



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

**PREFERENCE SYNONYMNÍCH KODONŮ V
MITOCHONDRIÁLNÍM GENOMU**

SYNONYMOUS CODON USAGE BIAS IN MITOCHONDRIAL GENOMES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Hedvika Urbanová

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Denisa Maděránková, Ph.D.

BRNO 2018

Bakalářská práce

bakalářský studijní obor **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Bc. Hedvika Urbanová

ID: 177787

Ročník: 3

Akademický rok: 2017/18

NÁZEV TÉMATU:

Preference synonymních kodonů v mitochondriálním genomu

POKYNY PRO VYPRACOVÁNÍ:

1) Proveďte literární rešerši na téma preference synonymních kodonů a popište parametry, které kodonovou preferenci charakterizují. 2) V libovolném programovém prostředí implementujte metody výpočtu preferenčních parametrů. 3) Implementaci doplňte uživatelským rozhraním a možností výpočtu preferenčních parametrů v plovoucím okně o nastavitelné velikosti a překryvu. 4) Pomocí vytvořené aplikace proveďte analýzu kodonové preference u vybraného souboru mitochondriálních genomů. 5) Výsledky diskutujte.

DOPORUČENÁ LITERATURA:

[1] SUN, Z.; WAN, D.; MURPHY, R.W.; MA, L.; ZHANG, X.; HUANG, D. Comparison of base composition and codon usage in insect mitochondrial genomes. *Genes Genom.* 2009, 31(1), 65–71.

[2] JIA, W. a P.G. HIGGS, Codon Usage in Mitochondrial Genomes: Distinguishing Context-Dependent Mutation from Translational Selection. *Mol. Biol. Evol.* 2008, 25(2), 339-351.

Termín zadání: 5.2.2018

Termín odevzdání: 25.5.2018

Vedoucí práce: Ing. Denisa Maděránková, Ph.D.

Konzultant:

prof. Ing. Ivo Provazník, Ph.D.
předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské 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 bakalářská práce se zabývá preferencí synonymních kodonů v mitochondriálním genomu. Jsou zde teoreticky rozebrány základní pojmy z oblastí bioinformatiky a genetiky. Kodonová preference je představena spolu s parametry, které ji charakterizují.

Praktická část bakalářské práce se zabývá návrhem funkcí pro výpočet parametrů kodonové preference. Mezi naprogramovanými parametry je relativní četnost synonymních kodonů, relativní adaptační index, kodonový adaptační index a homozygotnost kodonů. Dále je provedena analýza kodonové preference u mitochondriálního genomu *Homo sapiens*.

Klíčová slova

Preference kodonů, mitochondriální DNA, relativní četnost synonymních kodonů, relativní adaptační index, kodonový adaptační index, homozygotnost kodonů, efektivní počet kodonů

Abstract

This Bachelors thesis deals with synonymous codon usage bias in mitochondrial genomes. This thesis covers basic terms of bioinformatics and genetics. Codon usage bias is introduced together with parameters that characterize it.

Practical part of the thesis is dealing with programs that are designed to calculate parameters of codon usage bias. Programs that are described in this thesis are relative synonymous codon usage, relative adaptation index, codon adaptation index, homozygosity of codons, and effective number of codons. Synonymous codon usage bias in mitochondrial genome of *Homo sapiens* is analysed.

Keywords

Codon bias, mitochondrial DNA, relative synonymous codon usage, relative adaptation index, codon adaptation index, homozygosity of codons, effective number of codons

Bibliografická citace:

URBANOVÁ, H. Preference synonymních kodonů v mitochondriálním genomu. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 52 s. Vedoucí bakalářské práce Ing. Denisa Maděránková, Ph.D..

Prohlášení

Prohlašuji, že svou závěrečnou práci na téma Preference synonymních kodonů v mitochondriálním genomu jsem vypracovala samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce 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 části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne **25. května 2018**

.....
podpis autora

Poděkování

Děkuji vedoucí bakalářské práce Ing. Denise Maděránkové za její cenné rady a čas.

V Brně dne **25. května 2018**

.....
podpis autorky

Obsah

Úvod	1
1 Mitochondriální DNA	2
2 Preference kodonů	5
2.1 Vliv kodonové preference na genovou expresi	5
2.2 Relativní četnost synonymních kodonů	6
2.3 Relativní adaptační index	8
2.4 Kodonový adaptační index	9
2.5 Homozygotnost kodonů	10
2.6 Efektivní počet kodonů	11
2.7 Databáze GenBank	12
2.7.1 Fasta soubory	12
3 Programové řešení parametrů	13
3.1 Funkce pro RSCU	13
3.2 Funkce pro relativní adaptační index	15
3.3 Funkce pro kodonový adaptační index	17
3.4 Funkce pro homozygotnost kodonů	17
3.5 Funkce pro efektivní počet kodonů	19
4 Tvorba aplikace	21
4.1 Struktura aplikace	21
4.2 Popis fungování aplikace	23
5 Analýza kodonové preference	27
5.1 Výpočet efektivního počtu kodonů pro celý gen	27
5.2 Porovnání kodonové preference u aminokyselin	29
5.3 Porovnání genu ND1 a ND6 u obratlovců	31
6 Závěr	36
Literatura	37
Seznam symbolů, veličin a zkratk	39
A Vývojový diagram aplikace	40
B Výsledky analýzy	41

Seznam obrázků

Obr. 1.1: Struktura mitochondriální DNA.....	2
Obr. 4.1: Aplikace Kodonová preference po spuštění.....	22
Obr. 4.2: Vzor výpočtu RSCU pro celý gen.....	24
Obr. 4.3: Vzor grafu pro relativní četnost synonymních kodonů.....	25
Obr. 5.1: Kodonový adaptační index mtDNA <i>Homo sapiens</i> počítaný v okně.....	28
Obr. 5.2: Porovnání hodnot RSCU v okně pro asparagin genu ND1 <i>Homo sapiens</i> s <i>Antigonii capros</i>	32
Obr. 5.3: Porovnání hodnot RSCU v okně pro lysin genu ND1 <i>Homo sapiens</i> s <i>Antigonii capros</i>	32
Obr. 5.4: Porovnání hodnot RSCU v okně pro methionin genu ND1 <i>Homo sapiens</i> s <i>Antigonii capros</i>	33
Obr. 5.5: Porovnání hodnot RSCU v okně pro asparagin genu ND1 <i>Homo sapiens</i> s <i>Antigonii capros</i>	33
Obr. 5.6: Porovnání hodnot RSCU v okně pro lysin genu ND1 <i>Homo sapiens</i> s <i>Antigonii capros</i>	34
Obr. 5.7: Porovnání hodnot RSCU v okně pro methionin genu ND1 <i>Homo sapiens</i> s <i>Antigonii capros</i>	34

Seznam tabulek

Tab. 1.1: Geny mtDNA s krátkou specifikací.....	3
Tab. 1.2: Porovnání standardního genetického kódu s mitochondriálním obratlovců.....	4
Tab. 3.1: Hodnoty RSCU pro gen ND1 <i>Balaena mysticetus</i>	13
Tab. 3.2: Hodnoty relativního adaptačního indexu pro gen ND1 <i>Balaena mysticetus</i>	16
Tab. 3.3: Hodnoty homozygotnosti kodonů pro gen ND1 <i>Balaeny mysticetus</i>	18
Tab. 3.4: Hodnoty efektivních počtů kodonů pro gen ND1 <i>Balaeny mysticetus</i>	19
Tab. 5.1: Hodnoty efektivního počtu kodonů různých organismů.....	27

ÚVOD

Kodonová preference je výraz označující rozdílné používání synonymních kodonů v genech. Ve své podstatě se při počítání kodonové preference hledají rozdíly v četnostech používání kodonů synonymních, tedy kodonů kódujících stejnou aminokyselinu, v genech. Genetický kód je takzvaně degenerovaný, protože obsahuje 64 kodonů pro pouze 20 proteinogenních aminokyselin plus tři (v mitochondriální DNA dva) stop kodony. Některé aminokyseliny jsou proto kódovány více kodony.

Některé z těchto synonymních kodonů jsou v určitých genech používány častěji a některé vůbec. Protože má kodonová preference vliv na genovou expresi, počítají se hodnoty jejích parametrů. [1,2]

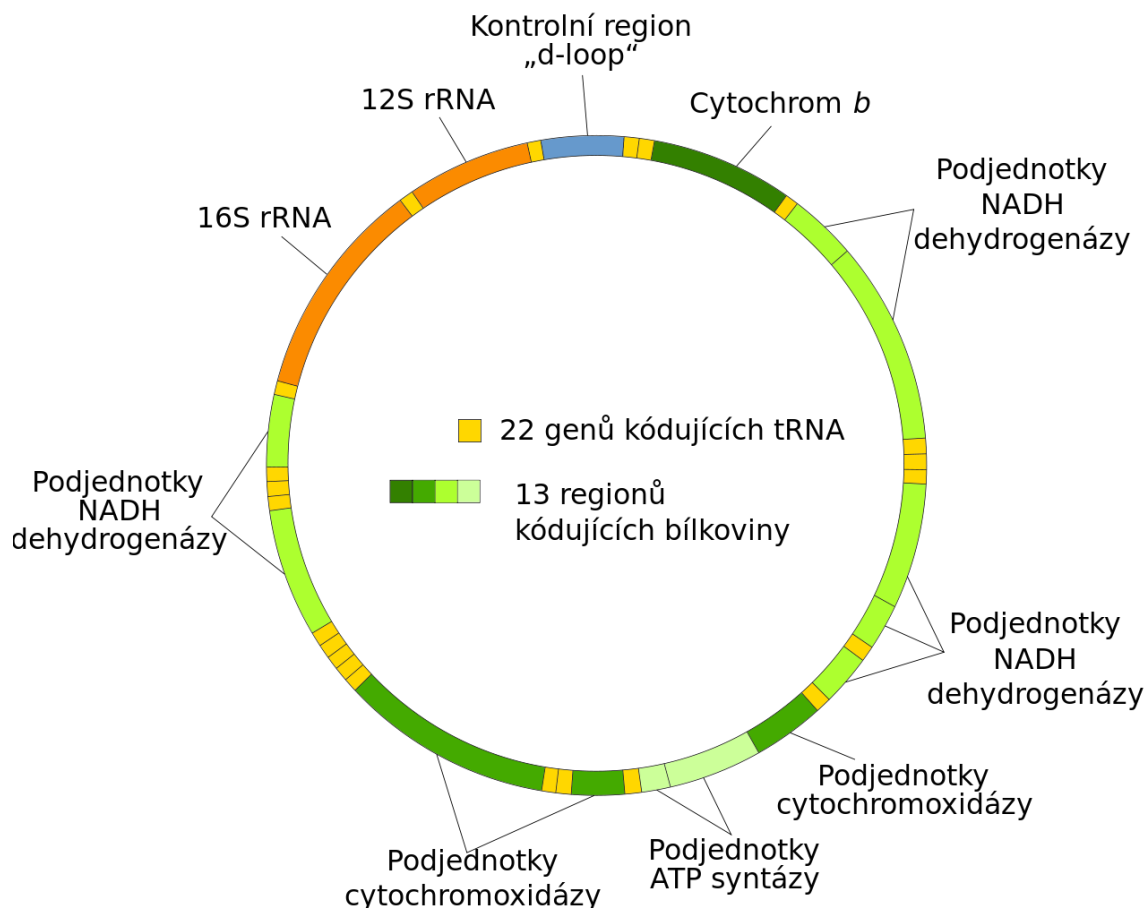
V této bakalářské práci se budu zabývat zejména preferencí kodonů v mitochondriálním genomu. Budou vysvětleny základní pojmy z oblasti genetiky a bioinformatiky, popsána základní charakteristika mitochondriálního genomu, co je to kodonová preference a parametry charakterizující kodonovou preferenci. Výpočty parametrů budou provedeny na krátké smyšlené sekvenci pro demonstraci postupu počítání jednotlivých parametrů. Parametry, kterými se práce bude zabývat, budou relativní četnost synonymních kodonů, relativní adaptační index, kodonový adaptační index, homozygotnost kodonů a efektivní počet kodonů.

Praktická část bakalářské práce bude obsahovat programování funkcí parametrů pro získání důležitých informací z mitochondriálního genomu. Mezi naprogramovanými parametry bude relativní četnost synonymních kodonů, relativní adaptační index, kodonový adaptační index, homozygotnost kodonů a efektivní počet kodonů. Tyto funkce budou využity při tvorbě aplikace s grafickým uživatelským rozhraním.

V závěru bude práce shrnuta z hlediska dosažených cílů. Zároveň bude vyhodnocena analýza kodonové preference.

1 MITOCHONDRIÁLNÍ DNA

Mitochondrie je semiautonomní organela, jejíž původ je, jak předpokládáme, pohlcení archebakterie, která se následně začlenila do buňky. Mitochondriální DNA je cirkulární dvouvláknová DNA prokaryotního typu. Uspořádání jednotlivých genů lze vidět na následujícím obrázku (viz obr. 2.1) [3].



Obr. 1.1: Struktura mitochondriální DNA [4].

Většina genů byla přemístěna do jaderné DNA. Mitochondriální DNA obsahuje jen 13 genů kódujících proteiny, dále 22 genů kódujících tRNA a 2 geny pro rRNA. V mitochondriích se nachází desítky kopií mitochondriální DNA, jelikož se cyklus replikace mitochondriální DNA neváže na buněčný cyklus. Díky častým replikacím a absenci opravných systémů (proof-reading) mitochondriální DNA snadno dochází k mutacím. Tyto mutace souvisí s chorobami MELAS (Mitochondria Myopathy, Encephalopathy, Lactic Acidosis, Stroke-like episodes) nebo MERF (Myoclonic Epilepsy with Ragged red Fibers) [3].

V následující tabulce (viz tab 1.1) je uveden seznam genů s krátkou specifikací jednotlivých genů. Všechny geny kódují proteiny, které jsou součástí dýchacího řetězce [3].

Tab. 1.1: Geny mtDNA s krátkou specifikací [3].

Gen	Název	Délka [bp]
ND1	NADH dehydrogenáza 1	955
ND2	NADH dehydrogenáza 2	1041
COX1	Cytochrom c oxidáza 1	1541
COX2	Cytochrom c oxidáza 2	683
ATP8	ATP syntáza 8	206
ATP6	ATP syntáza 6	680
COX3	Cytochrom c oxidáza 3	783
ND3	NADH dehydrogenáza 3	345
ND4L	NADH dehydrogenáza 4L	296
ND4	NADH dehydrogenáza 4	1377
ND5	NADH dehydrogenáza 5	1811
ND6	NADH dehydrogenáza 6	524
CYTB	Cytochrom b	1140

Genetický kód mitochondriální DNA je mírně odlišný od genetického kódu jaderné DNA. Jako příklad můžeme uvést kódování startkodonů (ATA a ATG) a dva stopkodony: AGA a AGG. U genetického kódu jaderné DNA je startkodon kódován pouze tripletem ATG a stopkodony jsou kódovány následovně: TAG, TGA a TAA. Je zajímavé, že genetický kód mitochondriální DNA se liší také mezi živočišnými druhy. Je rozdíl mezi genetickým kódem mtDNA obratlovců, bezobratlých, kvasinek, motolic atd. Rozdíl v genetickém kódu standardním a mitochondriálním u obratlovců je vidět v následující tabulce (tab. 1.2) [3].

V tabulce (tab. 1.2) si lze všimnout, že genetický kód se liší nejen ve start a stop kodonech, ale také v aminokyselinách jako je arginin, tryptofan a methionin.

Mitochondriální DNA je maternálně dědičná. Jinými slovy je mitochondriální genom děděn pouze po mateřské linii. Je to dáno polohou mitochondrií v gametách (spermie kolem bičíku a vajíčko uvnitř buňky) [3].

Tab. 1.2: Porovnání standardního genetického kódu s mitochondriálním obratlovců

Aminokyselina	Standardní	MtDNA obratlovců
Alanin	GCT, GCC, GCA, GCG	GCT, GCC, GCA, GCG
Arginin	CGT, CGC, CGA, CGG, AGA, AGG	CGT, CGC, CGA, CGG
Asparagin	AAT, AAC	AAT, AAC
Asparagová kyselina	GAT, GAC	GAT, GAC
Cystein	TGT, TGC	TGT, TGC
Glutamin	CAA, CAG	CAA, CAG
Glutamová kyselina	GAA, GAG	GAA, GAG
Glycin	GGT, GGC, GGA, GGG	GGT, GGC, GGA, GGG
Histidin	CAT, CAC	CAT, CAC
Isoleucin	ATT, ATC, ATA	ATT, ATC
Leucin	TTA, TTG, CTT, CTC, CTA, CTG	TTA, TTG, CTT, CTC, CTA, CTG
Lysin	AAA, AAG	AAA, AAG
Methionin	ATG	ATA, ATG
Fenylalanin	TTT, TTC	TTT, TTC
Prolin	CCT, CCC, CCA, CCG	CCT, CCC, CCA, CCG
Serin	TCT, TCC, TCA, TCG, AGT, AGC	TCT, TCC, TCA, TCG, AGT, AGC
Threonin	ACT, ACC, ACA, ACG	ACT, ACC, ACA, ACG
Tryptofan	TGG	TGA, TGG
Tyrosin	TAT, TAC	TAT, TAC
Valin	GTT, GTC, GTA, GTG	GTT, GTC, GTA, GTG
Stop kodony	TAA, TAG, TGA	TAA, TAG, AGA, AGG
Start kodony	TTG, CTG, ATG	ATT, ATC, ATA, ATG, GTG

2 PREFERENCE KODONŮ

Kodonová preference značí rozdíly v četnostech používání synonymních kodonů (kodonů kódujících stejnou aminokyselinu) v genech. Genetický kód obsahuje 64 různých kodonů, ale přiřazuje je jen 20 aminokyselinám plus třem (v mitochondriální DNA dvěma) stop kodonům. Kvůli nerovnosti těchto dvou čísel je genetický kód tzv. degenerovaný. Některé aminokyseliny jsou kódovány více kodony. V různých organismech lze nalézt různé četnosti v používání kodonů. Některé kodony jsou proto preferovány [1,2].

Dle Plotkina a Kudly (2010) se kodonová preference v současnosti považuje za stěžejní při genové expresi od transkripce k translaci. Kodonová preference je přirozeně všudypřítomná a může být velice silná. Některé kodony jako třeba UUG, CCA a GCA se například vůbec nevyskytují v organismu *Thermus thermophilus*. Kodonová preference se může lišit jak v rámci celého genomu, tak i v částech genu [1,4].

Mezi mnoha navrhovanými myšlenkami se rovnováha mezi mutacemi, přirozeným výběrem a genetickým driftem zdá být většinou akceptována. Vzájemně se ale nevylučují [2].

Parametry kodonové preference, kterými se tato bakalářská práce zabývá, jsou: relativní četnost synonymních kodonů, relativní adaptační index, kodonový adaptační index, homozygotnost kodonů a efektivní počet kodonů.

Při výpočtu základních parametrů preference kodonů se počítá relativní a absolutní četnost. Absolutní četnost je skutečný počet výskytu daného kodonu v genu. Relativní četnost kodonu je absolutní četnost dělená součtem absolutních četností všech synonymních kodonů pro danou aminokyselinu. Absolutní a relativní četnosti jsou základním kamenem při výpočtu parametrů kodonové preference [2,5].

V každé následující podkapitole budou vypočítány parametry kodonové preference pro zadanou sekvenci kodonů. Tento výpočet bude simulovat postup funkce při výpočtu těchto parametrů popsáném v následující kapitole.

Pro demonstraci postupu výpočtu parametrů kodonové preference bude použit následující příklad:

Sekvence kodonů: ATG TTT ATA ATT AAC ATC CTA ACA CTT ATC

Sekvence aminokyselin: M F M I N I L T L I

2.1 Vliv kodonové preference na genovou expresi

Dle Kurlanda (1991) kodonová preference ovlivňuje rychlost genové exprese. Rychlost, jakou se transkribuje každá tRNA, je ovlivněna frekvencí kodonu vyskytující se v mediátorové RNA. Čím je větší četnost kodonu, tím více tRNA pro daný kodon je transkribováno. Zároveň se pak při translaci rychleji tvoří peptidový

řetězec, než by tomu bylo při absenci kodonové preference. Dá se tedy říci, že díky kodonové preferenci se značně zefektivňuje proces genové exprese. [6]

Pokud buňka přenáší jediný druh mRNA, pak, jak tvrdí Kurland (1991), rychlost, s níž se provádí prodlužovací proces, může ovlivnit rychlost exprese proteinu, pokud jsou v omezujících množstvích přítomny ribozomy. Je tomu tak proto, že čím rychleji je polypeptidový řetězec dokončen, tím rychleji se ribozomy mohou vrátit k zahájení a dokončení jiného polypeptidového řetězce. Nicméně pokud buňka zpracovává stovky různých druhů mRNA, pak se mění rychlost, kterou jsou překládány, a nebude mít vliv na rychlost exprese příslušného proteinu. [6]

Dle Kurlanda (1991) bylo také prokázáno, že při změně sekvence mRNA, pouze však ve smyslu synonymních kodonů, byla rychlost genové exprese výrazně odlišná. [6]

Existuje tedy důvod se domnívat, tvrdí Kurland (1991), že interakce mezi ribozomální RNA a počátečními sekvencemi kódujícími mRNA ovlivňují účinnost translační iniciace. Nicméně statistická jednotnost frekvencí kodonů pozorovaných v počátečních kódujících sekvencích mRNA překládaných na velmi rozdílných úrovních exprese naznačuje, že při výběru tohoto intragenního zkreslení musí být zahrnuto něco jiného. Může se jednat o to, že jednotné, optimální oddělení ribozomů podél polyzomu, je dosaženo počátečními kodonovými řetězci. [6]

Zároveň můžeme kodonovou preferenci využít i při hledání exonů. Pokud nalezený exon obsahuje hodně vzácných kodonů, pak zřejmě nepůjde o exon. [6]

2.2 Relativní četnost synonymních kodonů

RSCU (Relative Synonymous Codon Usage) je parametr, u něhož se počítá relativní počet kodonů zastoupených v genu. Hodnota četností jednotlivých kodonů se vztahuje vůči hodnotě četnosti kodonu, která by byla dosažena, pokud by byl daný kodon jediným kodonem v genu pro danou aminokyselinu. Jinými slovy počítá četnost jednotlivých kodonů, která je násobena počtem synonymních kodonů pro danou aminokyselinu, a dělí tuto hodnotu součtem četností všech synonymních kodonů v genu pro danou aminokyselinu [7].

Hodnota RSCU rovna 1 neukazuje na kodonovou preferenci, naopak říká, že synonymní kodony pro danou aminokyselinu jsou zastoupeny v genu rovnoměrně. Pokud se jako výsledek objeví nula, ukazuje to na skutečnost, že daný kodon se v genu vůbec nevyskytuje. Jako preferenci kodonu můžeme označit situaci, při které jsou hodnoty RSCU větší než 1 [8].

Při počítání RSCU počítáme relativní četnost, jejíž vzorec je následující:

$$RF_{ac} = \frac{n_{ac}}{\sum_{c=1}^{d_a} n_{ac}} \quad (1)$$

Kde RF_{ac} je hodnota relativní četnosti pro daný kodon, n_{ac} je četnost daného kodonu v genu. Hodnota sumy je součet četností všech synonymních kodonů v genu pro danou aminokyselinu.

Vzorec pro výpočet hodnot RSCU je následující:

$$RSCU_{ac} = \frac{d_a \times n_{ac}}{\sum_{c=1}^{d_a} n_{ac}} = d_a \times RF_{ac} \quad (2)$$

Kde $RSCU_{ac}$ je hodnota RSCU pro daný kodon, d_a je počet synonymních kodonů pro danou aminokyselinu.

Příklad:

Sekvence kodonů: ATG TTT ATA ATT AAC ATC CTA ACA CTT ATC

Sekvence aminokyselin: M F M I N I L T L I

$$RSCU_{ATG} = \frac{d_M \cdot n_{ATG}}{n_{ATG} + n_{ATA}} = \frac{2 \times 1}{1 + 1} = 1$$

$$RSCU_{ATA} = \frac{d_M \cdot n_{ATA}}{n_{ATG} + n_{ATA}} = \frac{2 \times 1}{1 + 1} = 1$$

$$RSCU_{TTT} = \frac{d_F \cdot n_{TTT}}{n_{TTT}} = \frac{2 \times 1}{1} = 2$$

$$RSCU_{ATT} = \frac{d_I \cdot n_{ATT}}{n_{ATT} + n_{ATC}} = \frac{2 \times 1}{1 + 2} = \frac{2}{3}$$

$$RSCU_{ATC} = \frac{d_I \cdot n_{ATC}}{n_{ATT} + n_{ATC}} = \frac{2 \times 2}{1 + 2} = \frac{4}{3}$$

$$RSCU_{AAC} = \frac{d_N \cdot n_{AAC}}{n_{AAC}} = \frac{2 \times 1}{1} = 2$$

$$RSCU_{CTA} = \frac{d_L \cdot n_{CTA}}{n_{CTA} + n_{CTT}} = \frac{6 \times 1}{1 + 1} = 3$$

$$RSCU_{CTT} = \frac{d_L \cdot n_{CTT}}{n_{CTA} + n_{CTT}} = \frac{6 \times 1}{1 + 1} = 3$$

$$RSCU_{ACA} = \frac{d_T \cdot n_{ACA}}{n_{ACA}} = \frac{4 \times 1}{1} = 4$$

V uvedených výpočtech a výsledcích jsou hodnoty RSCU relevantní v delších sekvencích. Lze si všimnout, že hodnoty pro kodony, u kterých je vyšší stupeň degenerace, vycházejí vyšší hodnoty čísel než pro kodony, které mají stupeň degenerace nižší.

2.3 Relativní adaptační index

Hodnota relativního adaptačního indexu pro daný kodon se spočítá pomocí hodnot RSCU. Hodnota RSCU pro daný kodon je dělena maximální hodnotou RSCU (ze všech synonymních kodonů pro danou aminokyselinu). Vzorec pro výpočet je následující:

$$w_{ac} = \frac{RSCU_{ac}}{RSCU_{max}} \quad (3)$$

Kde $RSCU_{ac}$ je hodnota RSCU pro daný kodon příslušné aminokyseliny a $RSCU_{max}$ je maximální hodnota RSCU vybraná ze všech synonymních kodonů kódujících danou aminokyselinu. [8]

Hodnoty se tedy standardizují na hodnoty od nuly do jedné. Hodnotu relativního adaptačního indexu přiřadíme nejpreferovanějšímu synonymnímu kodonu. S hodnotou přibližující se k nule získáme méně používané kodony, jinými slovy, čím menší hodnota relativního adaptačního indexu, tím méně lze daný kodon v genu použít. Nulová hodnota relativního adaptačního indexu říká, že se kodon pro danou aminokyselinu v genu vůbec nevyskytuje.

Příklad:

Sekvence kodonů: ATG TTT ATA ATT AAC ATC CTA ACA CTT ATC

Sekvence aminokyselin: M F M I N I L T L I

$$\begin{aligned} w_{ATG} &= \frac{RSCU_{ATG}}{\max\{RSCU_M\}} = \frac{1}{1} = 1 \\ w_{ATA} &= \frac{RSCU_{ATA}}{\max\{RSCU_M\}} = \frac{1}{1} = 1 \\ w_{TTT} &= \frac{RSCU_{TTT}}{\max\{RSCU_F\}} = \frac{2}{2} = 1 \\ w_{ATT} &= \frac{RSCU_{ATT}}{\max\{RSCU_I\}} = \frac{\frac{2}{3}}{\frac{4}{3}} = \frac{1}{2} \\ w_{ATC} &= \frac{RSCU_{ATC}}{\max\{RSCU_L\}} = \frac{\frac{4}{3}}{\frac{4}{3}} = 1 \\ w_{AAC} &= \frac{RSCU_{AAC}}{\max\{RSCU_N\}} = \frac{2}{2} = 1 \\ w_{CTA} &= \frac{RSCU_{CTA}}{\max\{RSCU_L\}} = \frac{3}{3} = 1 \\ w_{CTT} &= \frac{RSCU_{CTT}}{\max\{RSCU_L\}} = \frac{3}{3} = 1 \end{aligned}$$

$$w_{ACA} = \frac{RSCU_{ACA}}{\max\{RSCU_T\}} = \frac{4}{4} = 1$$

2.4 Kodonový adaptační index

Kodonový adaptační index, zkratkou CAI (Codon Adaptation Index), je parametr kodonové preference. Při výpočtu se používá vypočítaný relativní adaptační index všech kodonů, které nemají relativní adaptační index roven nule [9].

Hledá se tzv. geometrický průměr. Geometrický průměr se dostane pomocí součinu nenulových hodnot relativního adaptačního indexu a následném odmocnění L-tou odmocninou, kde L je počet kodonů v genu. Počítání geometrického průměru má význam při výpočtu typické hodnoty čísel. Nahrazuje tedy soubor hodnot jednou hodnotou, která je typickým příkladem daného souboru čísel. V tomto případě soubor čísel představují nenulové hodnoty relativního adaptačního indexu [9].

Hodnota kodonového adaptačního indexu pro daný gen vychází pouze jedna. Hodnoty vycházejí mezi nulou a jedničkou. Čím více se hodnoty kodonového adaptačního indexu blíží jedné, tím víc se v genu objevuje výrazná kodonová preference. Naopak hodnoty blíží se nule značí to, že kodonová preference není v daném genu výrazná [9].

Vzorec pro výpočet je následující:

$$CAI = \sqrt[L]{\prod_{c=1}^L w_c} \quad (4)$$

Kde L je počet kodonů v počítané sekvenci, w_c je nenulová hodnoty relativního adaptačního indexu pro daný kodon [9].

Příklad

Sekvence kodonů: ATG TTT ATA ATT AAC ATC CTA ACA CTT ATC

Sekvence aminokyselin: M F M I N I L T L I

$$\begin{aligned} CAI &= \sqrt[10]{w_{ATG} \times w_{TTT} \times w_{ATA} \times w_{ATT} \times w_{AAC} \times w_{ATC} \times w_{CTA} \times w_{ACA} \times w_{CTT} \times w_{ATC}} \\ CAI &= \sqrt[10]{1 \times 1 \times 1 \times 0.5 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1} \\ CAI &= \sqrt[10]{0.5} \\ CAI &= 0.9330 \end{aligned}$$

Tato hodnota blíží se jedné by ukazovala na velkou kodonovou preferenci, ale tento parametr kodonové preference je vhodný počítat pro delší nukleotidové sekvence.

2.5 Homozygotnost kodonů

Homozygotnost kodonů je parametr kodonové preference, který se značí F . Tento parametr udává, jak rovnoměrně, či nerovnoměrně jsou rozloženy synonymní kodony v genu. Jinými slovy, s jakou frekvencí se objevují synonymní kodony dané aminokyseliny v genu [2].

Je-li rozložení kodonů pro danou aminokyselinu rovnoměrné, pak vychází malé hodnoty homozygotnosti (blíží se nule). Naopak pokud je nějaký kodon před těmi synonymními preferován, pak se hodnoty blíží 1. Vzorec pro homozygotnost kodonů je následující:

$$F_a = \frac{N_a \times \sum_{c=1}^N f_{ac}^2 - 1}{N_a - 1} = \frac{N_a \times \sum_{c=1}^N \left(\frac{n_{ac}}{N_a}\right)^2 - 1}{N_a - 1} \quad (5)$$

Kde N_a je počet dané aminokyseliny v genu a n_{ac} je absolutní četnost daného synonymního kodonu v genu [2].

Maximální hodnota homozygotnosti je jedna, což znamená, že je používán jen jeden ze synonymních kodonů. Z hodnoty homozygotnosti vychází další parametr, který je probrán v následující podkapitole. Jde o parametr efektivní počet synonymních kodonů.

Homozygotnost kodonů lze spočítat pouze tehdy, když je daná aminokyselina obsažena v počítané sekvenci alespoň jednou.

Příklad:

Sekvence kodonů: ATG TTT ATA ATT AAC ATC CTA ACA CTT ATC

Sekvence aminokyselin: M F M I N I L T L I

$$F_M = \frac{N_M \times \left(\frac{n_{ATG}^2}{N_M^2} + \frac{n_{ATA}^2}{N_M^2}\right) - 1}{N_M - 1} = \frac{2 \times (0,25 + 0,25) - 1}{2 - 1} = 0$$

$$F_L = \frac{N_L \times \left(\frac{n_{CTA}^2}{N_L^2} + \frac{n_{CTT}^2}{N_L^2}\right) - 1}{N_L - 1} = \frac{2 \times (0,25 + 0,25) - 1}{2 - 1} = 0$$

$$F_I = \frac{N_I \times \left(\frac{n_{ACA}^2}{N_I^2} + \frac{n_{ATC}^2}{N_I^2}\right) - 1}{N_I - 1} = \frac{3 \times \left(\frac{1}{9} + \frac{4}{9}\right) - 1}{3 - 1} = \frac{1}{3}$$

Z hodnot, které vyšly, je patrné, že u izoleucinu se vyskytuje kodonová preference. Lze si povšimnout, že hodnota absolutní četnosti u kodonu ATC vychází

dva a kodonu ACA jedna, už to napovídá, že se zde může vyskytovat kodonová preference. Jak moc bude kodonová preference signifikantní, je vyjádřeno pomocí parametru homozygotnosti. Díky tomu, že u methioninu a leucinu vyšla nula, se dá tvrdit, že kodonová preference u těchto aminokyselin pro tuto sekvenci neexistuje a synonymní kodony jsou rozloženy zcela rovnoměrně. Tyto výsledky ovšem nejsou příliš reprezentativní, jelikož byly počítány pouze pro sekvenci deseti kodonů.

2.6 Efektivní počet kodonů

Jak již bylo zmíněno v předchozí podkapitole, efektivní počet kodonů vychází z hodnot homozygotnosti. Hodnota efektivního počtu kodonů je převrácená hodnota homozygotnosti. Tato hodnota udává počet používaných kodonů v genu pro danou aminokyselinu. Nejmenší hodnota je jedna a největší hodnota je stupeň degenerace pro danou aminokyselinu. Pokud je hodnota jedna, značí to kodonovou preferenci. Pokud se hodnota blíží nebo je rovna stupni degenerace pro danou aminokyselinu, pak se kodonovou preferencí daná aminokyselina nevyznačuje. [9]

Vzorec pro výpočet efektivního počtu kodonů je následující:

$$N_c = \frac{1}{F_c} \quad (6)$$

Kde N_c je efektivní počet kodonů a F_c homozygotnost kodonů pro danou aminokyselinu.

Pokud by bylo cílem zjistit efektivní počet kodonů pro příklad spočítaný pro předchozí parametry, vycházelo by se pouze z jedné vypočtené hodnoty homozygotnosti. Efektivní počet kodonů lze počítat, pokud homozygotnost není rovna nule. Pokud se hodnota homozygotnosti nule rovná, považuje se hodnota efektivního počtu kodonů za počet všech synonymních kodonů pro danou aminokyselinu.

$$N_M=2$$

$$N_L=6$$

$$N_I = \frac{1}{\frac{1}{3}} = 3$$

Efektivní počet kodonů lze získat i pro celý gen. Pokud se spočítá hodnota pro celý gen, pak se hodnoty pohybují od 20 do 61. Udává počet používaných kodonů v genu pro všechny aminokyseliny. Vzorec pro mitochondriální genetický kód obratlovců je následující:

$$N = \frac{12}{F_2} + \frac{6}{F_4} + \frac{2}{F_6} \quad (7)$$

kde:

$$F_2 = \frac{F_N+F_D+F_C+F_Q+F_E+F_H+F_I+F_K+F_M+F_F+F_W+F_Y}{12} \quad (8)$$

$$F_4 = \frac{F_A+F_R+F_G+F_P+F_T+F_V}{6} \quad (9)$$

$$F_6 = \frac{F_L+F_S}{2} \quad (10)$$

Hodnotu efektivního počtu kodonů pro celou sekvenci nemá smysl počítat, jelikož je to sekvence velmi krátká a hodnota by neměla žádný význam.

2.7 Databáze GenBank

Genbank je databáze, která je veřejně přístupná, a obsahuje nukleotidové sekvence více než 260 000 druhů organismů [10].

Z databáze genbank je možné stahovat geny i celé genomy. Soubor genbank obsahuje název souboru, délku sekvence, počet vláken a topologii daného genetického materiálu (např: circular) [10].

Dále lze v souboru najít informace o typu sekvence (např. DNA, RNA) či datum poslední úpravy sekvence. Samozřejmě se v souboru vyskytuje informace o organismu a typu DNA, jestli je to celý genom, nebo jen samotný gen. Zároveň v něm lze dohledat specifikaci druhu z pohledu taxonomie, například [10]:

Heterodontus francisci
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Elasmobranchii; Galeomorphii; Heterodontoidea; Heterodontiformes;
Heterodontidae; Heterodontus.

Dá se najít i informace o zdroji, tedy odkud informace pocházejí (vědecké články atd.). Dále se zde nachází vlastnosti genu nebo genomu, co a na jakých pozicích se nachází včetně kódované tRNA a rRNA. [10]

Důležitými informacemi jsou pozice genů, které jsou použity pro extrahování exonů. Jak už bylo zmíněno v mitochondriálním genomu, je 13 genů kódujících proteiny, proto v souboru pro mitochondriální genom je 13 vektorů s pozicemi kódujících sekvencí [10].

Poslední informací, která se v souboru nachází, je samotná sekvence. Sekvence je dále bioinformaticky zpracovávána [10].

2.7.1 Fasta soubory

Fasta je formát souborů obsahující zpravidla genomická data organismů. Oproti genbank souboru obsahuje však méně informací.

Fasta soubor je tvořen hlavičkou a sekvencí. V prvním řádku fasta souboru se nachází hlavička, která je tvořena znakem „>“ a slovem *source*, pak následuje název organismu. Od druhého řádku pak začíná sekvence, která je také uvozena znakem „>“.

Fasta formát je používán zejména z důvodu jednoduchosti. Pomocí online konvertorů lze mezi sebou převádět genbank a fasta soubory. Lze tak nastavit, aby se z genbank souboru stal fasta soubor s oddělenými kódujícími částmi, které pak lze snadno indexovat.

3 PROGRAMOVÉ ŘEŠENÍ PARAMETRŮ

Tato kapitola se zabývá zejména programovým řešením výpočtu parametrů kodonové preference. Všechny výpočty byly provedeny v programovém prostředí MATLAB 2014a.

MATLAB je programovací prostředí a skriptovací jazyk čtvrté generace vytvořený společností MathWorks. MATLAB je zkratkou pro „Matrix Laboratory“, tedy maticová laboratoř. Toto programovací prostředí je vhodné pro práci s maticemi a vektory, které jsou základní datovou strukturou. Tento software je vhodným nástrojem pro vědeckotechnické výpočty [11].

V následujících podkapitolách jsou popsány funkce vytvořené pro výpočet parametrů kodonové preference. Každý parametr má vytvořenou vlastní funkci, kterou je možné volat z příkazového řádku. Pro zpracování dat sekvencí se pracuje s genbank soubory.

Při stažení genu či celého genomu lze soubor nahrát jako strukturu do MATLABu příkazem *genbankread*.

Pro další zpracování se musí z uložené sekvence extrahovat exony. K tomuto účelu byla vytvořena pomocná funkce *ExtractExon*, která z nahrané struktury vytvoří vektor s nukleotidovou sekvencí (jeden gen). K tomu, aby bylo možné použít funkci *ExtractExon*, je potřeba znát začátek a konec genu. Proto je nutné před použitím funkce *ExtractExon* použít pomocnou funkci *indices*, která ze struktury zjistí začátek a konec požadovaného genu.

Když je sekvence, kterou je cílem spočítat, uložena ve vektoru jako proměnná, je možné ji použít jako vstup do následujících funkcí. Jelikož je zpracovávána mitochondriální DNA, musí se nastavit genetický kód na 2, protože MATLAB má genetický kód pro mitochondriální DNA obratlovců uložený právě pod tímto číslem.

V každé z následujících podkapitol je pomocí funkcí spočítán určitý gen pro ukázkou typických hodnot následujících parametrů. Gen patří do mitochondriální DNA velryby grónské a kóduje NADH dehydrogenázu 1. Délka genu je 955 párů bází. NADH dehydrogenáza je součástí dýchacího řetězce.

3.1 Funkce pro RSCU

Tato funkce počítá pro každou aminokyselinu a každý synonymní kodon hodnotu relativní četnosti synonymních kodonů. Funkce má dva vstupy a jeden výstup. Výstupem je matice relativních četností a vstupem je počítaná sekvence a potom druh genetického kódu, se kterým pracujeme.

Funkce nejdříve příkazem *revgeneticcode* vypíše strukturu genetického kódu pro všechny aminokyseliny. Příkazem *struct2cell* je převedena struktura na buňkové pole. Vstupní sekvence uložena ve vektoru je z malých písmen převedena

na velká písmena. Poté je cyklus *for* buňkové pole procházeno spolu se vstupním vektorem sekvence a porovnáván každý kodon. Takto se spočítají absolutní četnosti jednotlivých kodonů.

Následuje cyklus *for*, který prochází matici četností a přepočítává je z absolutních četností na hodnoty relativních četností synonymních kodonů. Pro ukázkou hodnot relativních četností synonymních kodonů je sestavena následující tabulka (viz tab. 3.1).

Tab. 3.1: Hodnoty RSCU pro gen ND1 *Balaena mysticetus*

Alanin	GCT	GCC	GCA	GCG	-	-
	0,4000	1,8667	1,7333	0,0000	-	-
Arginin	CGT	CGC	CGA	CGG	-	-
	0,0000	0,5000	3,5000	0,0000	-	-
Asparagin	AAT	AAC	-	-	-	-
	0,5455	1,4545	-	-	-	-
Asparagová kyselina	GAT	GAC	-	-	-	-
	1,5000	0,5000	-	-	-	-
Cystein	TGT	TGC	-	-	-	-
	0,0000	2,0000	-	-	-	-
Glutamin	CAA	CAG	-	-	-	-
	2,0000	0,0000	-	-	-	-
Glutamová kyselina	GAA	GAG	-	-	-	-
	2,0000	0,0000	-	-	-	-
Glycin	GGT	GGC	GGA	GGG	-	-
	0,6667	1,3333	1,3333	0,6667	-	-
Histidin	CAT	CAC	-	-	-	-
	1,0000	1,0000	-	-	-	-
Isoleucin	ATT	ATC	-	-	-	-
	0,6400	1,3600	-	-	-	-
Leucin	TTA	TTG	CTT	CTC	CTA	CTG
	0,7500	0,0000	0,4286	1,0714	3,1071	0,6429
Lysin	AAA	AAG	-	-	-	-
	2,0000	0,0000	-	-	-	-
Methionin	ATA	ATG	-	-	-	-
	1,3684	0,6316	-	-	-	-
Fenylalanin	TTT	TTC	-	-	-	-
	0,2857	1,7143	-	-	-	-
Prolin	CCT	CCC	CCA	CCG	-	-
	0,9091	1,8182	1,2727	0,0000	-	-
Serin	TCT	TCC	TCA	TCG	AGT	AGC
	0,2857	3,1429	1,4286	0,0000	0,0000	1,1429
Threonin	ACT	ACC	ACA	ACG	-	-
	0,4286	1,1429	2,4286	0,0000	-	-
Tryptofan	TGA	TGG	-	-	-	-

	1,5556	0,4444	-	-	-	-
Tyrosin	TAT	TAC	-	-	-	-
	1,0909	0,9091	-	-	-	-
Valin	GTT	GTC	GTA	GTG	-	-
	0,0000	0,3636	3,6364	0,0000	-	-

V tabulce (viz tab. 3.1) si lze všimnout, že součty všech hodnot pro jednu aminokyselinu jsou stejné, jako je počet synonymních kodonů pro danou aminokyselinu. Též se dá vyvodit, že pro tři z uvedených kodonů objevuje 0. Dá se tedy říci, že se v genu vůbec neobjevuje. Naopak pro kodon CGA vyšla vysoká hodnota 3,5, což ukazuje na velkou kodonovou preferenci. Aminokyselina alanin má dva preferované kodony GCC a GCA.

Funkce pro počítání v okně

Tato funkce s názvem *RSCUok* je určena pro výpočet relativní četnosti synonymních kodonů v plovoucím okně s překryvem. Funkce má tři vstupy. Jedním z nich je samotná sekvence, druhým vstupem je velikost okna a třetím překryv v procentech. Funkce má dva výstupy ve formátu buňkových polí. Jedním z nich je kód a druhým je buňkové pole s vypočítanými hodnotami parametru.

Funkce funguje na podobném principu jako funkce pro výpočet celého genu. Ovšem jediným rozdílem je, že parametr počítá v plovoucím okně. Toho je dosaženo cyklem *while*, ve kterém se prochází vždy část sekvence od prvního nukleotidu až po velikost okna. Tato sekvence vstupuje do tří cyklů *for*, ve kterých se spočítají jednotlivé četnosti. Počítaná sekvence se poté smaže.

Na cyklus *while*, který spočítá četnosti kodonů, navazují cykly *for*, ve kterých se spočítá hodnota samotného parametru relativní četnosti synonymních kodonů. Výsledné buňkové pole má strukturu 20 řádků a 6 sloupců. V každém poli je uložen vektor s hodnotami. Hodnoty jdou po sobě stejně jako postupně části sekvencí, které jsou počítané.

3.2 Funkce pro relativní adaptační index

Tato funkce k výpočtu potřebuje i funkci popsány v předchozí podkapitole, a to funkci pro relativní četnost synonymních kodonů. Funkce má jeden výstup a dva vstupy, které se neliší od funkce popsané v předchozí kapitole. Výstupem funkce je matice relativního adaptačního indexu.

Ve funkci se prvně volá funkce pro relativní četnost synonymních kodonů. Ta se uloží do proměnné, ta poté prochází cykly *for* a přepočítává se na relativní adaptační index. Protože se může stát, že maximum v řádku matice RSCU bude rovno nule, je cyklus *for* ošetřen podmínkou *if*, která zamezí tomu, že by se ve jmenovateli objevila nula.

V následující tabulce je ukázka typických hodnot relativního adaptačního indexu pro gen ND1 mitochondriálního genomu velryby grónské. K výpočtu byla použita navržená funkce.

Tab. 3.2: Hodnoty relativního adaptačního indexu pro gen ND1 *Balaena mysticetus*

Alanin	GCT	GCC	GCA	GCG	-	-
	0,2143	1,0000	0,9286	0,0000	-	-
Arginin	CGT	CGC	CGA	CGG	-	-
	0,0000	0,1429	1,0000	0,0000	-	-
Asparagin	AAT	AAC	-	-	-	-
	0,3750	1,0000	-	-	-	-
Asparagová kyselina	GAT	GAC	-	-	-	-
	1,0000	0,3333	-	-	-	-
Cystein	TGT	TGC	-	-	-	-
	0,0000	1,0000	-	-	-	-
Glutamin	CAA	CAG	-	-	-	-
	1,0000	0,0000	-	-	-	-
Glutamová kyselina	GAA	GAG	-	-	-	-
	1,0000	0,0000	-	-	-	-
Glycin	GGT	GGC	GGA	GGG	-	-
	0,5000	1,0000	1,0000	0,5000	-	-
Histidin	CAT	CAC	-	-	-	-
	1,0000	1,0000	-	-	-	-
Isoleucin	ATT	ATC	-	-	-	-
	0,4706	1,0000	-	-	-	-
Leucin	TTA	TTG	CTT	CTC	CTA	CTG
	0,2414	0,0000	0,1379	0,3448	1,0000	0,2069
Lysin	AAA	AAG	-	-	-	-
	1,0000	0,0000	-	-	-	-
Methionin	ATA	ATG	-	-	-	-
	1,0000	0,4615	-	-	-	-
Fenylalanin	TTT	TTC	-	-	-	-
	0,1667	1,0000	-	-	-	-
Prolin	CCT	CCC	CCA	CCG	-	-
	0,5000	1,0000	0,7000	0,0000	-	-
Serin	TCT	TCC	TCA	TCG	AGT	AGC
	0,0909	1,0000	0,4545	0,0000	0,0000	0,3636
Threonin	ACT	ACC	ACA	ACG	-	-
	0,1765	0,4706	1,0000	0,0000	-	-
Tryptofan	TGA	TGG	-	-	-	-
	1,0000	0,2857	-	-	-	-
Tyrosin	TAT	TAC	-	-	-	-
	1,0000	0,8333	-	-	-	-
Valin	GTT	GTC	GTA	GTG	-	-
	0,0000	0,1000	1,0000	0,0000	-	-

Hodnoty v tabulce se pohybují od nuly do jedné. Nula vyšla pro kodony, které se v genu nevyskytují. Hodnota rovna jedné vyšla pro nejčastěji se vyskytující kodony – pro alanin GCC a pro arginin CGA (viz tab. 3.2).

Funkce pro počítání v okně

Tato funkce pracuje velice podobně jako funkce pro relativní četnost synonymních kodonů. Funkce obsahuje dvě vstupní proměnné. První je zadána sekvence, poté okno a překryv. Ve funkci je volána *RSCUok*. Buňkové pole relativních četností synonymních kodonů je pak přepočítáno na relativní adaptační index, jak bylo uvedeno v kapitole 2.

3.3 Funkce pro kodonový adaptační index

Funkce pro kodonový adaptační index je závislá na funkci pro relativní adaptační index. Tato funkce má jeden výstup a dva vstupy, které jsou stejné jako ve funkcích probraných v předchozích podkapitolách. Výstupem je jedna hodnota pro kodonový adaptační index platná pro celou počítanou sekvenci.

Funkce volá funkci relativního adaptačního indexu a ukládá hodnoty do proměnné. Poté následují cykly *for*, které prochází matici relativního adaptačního indexu a násobí je mezi sebou. Násobí se mezi sebou pouze nenulové indexy, což zabezpečuje podmínka *if*. Pomocí příkazu *length* se zjišťuje délku počítané sekvence. Když se tato délka podělí třemi, je získán počet kodonů v sekvenci. Tento údaj je poté použit při umocnění součinu $1/L$, kde L je počet kodonů v sekvenci. Hodnota kodonového adaptačního indexu počítaná pro gen ND1 velryby grónské je následující:

$$CAI_{ND1} = 0,9230$$

Tato hodnota je vysoká, což ukazuje na kodonovou preferenci.

Funkce pro počítání v okně

Tato funkce přepočítává hodnoty z funkcí předchozích. Je volána funkce *wok*, která vytvoří svým výstupem buňkové pole relativního adaptačního indexu. Pro nenulové hodnoty je poté spočítán geometrický průměr. Výstupem funkce je vektor hodnot kodonového adaptačního indexu počítaný v různých částech sekvence.

3.4 Funkce pro homozygotnost kodonů

Tato funkce má jeden výstup a dva vstupy. Vstupy jsou opět sekvence a genetický kód. Výstupem je vektor hodnot homozygotností synonymních kodonů pro každou z aminokyselin.

Funkce pomocí příkazů *nt2aa* převádí nukleotidové sekvence na sekvenci aminokyselin. Poté používá příkaz *aacount*, který spočítá absolutní četnosti aminokyselin. Jelikož je výstupem tohoto příkazu struktura, je třeba ji převést na

buňkové pole pomocí příkazu *struc2cell*. Poté je pomocí příkazu *revgeneticcode* a převedením na buňkové pole uložen genetický kód do lokální proměnné. Následně cykly *for* funkce počítá absolutní četnosti kodonů v sekvenci. Další cyklus *for* přepočítává hodnoty četností kodonů a aminokyselin na hodnoty homozygotnosti. Pro ilustraci hodnot homozygotnosti je vložena následující tabulka, která obsahuje hodnotu homozygotnosti pro gen ND1 *Balaeny mysticetus*.

Tab. 3.3: Hodnoty homozygotnosti kodonů pro gen ND1 *Balaeny mysticetus*

Aminokyselina	F
Alanin	0,3954
Arginin	0,7500
Asparagin	0,5636
Asparagová kyselina	0,5000
Cystein	0,0000
Glutamin	1,0000
Glutamová kyselina	1,0000
Glycin	0,2121
Histidin	0,3333
Isoleucin	0,5467
Leucin	0,3201
Lysin	1,0000
Methionin	0,5439
Fenylalanin	0,7429
Prolin	0,3290
Serin	0,3381
Threonin	0,4418
Tryptofan	0,6111
Tyrosin	0,4545
Valin	0,8182

Z výsledků v tabulce lze odvodit, že nejvýraznější kodonová preference je u glutaminu, kyseliny glutamové a lysinu. Nejrovnoměrnější rozložení synonymních kodonů má cystein.

Funkce pro počítání v okně

Tato funkce má opět tři vstupy a dva výstupy. Tři vstupy jsou stejné jako u ostatních funkcí s plovoucím oknem, a to sekvence, okno a překryv. Výstupem je pak kód a vektor s homozygotnostmi kodonů. Funkce funguje na podobném principu jako funkce pro relativní četnost synonymních kodonů. Obsahuje *while* cyklus, ve kterém jsou vloženy tři *for* cykly. V nich se počítají jak četnosti kodonů, tak i četnosti aminokyselin. Po ukončení cyklu *while* se hodnoty v dalších cyklech přepočítávají na hodnoty homozygotností kodonů.

3.5 Funkce pro efektivní počet kodonů

Tato funkce používá, jak už bylo řečeno, hodnoty homozygotnosti kodonů. Funkce s názvem *ef_poc_kod* má jeden vstup a dva výstupy. Vstupem je počítaná sekvence. První výstup představuje vektor hodnot pro jednotlivé aminokyseliny. Druhým výstupem se stává hodnota efektivního počtu kodonů pro celý gen.

Zprvu funkce volá funkci pro homozygotnost kodonů. Poté prochází vektor homozygotností a počítá jejich převrácenou hodnotu a ukládá do vektoru, pokud je hodnota homozygotnosti nulová, pak efektivní počet kodonů je roven stupni degenerace genetického kódu pro danou aminokyselinu. Stupeň degenerace se zjišťuje pomocí matlabovské funkce *revgeneticcode*.

Poté, co funkce spočítá efektivní počet kodonů pro jednotlivé aminokyseliny, počítá hodnoty pro celý gen. Pomocí vzorců uvedených v kapitole 3 spočítá hodnotu pro danou sekvenci. Jen místo hodnot homozygotností počítá s převrácenou hodnotou efektivních počtů kodonů jednotlivých aminokyselin, aby u krátkých sekvencí nevycházely hodnoty vyšší než 61.

Pro gen NADH dehydrogenázu 1 *Balaeny mysticetus* byly spočítány tyto hodnoty efektivního počtu kodonů jak pro celý gen, tak i pro jednotlivé aminokyseliny (viz tab. 3.4).

N=38,0316

V genu se aktivně používá pouze 38 kodonů z 61.

Tab. 3.4: Hodnoty efektivních počtů kodonů pro gen ND1 *Balaeny mysticetus*

Aminokyselina	N _c	Stupeň degenerace
Alanin	2,5291	4
Arginin	1,3333	4
Asparagin	1,7742	2
Asparagová kyselina	2,0000	2
Cystein	2,0000	2
Glutamin	1,0000	2
Glutamová kyselina	1,0000	2
Glycin	4,0000	4
Histidin	2,0000	2
Isoleucin	1,8293	2
Leucin	3,1233	6
Lysin	1,0000	2
Methionin	1,8387	2
Fenylalanin	1,3462	2
Prolin	3,0395	4
Serin	2,9577	6
Threonin	2,2635	4
Tryptofan	1,6364	2
Tyrosin	2,0000	2
Valin	1,2222	4

Pokud se porovnají hodnoty efektivního počtu kodonů se stupni degenerace pro danou aminokyselinu, dá se u některých aminokyselin pozorovat jistá kodonová preference a u některých (například tyrosin) není kodonová preference vůbec patrná. Čím je hodnota pro efektivní počet kodonů bližší stupni degenerace, tím je menší kodonová preference. Největší kodonová preference se vyskytuje, pokud je efektivní počet kodonů roven jedné.

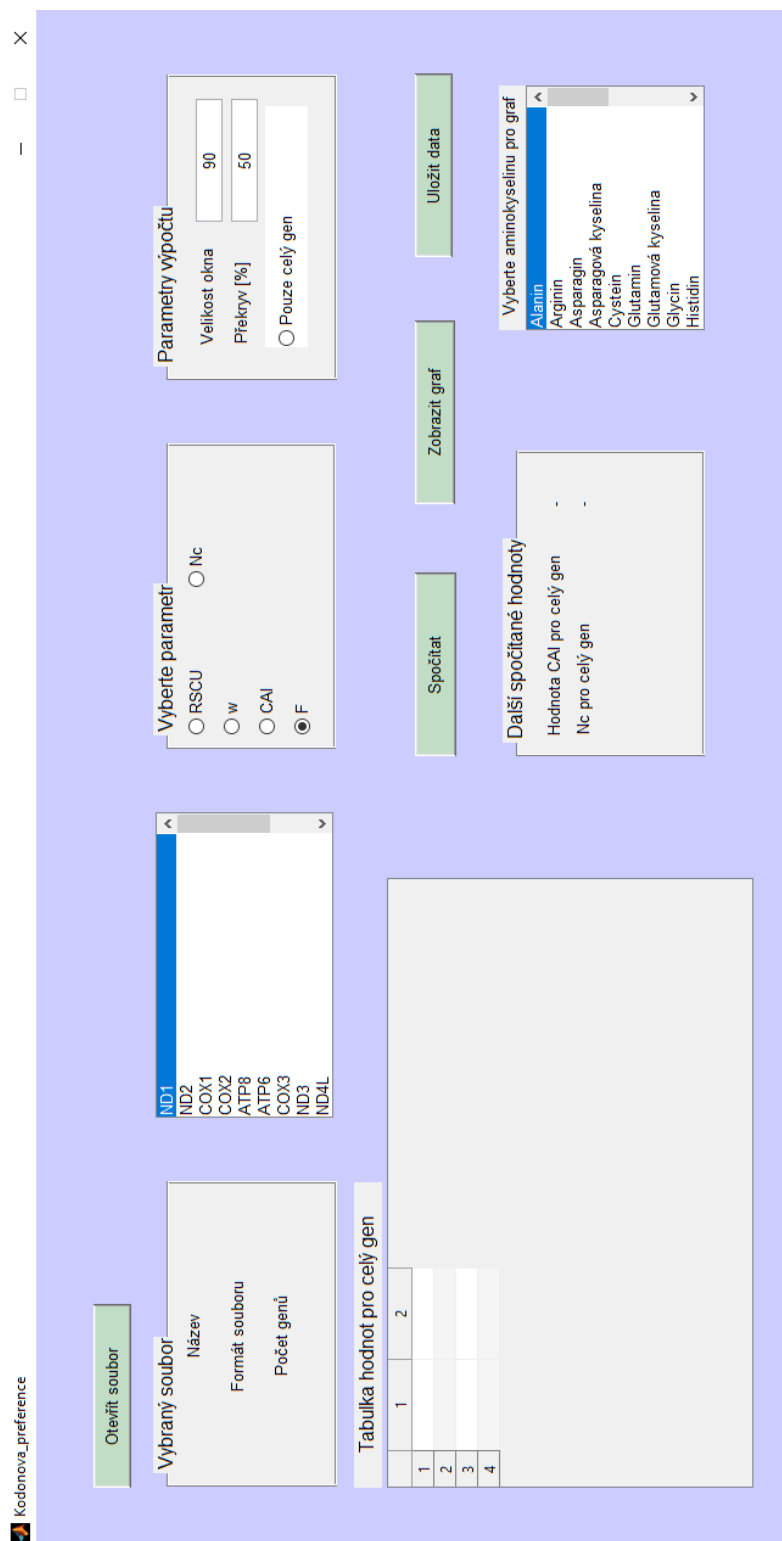
4 TVORBA APLIKACE

Aplikace byla vytvářena opět v programovém prostředí MATLAB 2014a. Do aplikace bylo zahrnuto počítání parametrů v plovoucím okně s nastavitelnou velikostí a nastavitelným překryvem. Aplikace byla tvořena s grafickým uživatelským rozhraním a slouží výhradně pro analýzu mitochondriální DNA. Po možné úpravě aplikace by bylo možno dosáhnout i analýzy genů jaderné DNA.

4.1 Struktura aplikace

Aplikace je vybavena grafickým uživatelským rozhraním. Rozhraní by mělo být jednoduché a předvídatelné. V levém horním rohu lze najít tlačítko *Otevřít*, které se při stisknutí uživatele zeptá, jaký soubor si přeje otevřít. Aplikace je vybavena šesti panely uspořádanými do dvou řad. V každém z panelů je třeba nastavit hodnoty nebo vybrat parametry výpočtu. V druhém řádku se kromě dvou panelů nachází také tabulka, ve níž se zobrazují některé výsledky.

Pro spuštění výpočtu, zobrazení grafu a uložení dat slouží tlačítka uprostřed. Jsou správně označena, aby uživatel snadno rozeznal, ke kterému účelu tlačítko slouží. Každý panel je popsán, tak aby uživatel věděl, co a v jakém panelu vybírá. Vše je dostatečně uspořádané pro snadnou orientaci. Na následujícím obrázku (viz obr. 4.1) je demonstrováno, jak vypadá aplikace po otevření.



Obr. 4.1: Aplikace Kodonová preference po spuštění

4.2 Popis fungování aplikace

Aplikace s názvem Kodonová preference umožňuje uživateli otevřít soubor formátu fasta. Program fasta soubor zpracuje – oddělí hlavičku od sekvence a vyhledá kódující sekvence, které uloží do globální proměnné CDS. Z hlavičky vezme název organismu, kterému sekvence patří, a zobrazí jej. Také zobrazí počet genů a formát souboru, se kterým pracuje.

Po načtení souboru je možné vybrat si analyzovaný gen a parametr, který je cílem spočítat. V neposlední řadě se dá nastavit velikost plovoucího okna a překryv, pokud je třeba počítat gen po částech, nebo zaškrtnout políčko pro spočítání parametru pro celý gen.

Pokud uživatel zaškrtně políčko pro celý gen, pak se výsledek zobrazí v tabulce pro parametry relativní četnosti synonymních kodonů, relativního adaptačního indexu a homozygotnosti kodonů. Parametr kodonový adaptační index pro celý gen se objeví v panelu s názvem další spočítané parametry.

Pro samotný výpočet musí uživatel kliknout na nastavení Spočítat. Toto tlačítko analyzuje, zda je zaškrtnuto políčko pro celý gen, pokud ano, provede výpočet funkcemi popsané v kapitole 4 a zobrazí je v tabulce či ve vedlejším panelu. Na následujícím obrázku (viz obr. 4.2) lze vidět tabulku s hodnotami pro relativní četnost synonymních kodonů pro gen ND1 *Balaeny mysticetus* pro aminokyselinu alanin.

Pokud políčko pro výpočet celého genu zaškrtnuté není, volá funkci *RSCUok*, *wok*, *CAIok*, nebo *F_parok*, která má velikost okna, překryv a sekvenci jako své vstupy. Ty funkcí *get* jsou získány z editovatelných textových polí, které uživatel vyplnil před stisknutím tlačítka. Výstupem funkcí *RSCUok*, *wok* a *F_parok* jsou dvě buňková pole. Jedno z nich obsahuje genetický kód, který je uložen do globální proměnné. Dalším výstupem je buňkové pole obsahující spočítaný parametr. Pouze u funkce *CAIok* je výstupem vektor hodnot kodonového adaptačního indexu a buňkové pole kódu.

Kodonova_preference

Otevřít soubor

Vybraný soubor

Balaena mysticetus

Formát souboru: Fasta

Počet genů: 13

ND1

ND2

COX1

COX2

ATP8

ATP6

COX3

ND3

ND4L

Vyberte parametr

☒ RSCU
☐ w
☐ CAI
☐ F

☒ Nc

Parametry výpočtu

Velikost okna

90

Překryv [%]

50

☒ Pouze celý gen

Uložit data

Zobrazit graf

Spočítat

Další spočítané hodnoty

Hodnota CAI pro celý gen

-

Nc pro celý gen

-

Vyberte aminokyselinu pro graf

Alanin

Arginin

Asparagin

Asparagová kyselina

Cystein

Glutamin

Glutamová kyselina

Glycin

Histidin

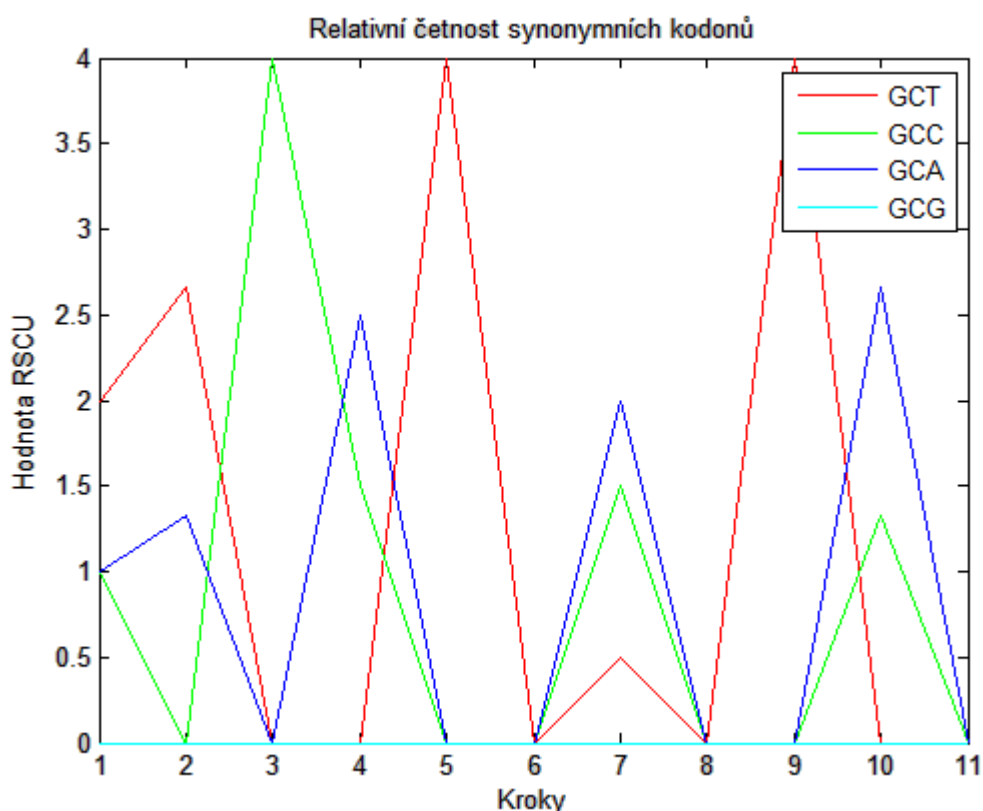
Tabulka hodnot pro celý gen

	1	2	3	4	5	6
1	0.4000	1.8667	1.7333	0	0	0
2	0	0.5000	3.5000	0	0	0
3	0.5455	1.4545	0	0	0	0
4	1.5000	0.5000	0	0	0	0
5	0	2	0	0	0	0
6	2	0	0	0	0	0
7	2	0	0	0	0	0
8	0.6667	1.3333	1.3333	0.6667	0	0
9	1	1	0	0	0	0
10	0.6400	1.3600	0	0	0	0
11	0.7500	0	0.4286	1.0714	3.1071	0.6429
12	2	0	0	0	0	0
13	1.3684	0.6316	0	0	0	0
14	0.2857	1.7143	0	0	0	0
15	0.6061	1.8182	1.2727	0	0	0

Obr. 4.2: Vzor výpočtu RSCU pro celý gen

Po provedení výpočtu se zobrazí graf. U parametrů, které mají hodnoty rozdílné pro různé aminokyseliny, je možné dodatečně vybrat jinou aminokyselinu a zobrazit další graf pomocí tlačítka. Poté, co kýženou aminokyselinu vybere, klikne na tlačítko pro zobrazení grafu. Graf se zobrazí a pro hodnoty RSCU a relativního adaptačního indexu se mu zobrazí křivky pro všechny synonymní kodony pro danou aminokyselinu.

Pro hodnotu homozygotnosti je zobrazena jedna křivka pro jednu aminokyselinu. Kodonový adaptační index má jednu křivku pro celý gen. V grafu si lze všimnout, že na ose y je zobrazen počítaný parametr a na ose x lze sledovat, v jaké části genu se nachází. Na následujícím obrázku (viz obr. 4.3) si lze všimnout, jak vypadá graf pro relativní četnost synonymních kodonů pro gen ND1 *Balaeny mysticetus* pro aminokyselinu alanin. Parametr byl počítán v okně o velikosti 180 kodonů.



Obr. 4.3: Vzor grafu pro relativní četnost synonymních kodonů

Další funkcionalita, kterou má uživatel možnost využít, je uložení dat do souboru xls. Je to soubor, jenž se dá spustit v tabulkovém procesoru *Excel*. Pro parametry relativní četnosti synonymních kodonů a relativního adaptačního indexu jsou data uložena tak, že v každém řádku jsou hodnoty pro právě jeden kodon, celkem tedy 69 řádků. Data pro homozygotnost kodonů jsou uložena do jednoho

řádku, každý sloupec reprezentuje jednu aminokyselinu. Kodonový adaptační index je rovněž uložen do řádkového vektoru.

Pro uložení dat stačí zmáčknout tlačítko s názvem Uložit data. Toto tlačítko upraví globální proměnné pro uložení tak, aby je tabulkový procesor snadno zpracoval a poté se matlabovskou funkcí *xlswrite* uloží do souboru.

5 ANALÝZA KODONOVÉ PREFERENCE

V této kapitole je provedena analýza kodonové preference za použití parametru efektivní počet kodonů, který je popsán v kapitole 2, CAI a RSCU. Pomocí tohoto parametru je možné určit úroveň kodonové preference u jednotlivých genů.

V aplikaci stačí zaškrtnout políčko pro výpočet Nc. Aplikace spočte, jak efektivní počet kodonů pro jednotlivé aminokyseliny, tak pro celý gen.

Kodonová preference je počítána u všech genů mitochondriálního genomu obratlovců a jednoho členovce. Genetický kód členovců se liší od mitochondriálního genetického kódu obratlovců. Proto byla funkce pro výpočet efektivního počtu kodonů upravena tak, aby správně spočítal hodnoty efektivního počtu kodonů u organismu *Bombyx mori* (bourec morušový). Genetický kód v programovém prostředí Matlab2014a pro genetický kód mitochondrií bezobratlých je označen číslem pět.

Mezi počítanými obratlovci je *Homo sapiens*, člověk. Dalším obratlovcem je nártoun západosundský, *Tarsius bancanus*, který patří mezi primáty. Posledním zkoumaným organismem je *Antigonia capros*, zástupce ryb čeledi ostnoploutví.

5.1 Výpočet efektivního počtu kodonů pro celý gen

Postup výpočtu je uveden v kapitole 2. Tento postup byl použit i při vytváření funkce popsané v kapitole 3. V následující tabulce (viz tab 5.1) jsou výsledky výpočtu efektivního počtu kodonů pro různé organismy. Je možné si všimnout velkého rozdílu úrovně kodonové preference u obratlovců a u členovců.

Tab 5.1: Hodnoty efektivního počtu kodonů různých organismů

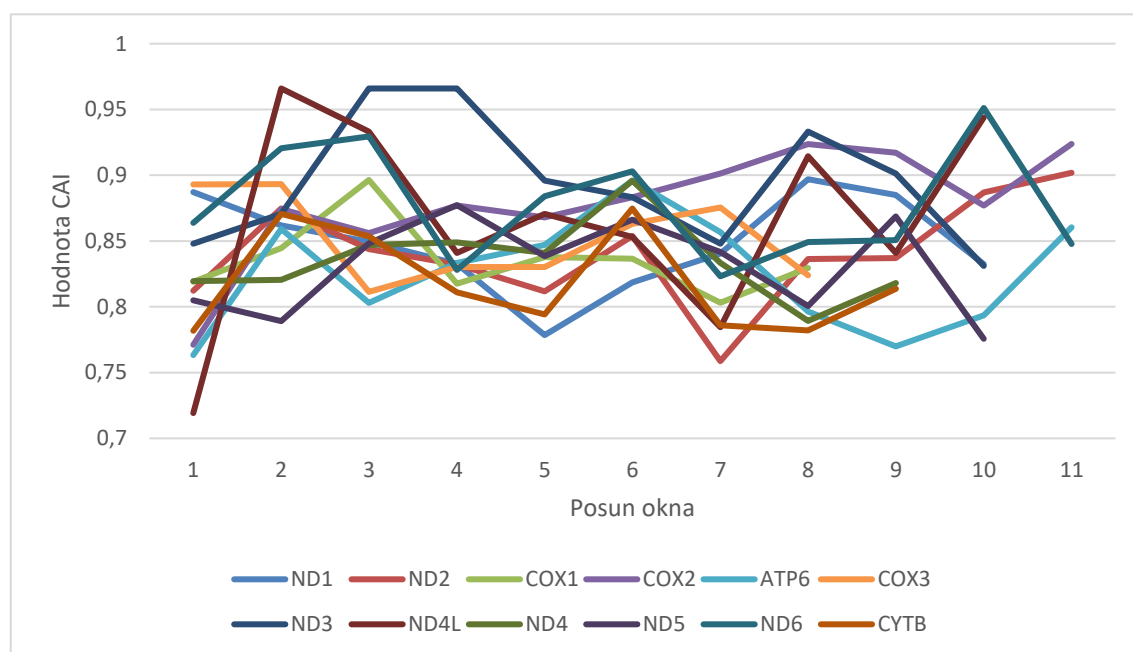
Geny	<i>Homo sapiens</i>	<i>Tarsius bancanus</i>	<i>Antigonia capros</i>	<i>Bombyx mori</i>	Délka genu [bp]
ND1	36,0	40,8	46,1	27,3	955
ND2	44,0	42,4	39,2	26,4	1041
COX1	43,1	42,1	47,3	28,4	1541
COX2	47,3	46,3	46,0	29,6	683
ATP8	53,9	45,6	48,0	39,2	206
ATP6	41,9	41,7	48,9	29,0	680
COX3	43,0	41,4	39,9	27,4	783
ND3	39,8	45,6	44,2	28,2	345
ND4L	46,4	35,5	42,2	31,1	296
ND4	38,3	41,0	46,3	30,2	1377
ND5	41,5	40,3	46,3	25,6	1811
ND6	38,5	40,5	41,3	27,1	524
CYTB	36,9	35,6	43,5	29,3	1140

Na první pohled je patrné, že kodonová preference je výraznější u bource morušového. V tabulce si lze všimnout i rozdílné úrovně kodonové preference mezi

geny. Zajímavá je podobnost mezi *Homo sapiens* a *Tarsius bancanus* v efektivním počtu kodonů u genu cytochromu B.

U genu ATP syntázy podjednotky 8 je dle efektivního počtu kodonů možné odvodit, že je u všech organismus nejmenší kodonová preference, to je dáno zejména délkou sekvence a tím, že se v sekvenci nevyskytují všechny aminokyseliny, to pak ovlivňuje výsledné skóre efektivního počtu kodonů. Pokud aminokyselina v sekvenci chybí, pak je efektivní počet kodonů počítán jako celkový počet synonymních kodonů pro danou aminokyselinu. Jelikož je výsledek pro ATP syntázu podjednotku 8 nereprezentativní v důsledku malého počtu kodonů, je tento gen vyloučen z další analýzy.

Další analýza je provedena pomocí kodonového adaptačního indexu. Kodonový adaptační index je počítán v okně o velikosti odpovídající přibližně jedné pětině počítané sekvence, velikost okna musí být po vynásobení setinou překryvu dělitelná třemi. V následujícím grafu je možné si všimnout, jak se kodonová preference mění v různých částech jednotlivých genů.



Obr. 5.1: Kodonový adaptační index mtDNA *Homo sapiens* počítaný v okně

Dle výsledků zřejmých z grafu (viz graf 5.1) se kodonová preference nemění v částech genu nijak výrazně. Hodnota kodonového adaptačního indexu se pohybuje v rozmezí 0,7 do 1, což značí jistou kodonovou preferenci. Zároveň si lze všimnout, že hodnoty indexu na začátku genů nejsou vyšší než v ostatních částech genu, takže nebyl potvrzen předpoklad, že je kodonová preference nejvýraznější na začátku genu.

5.2 Porovnání kodonové preference u aminokyselin

Pro porovnání kodonové preference u aminokyselin jednotlivých genů mitochondriálního genomu *Homo sapiens* bylo využito parametru relativní četnosti synonymních kodonů. Pro každý gen byla spočítána tabulka RSCU. Všechny výsledky jsou vloženy v přílohách na konci práce.

Zprvu bych se ráda zabývala aminokyselinami alaninem a argininem. U těchto aminokyselin se vyskytuje zajímavá shoda u kodonové preference všech genů, až na jeden. Ve všech tabulkách (kromě tabulky B.11) si lze všimnout toho, že kodony GCG a CGG mají hodnotu RSCU buď rovnu nule, nebo se hodnota nule blíží (viz tab. B.1 až 10, tab. B.12). Pouze u tabulky B.11 nepoužívané kodony jsou jiné. Mezi nepoužívané kodony u genu ND6 patří GCC a CGC. Tento rozdíl lze pozorovat i u jiných aminokyselin.

Dalšími aminokyselinami jsou asparagin a kyselina asparagová. Ačkoliv obě tyto aminokyseliny jsou kódovány pouze dvěma synonymními kodony, i zde se vyskytuje kodonová preference. U těchto aminokyselin je možné u všech genů kromě jednoho pozorovat, že kodony AAC (asparagin) a GAC (kyselina asparagová) jsou preferovány, a naopak AAT (asparagin) a GAT (kyselina asparagová) jsou téměř nepoužívány (viz tab. B.1 až 10, tab. B.12). Je zajímavé, že opět gen ND6 z tohoto pravidla vyčnívá a u něj je to přesně naopak (viz tab. B.11).

Aminokyselina cystein je kódována dvěma synonymními kodony: TGT a TGC. Pokud byly v genech kodony této aminokyseliny obsaženy, potom byl preferován kodon TGC. Kodon TGT je preferován pouze v genu NADH dehydrogenázy podjednotky 6 (viz tab. B.11). V některých genech se tyto kodony vůbec nevyskytovaly (např: NAHD dehydrogenáza podjednotka 1).

Aminokyselina glutamin je také kódována právě dvěma synonymními kodony: CAA a CAG. Tyto synonymní kodony nejsou používány stejně často. Kodon CAA je používán výrazně častěji než kodon CAG. U některých aminokyselin je používán výhradně kodon CAA pro kódování glutaminu. Mezi takové geny patří například ND1 nebo ATP6 (viz tab. B.1, tab. B.5). U genu ND6 je častějším kodonem pro kódování glutaminu CAG (viz tab. B.11).

Kyselina glutamová je aminokyselina se stupněm degenerace 2. Synonymní kodony používané pro tuto aminokyselinu jsou: GAA a GAG. Téměř všechny geny používají kodon GAA. Tento výsledek ovšem neplatí pro gen NADH dehydrogenázu podjednotku 6. Pro tento gen je preferován kodon GAG (viz tab. B.11).

Glycin patří mezi aminokyseliny, které jsou kódovány čtyřmi synonymními kodony: GGT, GGC, GGA a GGG. Při pohledu do tabulek pro jednotlivé geny je možné si všimnout jisté podobnosti v používání těchto kodonů. U většiny genů patří mezi nejčastěji používané kodon GGC. Méně používaný, ale přesto více než ostatní, je vyjma tří genů i GCA. V genu COX3 je nejčastěji používán kodon GGC, ovšem GGA a

GGG mají stejné hodnoty RSCU (viz tab. B.6). V genu ND1 je nejpoužívanější kodon rovněž GGC, ale jako druhý v pořadí je GGG (viz tab. B.1). Nejzajímavěji se opět projevil v hodnotách gen ND6, který obsahuje jako nejpoužívanější kodony GGT a GGG neboli kodony, které ostatní geny příliš nepoužívají (viz tab. B.11).

Obě dvě aminokyseliny, jak histidin, tak isoleucin, jsou kódovány dvěma synonymními kodony. I v tomto případě platí, že každá z těchto aminokyselin má preferovaný kodon. U histidinu je to zpravidla kodon CAC. Výjimkou je ND4, kde je preferovaný kodon CAT (viz tab. B.9). U genů, jako jsou ND3 a ND6 se histidin vůbec nevyskytuje. Isoleucin preferuje zpravidla kodon ATC. U genů, jako je ND3, COX3, ND4 a ND6 (viz tab. B.6, tab. B.7, tab. B.9 a tab. B.11) je to jinak. V genu pro COX3 je používání obou kodonů stejně časté. U genů pro ND3, ND4 a ND6 je nejčastějším kodonem ATT.

Aminokyselina leucin je jedna z mála, která je kódována 6 synonymními kodony: TTA, TTG, CTT, CTC, CTA a CTG. I u této aminokyseliny je možné pozorovat kodonovou preferenci napříč geny. Nejpoužívanější kodon je CTA, a to téměř ve všech počítaných genech. Výjimku tvoří pouze gen ND6, který má nejpoužívanější kodony TTA a TTG. Druhým nejpoužívanějším kodonem u všech genů vyjma ND6, COX2 a ND3 je kodon CTC. U genu ND3 je to TTA (viz tab. B.7) a gen COX2 používá častěji CTT (viz tab. B.4).

Aminokyselin lysin a methionin jsou aminokyseliny se stupněm degenerace dva. Lysin je kódován kodony AAA a AAG, methionin pak ATA a ATG. Obdobně, jak tomu bylo u předešlých aminokyselin je vždy jeden z nich preferován, a to u všech genů stejně, kromě genu ND6 (viz tab. B.11). Mezi preferovaný kodon lysinu patří kodon AAA a u methioninu ATA. V genu pro ND6 jsou oba kodony kódující lysin používány stejně a u methioninu je preferován kodon ATG (viz tab. B.11).

Fenylalanin je kódován dvěma synonymními kodony TTT a TTC. V každém z genů je preferován jiný z nich. Kodon TTT je preferován v genech ND1, COX3, ND4L a ND6 (viz tab. B.1, tab. B.6, tab. B.8, tab. B.11). Kodon TTC pak v genech ostatních (viz tab. B.2 až 5, tab. B.7, tab. B.9, tab. B.10, tab. B.12).

Další aminokyselinou je prolin, který je kódován čtyřmi kodony: CCT, CCC, CCA a CCG. Ve většině genů kromě ND4L, ND6 a CYTB je nejčastějším kodonem CCC. V genu ND6 je to ovšem CCT (viz tab. B.11). V genu ND4L je tomu taktéž jinak. V tomto genu jsou používány dva kodony stejně často – CCC a CCT (viz tab. B.8). U cytochromu B je tomu zase jinak, v tomto genu jsou používány stejně kodony CCC a CCA (viz tab. B.12).

Serin je další aminokyselina se stupněm degenerace šest. Kodony, které kódují serin, jsou: TCT, TCC, TCA, TCG, AGT a AGC. Tato aminokyselina se vyznačuje jistou kodonovou preferencí. Mezi nejpoužívanější kodony u většiny genů patří TCA a TCC. Výjimky tvoří geny ND6, ATP6 a COX1. Gen ND6 preferuje kodony TCT a AGT

(viz tab. B.11). Gen ATP6 pak kodony TCT a TCA (viz tab. B.5) a COX1 TCT a TCC (viz tab. B.3).

Threonin je kódován třemi kodony: ACT, ACC, ACA a ACG. Většina genů preferuje zejména kodony ACC kromě genů ND4, COX1 a COX3, které preferují kodon ACA (viz tab. B.3, tab. B.6 a tab. B.9). V genu ND6 lze nalézt shodu v počítání jednotlivých kodonů ACA a ACT, což jsou preferované kodony, jiné kodony kódující threonin v daném genu nenajdeme (viz tab. B.11).

Tyrosin a tryptofan jsou aminokyseliny se stupněm degenerace dva. Tryptofan je kódován TGA a TGG, tyrosin pak TAT a TAC. Ve všech genech je preferovaný kodon TGA pro tryptofan a TAC pro tyrosin. Jedinou výjimkou je opět gen ND6, který preferuje kodon TAT před TAC pro tyrosin. V kódování tryptofanu však v genu ND6 změna není. V genu ND4L se kodony pro tryptofan vůbec nevyskytují a kodony pro tyrosin jsou používány rovnoměrně (viz tab. B.8).

Poslední aminokyselinou je valin. Tato aminokyselina je kódována čtyřmi kodony: GTT, GTC, GTA a GTG. V používání této aminokyseliny není příliš patrná kodonová preference. Nejčastěji jsou v genech používány GTA, GTC a GTT. U genu ND6 patří k častějším i GTG (viz tab. B.11).

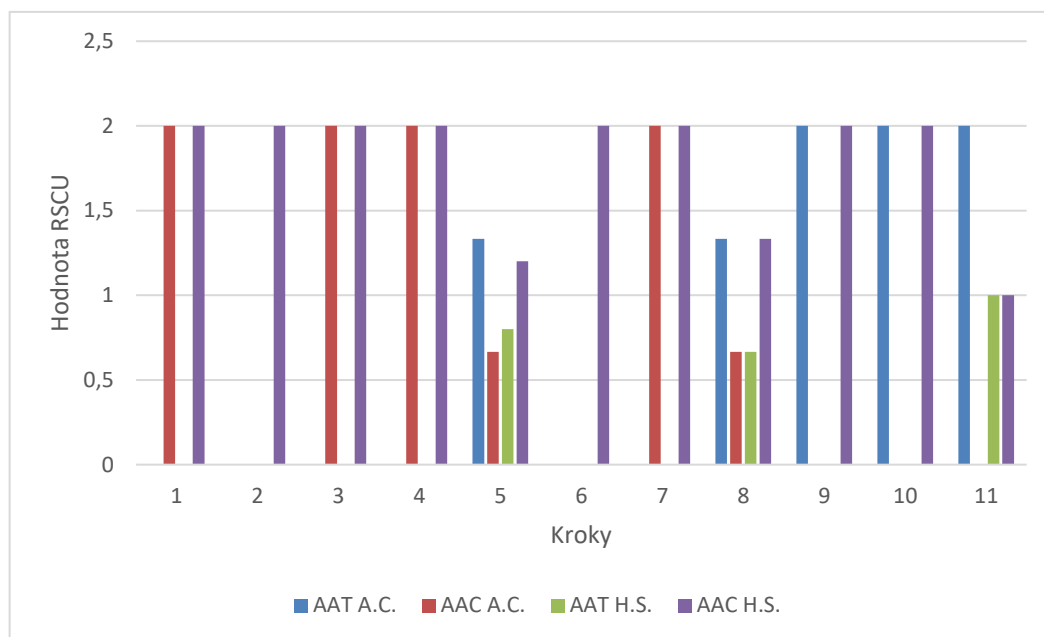
Celkově tak můžeme říci, že většina aminokyselin má preferované kodony a jsou to kodony stejné pro všechny geny mitochondriálního genomu *Homo sapiens*, kromě genu ND6.

5.3 Porovnání genu ND1 a ND6 u obratlovců

Pokud budeme vycházet z výsledků z přechozí podkapitoly, pak nejzajímavější gen byl bez pochyby gen ND6. Výsledky hodnot relativní četnosti synonymních kodonů vycházely odlišně než u ostatních genů.

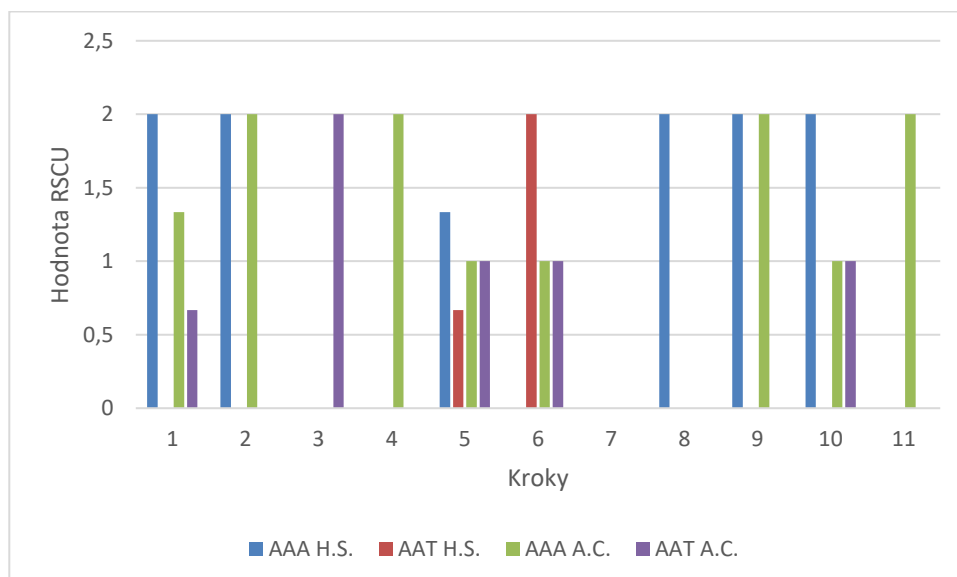
Další analýzou budou hodnoty RSCU pro různé aminokyseliny počítané v okně. Hodnoty tohoto parametru budou počítány rovněž pro organismus *Homo sapiens*. Analýzu budeme provádět pro zjištění, zda se mění preference jednotlivých kodonů pro danou aminokyselinu v průběhu genu, nebo zůstává preference stejná.

Pro porovnání s jiným obratlovcem byla zvolena *Antigonia capros*. Pro porovnání byly zvoleny aminokyseliny: asparagin, lysin a methionin. Analýza bude pouze orientační z důvodu menšího množství dat.



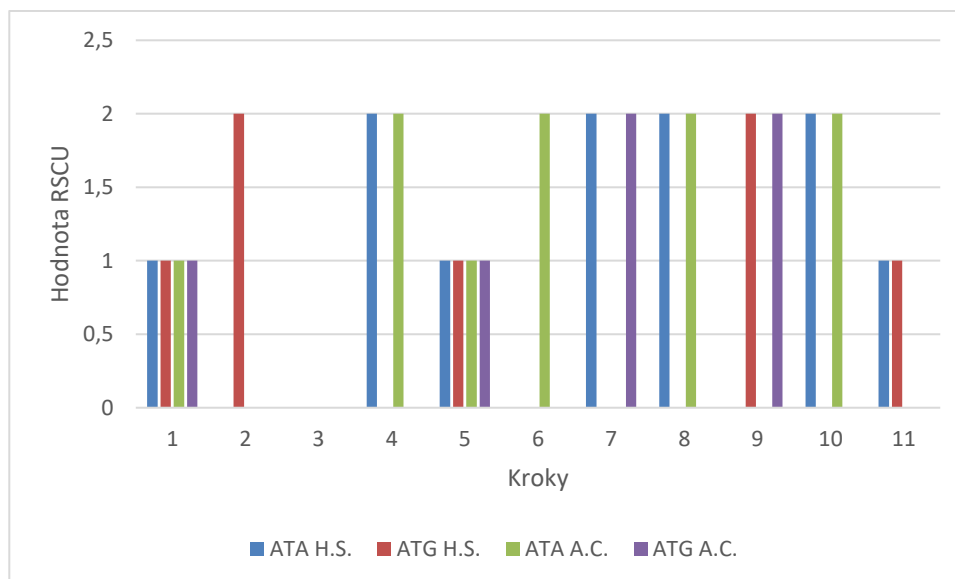
Obr. 5.2: Porovnání hodnot RSCU v okně pro asparagin genu ND1 *Homo sapiens* s *Antigonie capros*

V grafu 5.2 si můžeme všimnout, že kodonová preference u *Antigonie capros* se téměř nemění a pro asparagin je preferovaný kodon AAC v sekvenci téměř stejně. Rozdíl však zaznamenáme u *Homo sapiens*, kde na začátku sekvence je preferovaný také kodon AAC a ke konci sekvence kodon AAT.



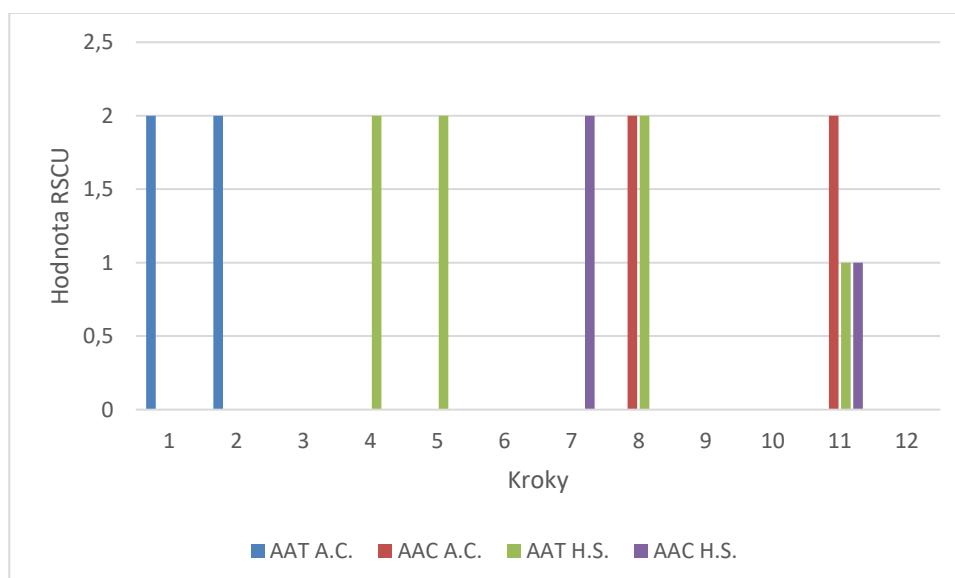
Obr. 5.3: Porovnání hodnot RSCU v okně pro lysin genu ND1 *Homo sapiens* a *Antigonie capros*

V grafu 5.3 vidíme, že preferovaný kodon u obou organismů je AAA, nejvíce pak na začátku a na konci sekvence.



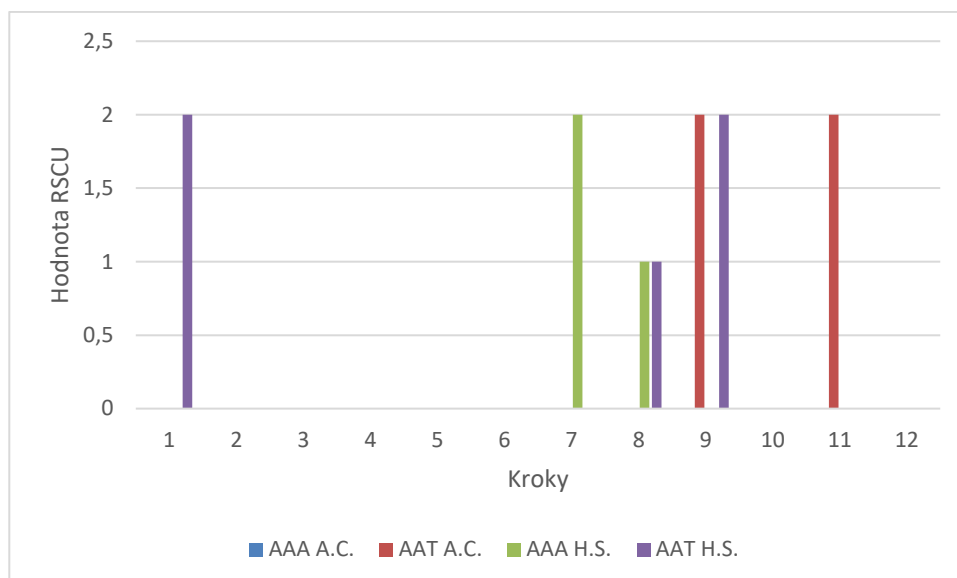
Obr. 5.4: Porovnání hodnot RSCU v okně pro methionin genu ND1 *Homo sapiens* a *Antigonie capros*

V grafu 5.4 si můžeme všimnout, že se kodonová preference výrazně mění v částech sekvence.



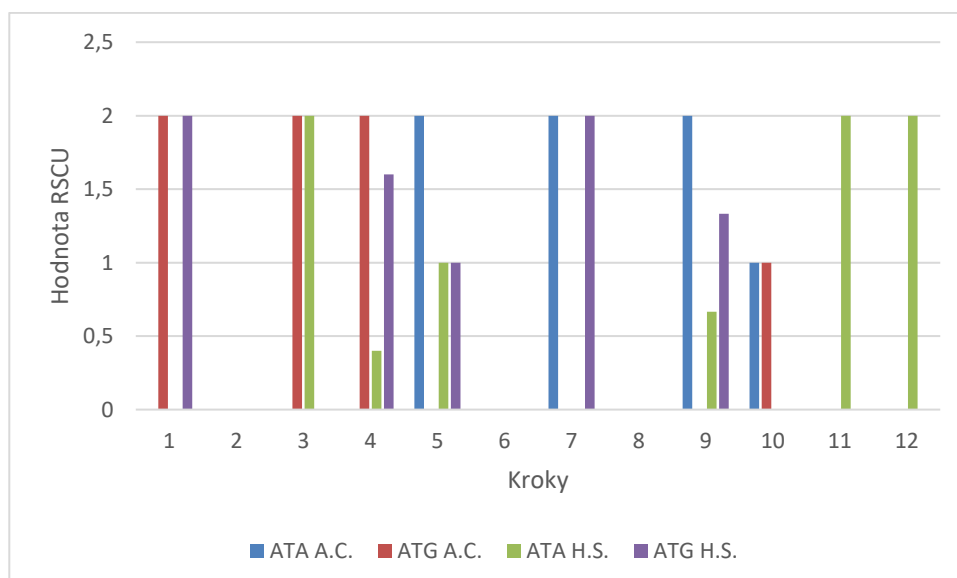
Obr. 5.5: Porovnání hodnot RSCU v okně pro asparagin genu ND6 *Homo sapiens* a *Antigonie capros*

Již od prvního pohledu na graf 5.5 je poznat, že se v genu aminokyselina asparagin příliš nevyskytuje. V první polovině sekvence se vyskytuje kodon AAT u obou organismů a v druhé polovině je preferovanější kodon AAC.



Obr. 5.6: Porovnání hodnot RSCU v okně pro lysin genu ND6 *Homo sapiens* a *Antigonie capros*

Je možné si všimnout v grafu 5.6, že lysin není nejčastější aminokyselina v genu ND6. V první polovině se téměř nevyskytuje. Z většiny je však u obou organismů preferován kodon AAT.



Obr. 5.7: Porovnání hodnot RSCU v okně pro methionin genu ND6 *Homo sapiens* a *Antigonie capros*

V grafu 5.7 je možné vidět, že u obou organismů se kodonová preference mění, a to zcela odlišně. U organismu *Homo sapiens* se preferuje kodon ATG v začátku sekvence, poté ho však vystřídá kodon ATA. U *Antigonie capros* se mění zcela nezávisle na části sekvence.

U obou organismů bylo na základě uvedených grafů zjištěno, že gen ND6 je u *Homo sapiens* a *Antigone capros* odlišný z hlediska kodonové preference oproti genu ND1. Jelikož je závěr stanoven na základě malého množství dat, není tento závěr signifikantní. Pokud práce podnítí další výzkum na toto téma, bylo by velice zajímavé otestovat velké množství dat a zjistit, zda se tato kodonová preference vyskytuje napříč všemi geny s výjimkou ND6 i u jiných obratlovců.

6 ZÁVĚR

V této práci jsem se zabývala kodonovou preferencí v mitochondriálním genomu. V práci byly teoreticky rozebrány základní pojmy z oblasti genetiky a bioinformatiky. Dále pak byla představena kodonová preference a parametry kodonovou preferenci charakterizující.

V praktické části byly vytvořeny funkce pro výpočet parametrů kodonové preference, které byly dále využity při tvorbě aplikace, pomocí které by bylo možno spočítat jakýkoliv z parametrů kodonové preference. Tato aplikace spočítá hodnoty parametrů, jak pro celý gen, tak pro jeho části v plovoucím okně. Aplikace využívá uživatelského rozhraní programového prostředí MATLAB 2014a. Aplikace by měla být interaktivní, měla by ukazovat měnící se parametry kodonové preference v rámci částí genu v přehledném grafu. Aplikace má nastavitelné plovoucí okno i překryv. Aplikace má možnost uložení dat do souboru, který je možný otevřít pomocí tabulkového procesoru. Pomocí aplikace byla provedena analýza mitochondriálního genomu *Homo sapiens*.

Analýza byla prováděna pomocí efektivního počtu kodonů, dále pomocí kodonového adaptačního indexu a relativní četnosti synonymních kodonů. Zprvu bylo pomocí efektivního počtu kodonů zjištěno, zda se u *Homo sapiens* kodonová preference vyskytuje.

Poté bylo zjištěno, zdali se kodonová preference mění v různých částech genu pomocí kodonového adaptačního indexu. Výsledek této krátké analýzy ukázal, že se kodonová preference v částech genů výrazně nemění, pokud se soustředíme na gen jako celek.

V následné analýze mitochondriálního genomu *Homo sapiens* pomocí relativní četnosti synonymních kodonů bylo zjištěno, že kodonová preference se vyskytuje napříč všemi geny mitochondriálního genomu. Tato kodonová preference se u většiny genů shodovala, jinými slovy byly preferovány stejné geny ve většině genů. Jediným odlišným genem je gen NADH dehydrogenázy podjednotky 6, v tomto genu se zpravidla vyskytovala kodonová preference jiného kodonu než u genů ostatních.

Výsledek byl otestován i v plovoucím okně a výsledek *Homo sapiens* porovnán s *Antigonií capros*. Většina kodonových preferencí byla velice podobná u obou organismů. Závěr však vychází z menšího množství dat.

Literatura

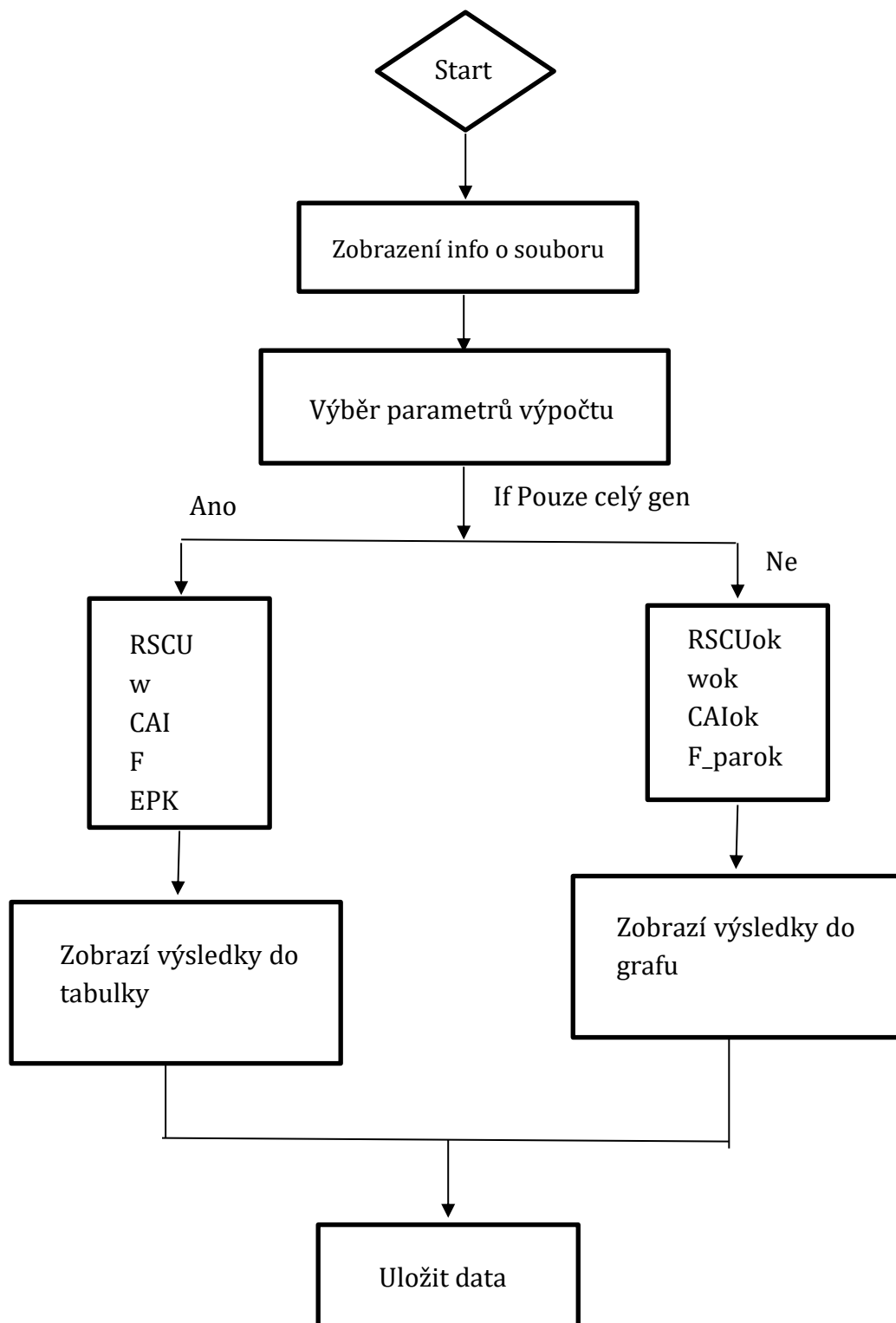
- [1] PLOTKIN, Joshua B. a Jonathan DUSHOFF. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus [online]. *PNAS*. 2003, **100**(12), 7152–7157. [cit. 2017-12-27].
Dostupné z: <http://www.pnas.org/content/100/12/7152.full>
- [2] PLOTKIN, Joshua B., Jonathan DUSHOFF, Michael M. DESAI a Hunter B. FRASER. Codon Usage and Selection on Proteins [online]. *Molecular Evolution*. 2006, **63**(5), 635-653. [cit. 2017-12-27].
Dostupné z: <https://www.ncbi.nlm.nih.gov/pubmed/17043750>
- [3] ST.JOHN, Justin C. *Mitochondrial DNA, Mitochondria, Disease and Stem Cells*. 1.Australia: Springer Science & Business Media, 2012. ISBN 9781627031011.
- [4] Structure of the human mitochondrial genome. In: *Wikipedia* [online]. 2009 [cit. 2018-01-03]. Dostupné z: https://upload.wikimedia.org/wikipedia/commons/thumb/2/24/Mitochondrial_DNA_cs.svg/1267px-Mitochondrial_DNA_cs.svg.png
- [5] PLOTKIN, Joshua B. a Grzegorz KUDLA. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* [online]. 2011, **12**(1), 32-42 [cit. 2017-12-27].
Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074964/>
- [6] KURLAND, C.G. Codon bias and gene expression. *FEBS Letters*. ADONIS, 1991, **283**(2), 165-169.
- [7] LAFAY, Bénédicte, John C. ATHERTON a Paul M. SHARP. Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. *Microbiology* [online]. 2000, **146**(4), 851-860 [cit. 2017-12-27]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pubmed/10784043>
- [8] NABIYOUNI, Maryam. *Mega-scale bioinformatics investigation of codon bias in vertebrates* [online]. Toledo, 2011 [cit. 2017-12-27]. Dostupné z: <http://utdr.utoledo.edu/cgi/viewcontent.cgi?article=1677&context=theses-dissertations>. Diplomová práce. The University of Toledo.

- [9] LEE, Soohyun, Seyeon WEON, Sooncheol LEE a Changwon KANG. Relative Codon Adaptation Index, a Sensitive Measure of Codon Usage Bias. *Evolutionary Bioinformatics* [online]. Daejeon: Libertas Academica, 2010, **5**(6), 47-55 [cit. 2017-12-27].
Dostupné z: <https://www.ncbi.nlm.nih.gov/pubmed/20535230>
- [10] GenBank Overview. *NCBI* [online]. Bethesda: U.S. National Library of Medicine, 2017 [cit. 2018-01-03].
Dostupné z: <https://www.ncbi.nlm.nih.gov/genbank/>
- [11] *What Is MATLAB?* [online]. [cit. 2018-01-03]. Dostupné z: <http://cimss.ssec.wisc.edu/wxwise/class/aos340/spr00/whatismatlab.htm>

Seznam symbolů, veličin a zkratek

RSCU	-	Relativní četnost synonymních kodonů
w	-	Relativní adaptační index
CAI	-	Kodonový adaptační index
F	-	Homozygotnost kodonů
n	-	Absolutní četnost
RF	-	Relativní četnost
Nc	-	Efektivní počet kodonů
ND1	-	NADH dehydrogenáza 1
ND2	-	NADH dehydrogenáza 2
ND3	-	NADH dehydrogenáza 3
ND4	-	NADH dehydrogenáza 4
ND4L	-	NADH dehydrogenáza 4L
ND5	-	NADH dehydrogenáza 5
ND6	-	NADH dehydrogenáza 6
COX1	-	Cytochrom c oxidáza 1
COX2	-	Cytochrom c oxidáza 2
COX3	-	Cytochrom c oxidáza 3
ATP8	-	ATP syntáza 8
ATP6	-	ATP syntáza 6
CYTB	-	Cytochrom b

A Vývový diagram programu



B Výsledky analýzy

Tab. B.1: Výsledky RSCU pro *Homo sapiens* genu ND1

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,444	2,074	1,481	0,000	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	1,143	2,857	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,000	2,000	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,000	2,000	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	0,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	2,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,455	0,545	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,000	2,667	0,333	1,000	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,000	2,000	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,870	1,130	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,476	0,095	0,571	1,810	2,571	0,476
	'AAA'	'AAG'	-	-	-	-
K	1,714	0,286	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,625	0,375	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,500	1,500	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,364	3,091	0,364	0,182	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,571	2,286	2,286	0,000	0,286	0,571
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,114	2,286	1,600	0,000	-	-
	'TGA'	'TGG'	-	-	-	-
W	2,000	0,000	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,714	1,286	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,800	2,000	1,200	0,000	-	-

Tab. B. 2: Výsledky RSCU pro *Homo sapiens* genu ND2

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,800	1,400	1,600	0,200	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	1,000	3,000	0,000	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,600	1,400	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,000	0,000	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	0,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	1,600	0,400	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,667	0,333	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,308	2,154	1,538	0,000	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,500	1,500	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,500	1,500	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,750	0,094	0,656	1,688	2,438	0,375
	'AAA'	'AAG'	-	-	-	-
K	1,667	0,333	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,917	0,083	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	1,067	0,933	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,870	2,087	0,696	0,348	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,857	2,357	1,500	0,214	0,214	0,857
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,465	2,233	1,023	0,279	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,818	0,182	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,400	1,600	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	1,500	1,000	1,500	0,000	-	-

Tab. B.3: Výsledky RSCU pro *Homo sapiens* genu COX1

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	1,100	1,500	1,300	0,100	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	1,000	1,000	2,000	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,471	1,529	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,267	1,733	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	2,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	1,667	0,333	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,400	0,600	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,609	1,478	1,652	0,261	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,444	1,556	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,632	1,368	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,677	0,000	0,484	1,258	3,194	0,387
	'AAA'	'AAG'	-	-	-	-
K	1,800	0,200	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,563	0,438	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,585	1,415	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,690	2,483	0,828	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	1,688	1,875	1,313	0,375	0,000	0,750
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,588	1,294	1,882	0,235	-	-
	'TGA'	'TGG'	-	-	-	-
W	2,000	0,000	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,364	1,636	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,444	1,333	1,889	0,333	-	-

Tab. B.4: Výsledky RSCU pro *Homo sapiens* genu COX2

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	1,143	1,429	1,143	0,286	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	1,333	0,667	2,000	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,571	1,429	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,545	1,455	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,667	1,333	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	1,714	0,286	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,455	0,545	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,727	1,455	1,091	0,727	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,667	1,333	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,636	1,364	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,727	0,182	1,273	1,091	2,182	0,545
	'AAA'	'AAG'	-	-	-	-
K	2,000	0,000	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,600	0,400	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,600	1,400	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,267	2,400	0,800	0,533	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,600	2,400	2,400	0,000	0,600	0,000
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,952	1,524	1,524	0,000	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,500	0,500	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,222	1,778	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,308	2,154	1,538	0,000	-	-

Tab. B.5: Výsledky RSCU pro *Homo sapiens* genu ATP6

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,632	2,526	0,632	0,211	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	2,000	2,000	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,364	1,636	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,000	2,000	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	0,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	2,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	2,000	0,000	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,000	2,500	1,500	0,000	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,000	2,000	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,897	1,103	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,545	0,136	0,682	0,955	2,864	0,818
	'AAA'	'AAG'	-	-	-	-
K	1,667	0,333	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,333	0,667	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,889	1,111	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,857	2,286	0,857	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	1,846	0,923	1,846	0,000	0,000	1,385
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,615	1,846	1,538	0,000	-	-
	'TGA'	'TGG'	-	-	-	-
W	2,000	0,000	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,667	1,333	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	1,000	0,500	2,000	0,500	-	-

Tab. B.6: Výsledky RSCU pro *Homo sapiens* genu COX3

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,267	2,933	0,800	0,000	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	1,600	2,400	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,333	1,667	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	1,333	0,667	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	2,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	2,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,429	0,571	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,211	2,105	0,842	0,842	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,588	1,412	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	1,000	1,000	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,529	0,000	0,529	1,588	2,824	0,529
	'AAA'	'AAG'	-	-	-	-
K	2,000	0,000	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,455	0,545	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	1,130	0,870	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,615	1,846	1,231	0,308	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,571	2,000	2,286	0,000	0,286	0,857
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,333	1,667	1,833	0,167	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,500	0,500	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,727	1,273	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,615	1,231	1,846	0,308	-	-

Tab. B.7: Výsledky RSCU pro *Homo sapiens* genu ND3

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,500	3,500	0,000	0,000	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	4,000	0,000	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,000	2,000	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,667	1,333	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	2,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	2,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,200	0,800	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,000	2,667	1,333	0,000	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,000	0,000	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	1,111	0,889	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	2,143	0,000	0,214	0,857	2,357	0,429
	'AAA'	'AAG'	-	-	-	-
K	2,000	0,000	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,750	0,250	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,500	1,500	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	1,500	1,000	1,500	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,000	4,000	1,000	0,000	1,000	0,000
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,571	2,286	1,143	0,000	-	-
	'TGA'	'TGG'	-	-	-	-
W	2,000	0,000	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,667	1,333	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	1,333	1,333	1,333	0,000	-	-

Tab. B.8: Výsledky RSCU pro *Homo sapiens* genu ND4L

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,444	2,222	0,889	0,444	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	4,000	0,000	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,667	1,333	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,000	2,000	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	2,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	2,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	2,000	0,000	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,000	2,000	2,000	0,000	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,667	1,333	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	1,429	0,571	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,261	0,000	0,261	1,565	3,652	0,261
	'AAA'	'AAG'	-	-	-	-
K	0,000	0,000	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,800	0,200	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	1,333	0,667	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	2,000	2,000	0,000	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,000	3,000	2,250	0,750	0,000	0,000
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,800	2,400	0,800	0,000	-	-
	'TGA'	'TGG'	-	-	-	-
W	0,000	0,000	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	1,000	1,000	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,000	1,333	1,333	1,333	-	-

Tab. B.9: Výsledky RSCU pro *Homo sapiens* genu ND4

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,923	1,846	1,231	0,000	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	2,222	1,778	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,087	1,913	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,000	2,000	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,667	1,333	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	1,800	0,200	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	2,000	0,000	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,235	2,118	0,941	0,706	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,154	1,846	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,821	1,179	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,500	0,063	0,625	1,938	2,625	0,250
	'AAA'	'AAG'	-	-	-	-
K	1,818	0,182	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,778	0,222	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,900	1,100	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,522	2,435	1,043	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,698	2,372	1,395	0,140	0,279	1,116
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,667	1,417	1,833	0,083	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,846	0,154	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,308	1,692	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,000	1,231	2,462	0,308	-	-

Tab. B.10: Výsledky RSCU pro *Homo sapiens* genu ND5

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,455	1,727	1,818	0,000	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,500	2,000	1,500	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,364	1,636	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,545	1,455	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,333	1,667	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	1,700	0,300	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	1,778	0,222	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,615	1,692	1,538	0,154	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,143	1,857	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,667	1,333	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,404	0,115	0,635	1,788	2,596	0,462
	'AAA'	'AAG'	-	-	-	-
K	1,905	0,095	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,769	0,231	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,263	1,737	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,875	2,000	1,125	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,245	2,449	1,592	0,122	0,367	1,224
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,800	1,785	1,354	0,062	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,833	0,167	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,750	1,250	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	1,600	1,067	1,333	0,000	-	-

Tab. B.11: Výsledky RSCU pro *Homo sapiens* genu ND6

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	2,000	0,000	0,500	1,500	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	1,333	0,000	0,000	2,667	-	-
	'AAT'	'AAC'	-	-	-	-
N	1,500	0,500	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	2,000	0,000	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	2,000	0,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	0,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	0,200	1,800	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	1,103	0,000	0,828	2,069	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,000	0,000	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	2,000	0,000	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	2,526	2,526	0,000	0,000	0,000	0,947
	'AAA'	'AAG'	-	-	-	-
K	1,000	1,000	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	0,400	1,600	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	2,000	0,000	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	4,000	0,000	0,000	0,000	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	1,800	0,000	1,200	0,000	1,800	1,200
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	2,000	0,000	2,000	0,000	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,200	0,800	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	2,000	0,000	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	1,161	0,387	1,161	1,290	-	-

Tab. B.12: Výsledky RSCU pro *Homo sapiens* genu CYTB

	GCT'	'GCC'	'GCA'	'GCG'	-	-
A	0,000	2,720	1,280	0,000	-	-
	'CGT'	'CGC'	'CGA'	'CGG'	-	-
R	0,000	1,714	2,286	0,000	-	-
	'AAT'	'AAC'	-	-	-	-
N	0,400	1,600	-	-	-	-
	'GAT'	'GAC'	-	-	-	-
D	0,182	1,818	-	-	-	-
	'TGT'	'TGC'	-	-	-	-
C	0,000	2,000	-	-	-	-
	'CAA'	'CAG'	-	-	-	-
Q	2,000	0,000	-	-	-	-
	'GAA'	'GAG'	-	-	-	-
E	2,000	0,000	-	-	-	-
	'GGT'	'GGC'	'GGA'	'GGG'	-	-
G	0,000	2,000	1,500	0,500	-	-
	'CAT'	'CAC'	-	-	-	-
H	0,333	1,667	-	-	-	-
	'ATT'	'ATC'	-	-	-	-
I	0,526	1,474	-	-	-	-
	'TTA'	'TTG'	'CTT'	'CTC'	'CTA'	'CTG'
L	0,656	0,188	0,656	1,969	2,344	0,188
	'AAA'	'AAG'	-	-	-	-
K	1,778	0,222	-	-	-	-
	'ATA'	'ATG'	-	-	-	-
M	1,600	0,400	-	-	-	-
	'TTT'	'TTC'	-	-	-	-
F	0,583	1,417	-	-	-	-
	'CCT'	'CCC'	'CCA'	'CCG'	-	-
P	0,696	1,565	1,565	0,174	-	-
	'TCT'	'TCC'	'TCA'	'TCG'	'AGT'	'AGC'
S	0,000	2,276	2,897	0,000	0,207	0,621
	'ACT'	'ACC'	'ACA'	'ACG'	-	-
T	0,516	1,677	1,677	0,129	-	-
	'TGA'	'TGG'	-	-	-	-
W	1,833	0,167	-	-	-	-
	'TAT'	'TAC'	-	-	-	-
Y	0,706	1,294	-	-	-	-
	'GTT'	'GTC'	'GTA'	'GTG'	-	-
V	0,400	1,200	2,400	0,000	-	-