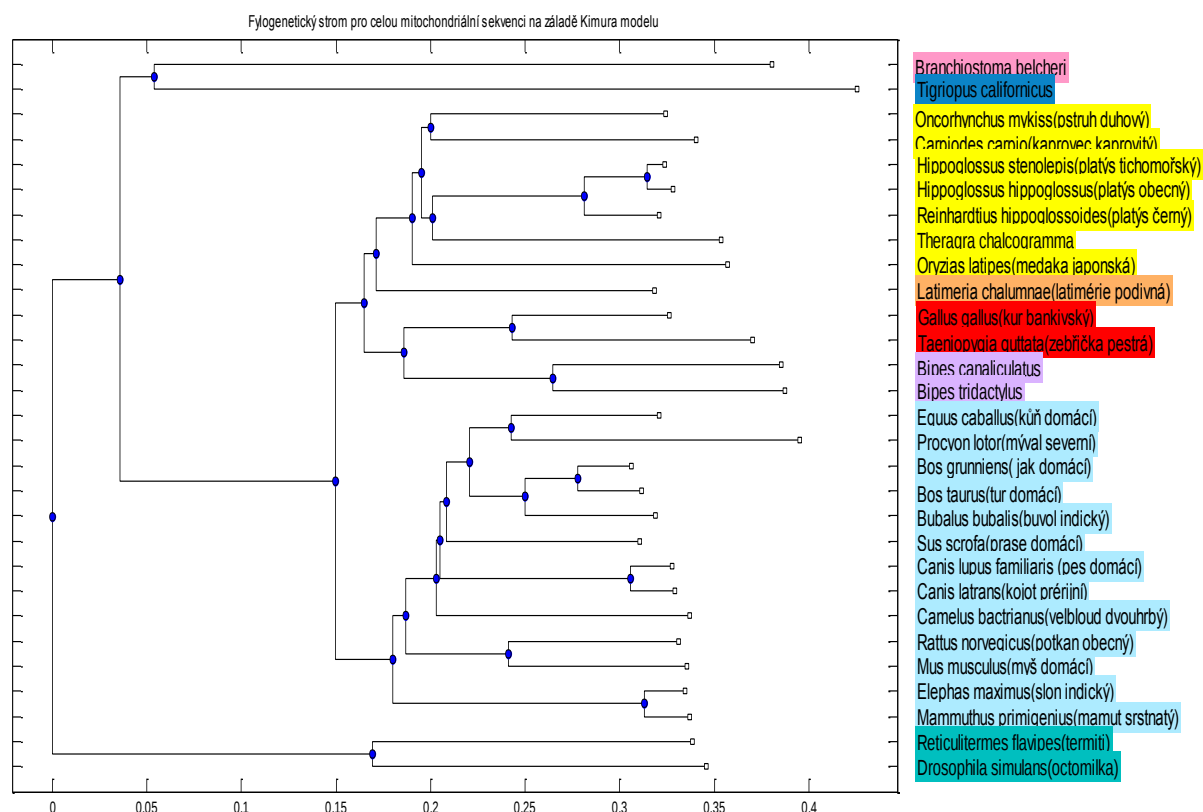
Tab. 7: Správně určené dendogramy pro jednotlivé regiony

nad1	nad2	cox1	cox2	atp8	atp6	cox3	nad3	nad4L	nad4	nad5	nad6	cob	barcod	celá sekvence
1	x	x	1	x	1	1	x	x	x	x	x	x	x	1
959	1039	1536	682	162	680	784	346	297	1373	1788	525	1147	648	16523

Proměnné: 1-správně, x-špatně, průměrná délka genů [bp]

6.2 Analýza celého mitochondriálního genomu

DNA barcoding by měl umožnit rychle, spolehlivě, automatizovaně a levně identifikovat druhy uživatelem bez žádných taxonomických zkušeností. Použití celé sekvence, jako nástroj pro identifikaci druhů by sice ještě do nedávna nesplňovalo podmínku rychlosti, ovšem, co se týče spolehlivosti, byl by na prvních místech před všemi 13 geny. Na druhou stranu by nebylo potřeba řešit získání jen jednoho genu z této celé sekvence, což by rychlost automatizované identifikace, jak je plánováno, zvýšilo. Nebylo by třeba použití specifických primerů k získání jednotlivých genů. Také by se nemuseli zavádět nové databáze speciálně pro DNA barcoding, protože dosavadní databáze obsahují nesčetné množství kompletních mitochondriálních genomů. S nástupem nové generace sekvenátorů není získávání celých mitochondriálních sekvencí ničím limitováno a DNA barcoding by se měl zaměřit právě na porovnávání celého mitochondriálního genomu, než jen jednoho velmi omezeného úseku.



Obr. 29: Dendrogram pro celé mitochondriální genomy

Celý mitochondriální genom vícebuněčných živočichů je dlouhý průměrně 16500 pb, což je skoro 10 krát víc než délka největšího mitochondriálního genu, proto výpočty vzdáleností mezi sekvencemi trvaly déle. Navíc se vyskytl problém se zarovnáním těchto dlouhých sekvencí. Protože funkce, která byla použita na vícenásobné zarovnání jednotlivých

relativně krátkých genů (*multialing*), nemohla být použita pro celé sekvence. Místo toho byly celé mitochondriální sekvence zarovnávány párově funkcí *nwaling*, což je funkce pro globální zarovnávání Needlemanovým-Wunchovým algoritmem. Všechny možné dvojice celých mt sekvencí byly globálně zarovnány, ze zarovnání se spočítaly p-distance s korekcí 2-parametrickým Kimurovým evolučním modelem. Dá se říct, že konečná hodnot mezidruhové variability je průměrem hodnot vzdáleností mezi každými dvěma sekvencemi jiného druhu. Díky tomuto problému nebyla vnitrodruhová variabilita počítána pomocí konsensu, ale stejně, jak bylo popsáno.

Průměrná hodnota vnitrodruhové variability pro kompletní mitochondriální sekvenci je 0,01995, což je menší hodnota než u většiny genů a mezidruhová variabilita byla průměrně 0,41113. Směrodatné odchylky (0,04213 a 0,10243) jsou, s ohledem na ostatní výsledky, jedny z nejmenších. Z výsledků vyplývá, že hodnota mezidruhové variability je víc jak 20 krát vyšší než hodnota vnitrodruhové.

Dalším plusem pro analýzu kompletního mitochondriálního genomu je vytvořený dendrogram na obr. 25. Strom je podobný jako u genů ATP6, COX2 a NAD4, tedy správný a nejvíce podobný skutečné příbuznosti organismů, jež byla určena na základě morfologických, behaviorálních a environmentálních parametrů.

Závěr

Cílem této práce na téma „Analýza mitochondriálních genů živočichů pro DNA barcoding“ bylo vypracování rešerše o mitochondriálním genomu a o taxonomické metodě DNA barcodingu, sestavení datasetu mitochondriálních sekvencí, vytvoření funkce pro extrakci jednotlivých genů, výpočet mezidruhových a vnitrodruhových variabilit a zhodnocení účinnosti těchto genů.

Od svého vzniku v roce 2003 se přístup DNA barcodingu vyvinul v široce používanou metodu pro identifikaci druhů a hraje klíčovou roli v mnoha mezinárodních projektech. V současnosti převládá názor, že DNA barcoding je třeba používat nikoli samostatně, ale ve spolupráci s dalšími taxonomickými metodami, aby se předešlo mylným identifikacím a závěrům. Všechny tyto pro a proti jsou v této práci rozebírány. Doposud DNA barcoding pracuje na základě jednoho genu obsaženého v mitochondriálním genomu. Moje další práce bude spočívat v určení vnitrodruhových a mezidruhových variabilit pro všechny geny v genomu a na základě výsledků zhodnotit, zda nejsou vhodné pro tuto metodu i další geny či jejich části.

Byl sestaven dataset sekvencí stažených na stránkách NCBI jako formát genbank, obsahující 32 živočišných druhů a 1037 jejich kompletních mitochondriálních sekvencí. Tato databáze sice není stoprocentně spolehlivá, ovšem stejně na tom jsou všechny jiné databáze, kde je vědcům umožněno vkládat molekulární informace. Pro formát genbank musela být vytvořena funkce na extrakci jednotlivých genů i celé sekvence, protože *genbankread* byl ve většině případů nespolehlivý a nenačetl všechny uvedené informace z důvodu nestandardizovaného zápisu. Nejsou dána žádná striktní pravidla pro zápis těchto informací, a když ano, tak zřejmě nejsou dodržována.

Výpočet variabilit proběhl buď pomocí Kimura modelu nebo byl spočítán podíl rozdílných míst k celkové délce sekvencí tzv. p-distance. Hodnoty vnitrodruhových a mezidruhových variabilit jsou uvedeny v tabulce č. 6, kde se kromě hodnot vzdáleností nachází i směrodatné odchylky a průměrná délka jednotlivých regionů.

Následně byly, pro lepší pochopení a zorientování se ve výsledcích, sestrojeny dendogramy pro každý zkoumaný region. Každý tento region je podrobně popsán a v závěru je uvedena tabulka č. 7, která hodnotí použitelnost všech 13 mitochondriálních genů, včetně barcodu a celého genomu.

Jako vhodné geny pro DNA barcoding, bereme-li v úvahu hodnoty vypočtených variabilit a dodržení pravidla, že mezidruhová variabilita musí být minimálně desetkrát větší

než vnitrodruhová, dále pak délku zkoumaného regionu a správně vypočítaný dendogram, bych doporučila k dalšímu zkoumání geny NAD1, COX2, ATP6, NAD4 a hlavně celý mitochondriální genom.

V závěru je nutné říci, že tyto výsledky neznamenaají, že COX1 nemůže být použit pro identifikaci druhů, ale dokazují, že pro stanovení mnoho taxonů budeme muset spoléhat nejen na molekulární metodu DNA barcodingu, ale použít i jiné.

Literatura

[1] SCHEFFLER, Immo E. Mitochondria. 2nd ed., Wiley-Blackwell, 2007, 472 s. ISBN 978-0-470-04073-7.

[2] BLAXTER, Mark. The promise of a DNA taxonomy. Phil. Trans. R. Soc. Lond. B., 2004, vol. 359, pp. 669-679.

[3] ALBERTS, Bruce, et al. Základy buněčné biologie: Úvod do molekulární biologie. 2. vyd. Translation Prof. RNDr. Arnošt Kotyk, DrSc. Ústí nad Labem: Espero publishing, 1998. ISBN 80-902906-2-0.

[4] EFENBERK, Aleš. MIMOJADERNÁ DĚDIČNOST U ČLOVĚKA. Brno, 2008. 31 s. MASARYKOVA UNIVERZITA; Přírodovědecká fakulta; Ústav experimentální biologie; Oddělení genetiky a molekulární biologie. Vedoucí bakalářské práce prof. RNDr. Jiřina Relichová, CSc.

[5] VŠCHT PRAHA, Ústav organické technologie. Mitochondriální dědičnost. [online]. [cit. 12. prosince 2013]. Dostupné na WWW: <<http://www.vscht.cz/kot/resources/studijni-materialy/bc-skripta/kapitola04.pdf>>

[6] SNUSTAD, P.D. and SIMMONS, M.J. Principles of Genetics. Wiley, 5. vyd., 2008, 848s. ISBN: 978-0470388259

[7] PIERCE, B.A. Genetics – A Conceptual Approach. W. H. Freeman; Third Edition edition 2007, 832s, ISBN: 0716779285

[8] WEAVER, R.F. and HEDRICK, P.W. Genetics. Third Edition. 1997, Wm. C. Brown Publishers, 638s. ISBN:0071148752

[9] MUDr. Antonín ŠÍPEK. *Genetika-Biologie* [online]. 2011, [cit.2013-12-12] Dostupné z: < <http://www.genetika-biologie.cz/mutace>>.

[10] Prof. RNDr. Jaroslav FLÉGR, CSc. Molekulární taxonomie [online], [cit.2013-12-12] Dostupné z: < web.natur.cuni.cz/~flegr/dokumenty/moltaxanotace.do>.

[11] Barcode of life : National Center for Biotechnology Information, US National Library of Medicine [online]. [cit. 2013-12-23]. Text v angličtině. Dostupný z WWW: <<http://www.ncbi.nlm.nih.gov/Genbank/barcode.html>>.

- [12] HERBERT, Paul. D. N., STOECKLE, Mark Y. Čárový kód života. *Scientific American*. 2008, č.lisopad 2008, s. 24-29. české vydání.
- [13] HAJIBABAEI, M.; SMITH, M. A.; JANZEN, D. H.; RODRIGUEZ, J. J.; WHITFIELD, J. B.; HEBERT, P. D. N. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 6: 959-964
- [14] JOHNSON, N. K.; CICERO, C. 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution* 58 (5): 1122–30
- [15] MENDELSON, T. C. & SHAW, K. L. 2005. Rapid speciation in an arthropod: The likely force behind an explosion of new Hawaiian cricket species revealed. *Nature* 433:375–376.
- [16] RUBINOFF, D, 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology* [online]. 10.3., vol. 20, no. 4, pp. 1026–1033 [accessed. 24. March 2014]. ISSN 08888892. Retrieved z: doi:10.1111/j.1523-1739.2006.00372.x
- [17] RATNASINGHAM, S. & HEBERT, P. D. N. BOLD: The Barcode of Life Data system. [online]. Barcode of Life Secretariat c/o National Museum of Natural History. [cit. 2013-12-27]. Dostupné ve formátu PDF na. <<http://www.boldsystems.org/libhtml/docs/bold.pdf>>.
- [18] HEBERT PDN, STOECKLE MY, ZEMLAK TS, FRANCIS CM (2004) Identification of Birds through DNA Barcodes. *PLoS Biol* 2(10): e312. doi:10.1371/journal.pbio.0020312
Dostupné online: <http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0020312>
- [19] HEBERT, P. D. N.; CYWINSKA, A.; BALL, S. L.; deWAARD, J. R. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270: 313-321
- [20] Consortium for the barcode of Life [online]. Washington [USA] : CBOL, May 2004 [cit. 20013-12-23]. Text v angličtině. Dostupný z WWW: <<http://www.barcoding.si.edu/>>.
- [21] CASIRAGHI, Maurizio, Massimo LABRA, Emanuele FERRI, Andrea GALIMBERTI and Fabrizio DE MATTIA, 2010. DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in bioinformatics* [online]. 7., vol. 11, no. 4, pp. 440–453 [accessed. 28. March 2014]. ISSN 1477-4054. Retrieved z: doi:10.1093/bib/bbq003

- [22] LUO, Arong, Aibing ZHANG, Simon Yw HO, Weijun XU, Yanzhou ZHANG, Weifeng SHI, Stephen L CAMERON and Chaodong ZHU, 2011a. Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC genomics* [online]. B.m.: BioMed Central Ltd, 1., vol. 12, no. 1, p. 84 [accessed. 28. March 2014]. ISSN 1471-2164. Retrieved z: doi:10.1186/1471-2164-12-84
- [23] LIU, Jun, Qi LI and Lingfeng KONG, 2011. Identifying the true oysters (Bivalvia: Ostreidae) with mitochondrial phylogeny and distance-based DNA barcoding [online]. vol. 11, no. 5, pp. 820–830 [accessed. 23. April 2014]. Retrieved z: http://apps.isiknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=Q2xAStVTgRtrtQwfzqs&page=1&doc=9
- [24] CHAPPLE, David G and Peter A RITCHIE, 2013b. A retrospective approach to testing the DNA barcoding method. *PloS one* [online]. 1., vol. 8, no. 11, p. e77882 [accessed. 28. March 2014]. ISSN 1932-6203. Retrieved z: doi:10.1371/journal.pone.0077882
- [25] GOLDSTEIN, Paul Z and Rob DESALLE, 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *BioEssays : news and reviews in molecular, cellular and developmental biology* [online]. 3., vol. 33, no. 2, pp. 135–47 [accessed. 31. March 2014]. ISSN 1521-1878. Retrieved z: doi:10.1002/bies.201000036
- [26] DESALLE, Rob, Mary G EGAN and Mark SIDDALL, 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* [online]. 29.10., vol. 360, no. 1462, pp. 1905–16 [accessed. 19. March 2014]. ISSN 0962-8436. Retrieved z: doi:10.1098/rstb.2005.1722
- [27] BLAXTER, Mark, Jenna MANN, Tom CHAPMAN, Fran THOMAS, Claire WHITTON, Robin FLOYD and Eyualet ABEBE, 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* [online]. 29.10., vol. 360, no. 1462, pp. 1935–43 [accessed. 19. March 2014]. ISSN 0962-8436. Retrieved z: doi:10.1098/rstb.2005.1725

[28] BERGM, Tjard, Heike HADRYŠ and Gerhard BREVES, 2009. Character-based DNA barcoding : a superior tool for species classification [online]. vol. 12, pp. 446–450. Retrieved z: doi:10.2376/0005-9366-122-446

[29] Vysoké učení technické v Brně [online]. Odpovědnost prof.Ing.Ivo Provazník, Ph.D. © 2013 [cit. 2014-05-20]. Analýza biologických sekvencí. Dostupné z: <<https://www.vutbr.cz/elearning/mod/resource/view.php?id=253766>>.

Seznam zkratek

DNA – deoxyribonukleová kyselina

mtDNA – mitochondriální deoxyribonukleová kyselina

RNA – ribonukleová kyselina

tRNA – transferová ribonukleová kyselina

ATP – adenosintrifosfát

ADP – adenosindifosfát

CBOL – The Consortium for the Barcode of Life

NCBI – National Center for Biotechnology Information

BOLD – University of Guelph's Barcode of Life Database

COI - cytochrom oxydáza

EMBL - Evropská molekulárně biologická laboratoř

PCR - Polymerase Chain Reaction

kb – kilobáze

bp – párů bází

AK – aminokyselina

NK – nukleová kyselina

DDBJ – The DNA Data Bank of Japan

NIH - National Institutes of Health

NAD1 – gen NADH dehydrogenázy podjednotky 1

NAD2 – gen NADH dehydrogenázy podjednotky 2

NAD3 – gen NADH dehydrogenázy podjednotky 3

NAD4 – gen NADH dehydrogenázy podjednotky 4

NAD4L – gen NADH dehydrogenázy podjednotky 4L

NAD5 – gen NADH dehydrogenázy podjednotky 5

NAD6 – gen NADH dehydrogenázy podjednotky 6

ATP8 – gen ATP syntázy podjednotky 8

ATP6 – gen ATP syntázy podjednotky 6

COX1, COI – gen cytochromoxidázy podjednotky 1

COX2 – gen cytochromoxidázy podjednotky 2

COX3 – gen cytochromoxidázy podjednotky 3

COB, CYTB- gen cytochrom b

K2P – dvoupřímý Kimura model

Seznam příloh

- CD s elektronickou verzí diplomové práce, kompletními vytvořenými programy a funkcemi, tabulky vnitrodruhových a mezidruhových vzdáleností pro všechny geny, barcod a celý mitochondriální genom, dendrogramy