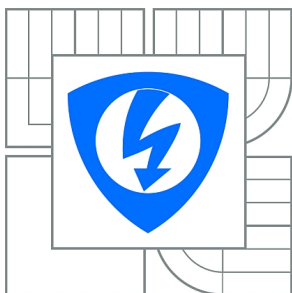


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF BIOMEDICAL ENGINEERING

ANALÝZA VARIABILITY INTRONŮ

ANALYSIS OF INTRON SEQUENCE VARIABILITY

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

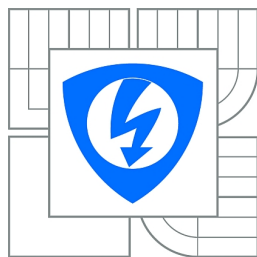
MDDr. HANA KUKAČKOVÁ

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. VLADIMÍRA KUBICOVÁ

BRNO 2013



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav biomedicínského inženýrství

Bakalářská práce

bakalářský studijní obor

Biomedicínská technika a bioinformatika

Studentka: MDDr. Hana Kukačková

ID: 125046

Ročník: 3

Akademický rok: 2012/2013

NÁZEV TÉMATU:

Analýza variability intronů

POKYNY PRO VYPRACOVÁNÍ:

1) Vytvořte literární rešerši o genové expresi, o metodách vyhledávání intronů a exonů u eukaryotických organismů. Popište dosavadní poznatky o konzervativnosti intronů. 2) Navrhněte postup, kterým budete zjišťovat konzervativnost DNA sekvence v intronech a v okolí rozmezí exonů a intronů. 3) Pomocí navrženého postupu proveďte analýzu DNA sekvencí vybraných organismů z různých druhů a v rámci jednoho druhu. Využijte anotované sekvence z NCBI. 4) Proveďte diskusi získaných výsledků.

DOPORUČENÁ LITERATURA:

[1] SNUSTAD, D. Peter a SIMMONS, Michael J.. Genetika. 5. vyd. Překlad Jiřina Relichová. Brno: Masarykova univerzita, 2009, ISBN 978-802-1048-522.

[2] KATAYAMA, Susumu a TOMOSHIGE, Yujiro. Base Selection Method for Recognition of Exon-Intron Boundaries. In: MAMECTIS'08 Proceedings of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems. 2008, p. 283-288. ISSN 1790-2769.

Termín zadání: 11.2.2013

Termín odevzdání: 31.5.2013

Vedoucí práce: Ing. Vladimíra Kubicová

Konzultanti bakalářské práce:

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

Bakalářská práce se v teoretické části zabývá genovou expresí, významem a charakteristikou intronů a metodami jejich rozpoznávání. V praktické části se zaměřuje na analýzu intronů zvolených genů a vyhledávání konzervativních sekvencí .

Nejprve je rozebrána struktura nukleových kyselin a popis jejich chemických a fyzikálních vlastností a genová exprese, její jednotlivé kroky – transkripce, posttranskripční úpravy a translace. Dále se práce zaměřuje na význam intronů, teorie jejich evoluce, podobnosti mezi různými skupinami organismů a některé metody jejich vyhledávání.

Praktická část se zabývá analýzou počátečních a koncových oblastí intronů a ověřování konzervativnosti hraničních dinukleotidových sekvencí, dále analýzou hraničních oblastí intron – exon a exon - intron a dále se snaží navrhnout nové potenciálně konzervativní sekvence – motivy, vyhledat je a ověřit jejich konzervativnost.

Klíčová slova

Nukleová kyselina, DNA, RNA, genová exprese, intron, metody rozpoznávání intronů, konzervativní úseky intronů

Abstract

This bachelor thesis reviews genetic expression, the main characteristics of introns, their importance, and the methodology of their recognition. Practical part focuses on intron variability analysis and conservative sequences searching.

In the beginning, there is a structure of nucleic acids with their chemical and physical characteristics reviewed, follow by gene expression description with all steps – transcription, posttranscriptional splicing and translation. Afterwards, there is an analysis of introns importance, introns evolution theories and similarities among different organisms mentioned, follow by some introns recognition methods.

Practical part is focused on the initial and terminal regions of introns and conservative dinucleotide sequences verifying. The thesis studies intron – exon and exon – intron boundaries. The last part tries to suggest new potentially conservative sequences – patterns, search for them and check their conservativeness.

Key words

Nucleic acid, DNA, RNA, gene expression, intron, intron recognition methods, intron conservative sequences

Bibliografická citace

KUKAČKOVÁ, H. Analýza variability intronů. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2013. 52 s. Vedoucí bakalářské práce Ing. Vladimíra Kubicová.

Prohlášení o původnosti

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně, pod vedením vedoucí bakalářské práce Ing. Vladimíry Kubicové a s použitím uvedených zdrojů.

V Brně dne 29. 5. 2013

.....

MDDr. Hana Kukačková

Poděkování

Ráda bych na tomto místě poděkovala svojí vedoucí Ing. Vladimíře Kubicové za vstřícný přístup a rady a podněty, které mi velice pomohly při zpracování této bakalářské práce.

Obsah

Úvod.....	9
1 Nukleové kyseliny	10
2 Genová exprese.....	13
2.1 Typy RNA	13
Mediátorová RNA (mRNA)	14
Transferová RNA (tRNA).....	14
Ribozomální RNA (rRNA).....	14
Další typy RNA.....	14
2.2 Syntéza RNA	14
2.3 Transkripce	14
2.4 Posttranskripční úpravy	15
2.5 Translace.....	17
2.5.1 Ribozom	17
2.5.2 Inicie translace	17
2.5.3 Elongace	18
2.5.4 Terminace.....	18
3 Význam a vývoj intronů.....	19
4 Metody vyhledávání intronů a exonů u eukaryotických organismů	20
4.1 Metody vyhledávání intronů a exonů založené na vyhledání určité sekvence.....	20
4.1.1 Metoda Katayamy a Tomoshige	20
4.1.2 Genetické programování	20
4.1.3 Hidden Markov Model	20
4.2 Metody vyhledávání intronů a exonů pomocí Fourierovy transformace	21
5 Popis metod.....	22
5.1 Získání dat pro analýzu.....	22
5.1.1 Vyhledavac.....	22
5.1.2 Vyhledávání hranic.....	22
5.2 Ověřování konzervativnosti na rozhraní intronů a exonů.....	23
5.3 Vyhledávání konzervativních motivů.....	24

5.3.1	Funkce hledani_motivu	24
5.3.2	Funkce vyhodnoceni	24
5.3.3	Funkce vyhodnoceni2	24
6	Analýza genů	25
6.1	Ověřování konzervativnosti dinukleotidových sekvencí AG a GT	25
6.1.1	Homo sapiens	25
6.1.2	Mus musculus.....	27
6.1.3	Gallus gallus.....	29
6.1.4	Apis mellifera	31
6.2	Analýza hraniční oblasti intronů – exonů.....	33
6.2.1	Homo sapiens	33
6.2.2	Mus musculus.....	36
6.2.3	Gallus gallus.....	38
6.2.4	Apis mellifera	40
6.3	Vyhledávání konzervativních motivů.....	42
6.3.1	Homo sapiens	42
6.3.2	Mus musculus.....	43
6.3.3	Gallus gallus.....	43
6.3.4	Apis mellifera	44
	Závěr.....	45
	Seznam použité literatury	47
	Seznam použitých zkratek.....	50
	Seznam obrázků	51
	Seznam tabulek	52

Úvod

Genetika je v dnešní době oborem velice moderním se stále se rozšiřujícími aplikacemi v praxi. Je zmapován genom řady organismů, známe funkce a produkty genové exprese a dokážeme interpretovat i vyhledávat různé anomálie a mutace a přiřadit jim konkrétní poruchu. Jsme schopni do genetické výbavy i zasahovat a měnit ji v určitých parametrech dle našich představ.

Oblast, která zatím není tak prozkoumaná, je negenová oblast DNA – introny. Jedná se o značně variabilní oblasti DNA, o jejíž úloze při genetické expresi, původu a evoluci víme jen málo a mnohdy se spíše domníváme. Literatura poskytuje důkazy o konzervativnosti určitých oblastí v intronech a domnívá se, že při bližším zkoumání by se mohlo zjistit, že mutace v těchto konzervativních úsecích mohou mít souvislost s nějakým defektem nebo onemocněním.

V bakalářské práci budou shrnuty informace o struktuře nukleových kyselin – DNA a RNA, jako jedinečných látek, které zajišťují přenos genetické informace na další generace. Bude se věnovat genetické expresi, jako procesu, který informaci v nukleových kyselinách uchovanou převádí do funkčních nebo regulačních molekul. Shrne dosavadní poznatky o intronech, jejich významu, evolučních teoriích a původu. Zaměří se také na způsoby, jakými lze introny vyhledávat.

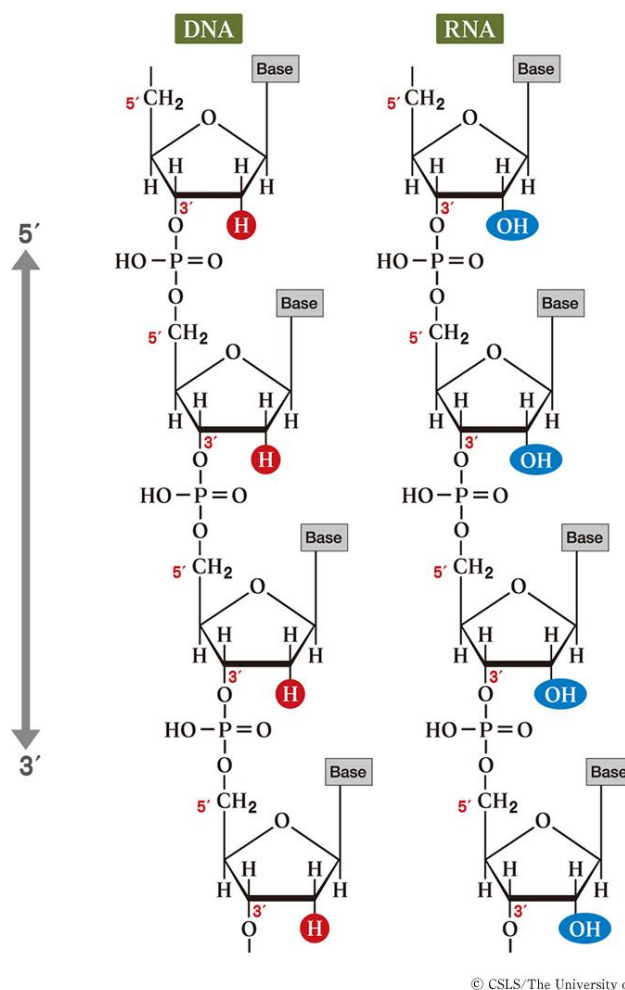
Cílem práce je ověřit konzervativnost oblastí na počátcích a koncích intronů u člověka a dalších organismů a vyhodnotit, do jaké míry jsou konzervativní úseky mezi organismy podobné. Předpokládáme, že sekvence na začátku i konci intronů budou velmi podobné a to napříč různými organismy.

Dále se bude práce věnovat analýze rozhraní mezi kódujícími a nekódujícími úseky. Zajímá nás, zda najdeme opakující se nukleotidy pouze v intronech nebo nalezneme pravidla i v exonech.

V posledním bodě se práce bude pokoušet vyhledat další oblasti nebo sekvence v intronech, které by mohly být konzervativní. Poté tyto motivy bude vyhledávat v celé sadě intronů a bude se snažit najít podobnost v délce sekvencí a v jejich poloze v intronu.

1 Nukleové kyseliny

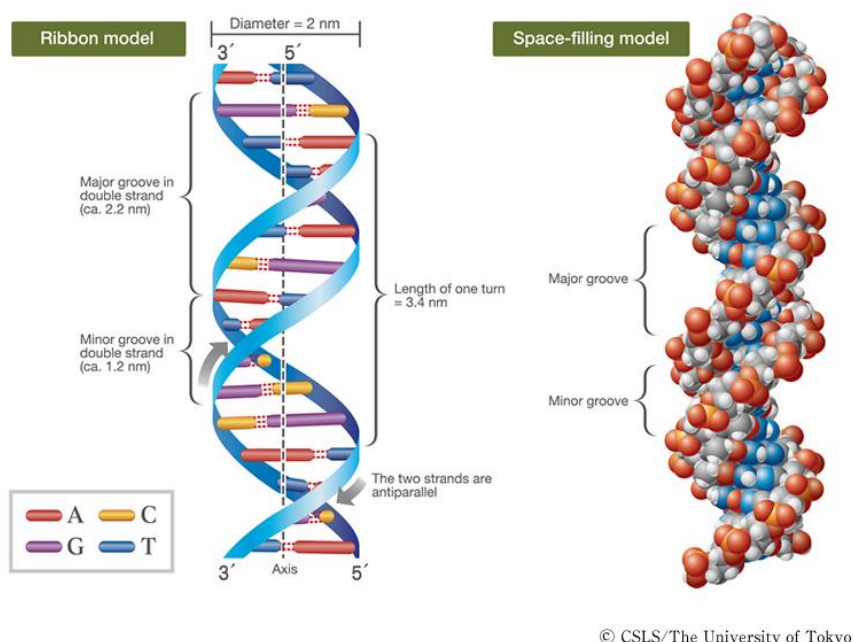
Nukleové kyseliny jsou chemické látky zajišťující uchování a přenos genetické informace. Pro tuto funkci jsou specificky uzpůsobeny. Nukleové kyseliny se skládají z velkého počtu nukleotidů, podjednotek, z níž každá obsahuje tři struktury – pětuhlíkatého sacharidu - pentózy, dusíkaté heterocyklové báze a fosfátové skupiny. Pentózy jsou zastoupeny v nukleových kyselinách dvě – 2-deoxy-D-ribóza v případě DNA a ribóza v případě RNA. Heterocyklové báze můžeme dále rozdělit do dvou skupin – purinové a pyrimidinové. Purinové báze jsou tvořeny dvěma cykly a jsou zastoupeny adeninem a guaninem, pyrimidinové obsahují cyklus pouze jeden a řadíme k nim thymin, cytosin a uracil. Purinové báze jsou zastoupeny v RNA i DNA, z pyrimidinových je v DNA zastoupen thymin a cytosin, v RNA je na místo thyminu uracil. [1], [3], [4]



© CSLS/The University of Tokyo

Obrázek 1 Primární struktura DNA a RNA [14]

DNA se v buňkách nachází nejčastěji ve formě dvoušroubovice. Tuto strukturu udržují především vodíkové můstky – nekovalentní vazby vždy mezi purinovou a pyrimidinovou bází. Tyto vazby nedosahují síly vazby fosfodiesterové, která poutá jednotlivé nukleotidy do řetězce nukleové kyseliny. Mezi adeninem a thyminem se tvoří dva vodíkové můstky, mezi cytosinem a guaninem tři. Výsledkem jsou do spirály navzájem stočené dva řetězce DNA, páry bází jsou uspořádány nad sebou a jsou orientovány kolmo k ose šroubovice. Párování bází je specifické, za normálního stavu se vodíkové můstky nebudou tvořit mezi adeninem a cytosinem nebo guaninem a thyminem. Známe-li tedy pořadí bází v jednom řetězci DNA, známe strukturu druhého vlákna, které je k prvnímu tzv. komplementární. Právě komplementarita je základem uchování a přenosu genetické informace. [1], [3], [4]



Obrázek 2 Dvouvláknová DNA [14]

Šroubovice DNA je velice stabilní i ve vodném prostředí buněk. Tato stabilita je ovlivněna tím, že baze jsou ve svém postavení poměrně nepolární a tudíž se ve vodě nerozpouští. DNA šroubovice má dva žlábký – jeden velký a druhý malý. To hraje svoji roli při genové expresi, protože regulační proteiny se specificky vážou buď do velkého, nebo malého žlábků. [1], [3], [4]

DNA se může vyskytovat v několika konformacích. Nejčastější z nich je B-DNA, kterou tato nukleová kyselina zaujímá ve fyziologických podmínkách. V této konformaci je DNA šroubovice pravotočivá a obsahuje 10,4 nukleotidových párů na otáčku. Další konformací je

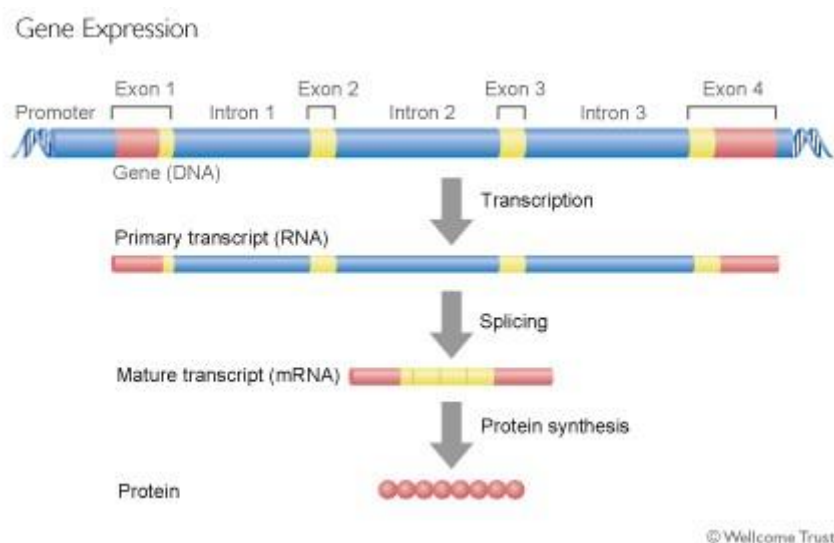
A-DNA, kterou lze pozorovat, je-li DNA v koncentrovaném roztoku soli nebo je-li částečně dehydratována. Ta je rovněž pravotočivá, na otáčku však připadá 11 nukleotidů. Předpokládá se však, že v prostředí in vivo se tato konformace nenachází. Jinou konformací přítomnou v živých organismech je Z-DNA, která je levotočivá, má 12 párů bazí na otáčku, tvoří pouze jeden žlábek a vyskytuje se nejčastěji tam, kde je DNA bohatá na páry C-G. Její význam je prozatím nejasný. [1], [3], [4]

2 Genová exprese

Genová exprese je proces, kterým je genetická informace zakódovaná v nukleové kyselině použita k syntéze funkčního genového produktu. Takovým produktem bývá velmi často protein, ale může to být i funkční RNA. [1], [3], [4]

Genová exprese zahrnuje dva kroky. Prvním je transkripce, kdy se na základě jednoho templátového vlákna DNA na principu komplementarity syntetizuje vlákno RNA. Ve druhém kroku je na základě takto vytvořené RNA tvořen polypeptidový řetězec procesem označovaným jako translace. [1], [3], [4]

Tento přenos genetické informace je vyjádřen jako centrální dogma molekulární biologie. DNA se přenáší do DNA při předávání genetické informace mezi generacemi a z DNA do fenotypového projevu organismu. RNA se může přenášet do RNA při replikaci RNA virů. Přenos je možný i z RNA do DNA za účasti speciálních enzymů. Z toho vyplývá, že přenos genetické informace mezi DNA a RNA je za jistých podmínek vratný, přenos z RNA do proteinů je ireverzibilní. [1], [3], [4]



Obrázek 3 Genová exprese [15]

2.1 Typy RNA

Při genové expresi hraje RNA velice významnou roli. Vzniká transkripcí DNA a vyskytuje se v několika formách.

Mediátorová RNA (mRNA)

nese informaci o struktuře proteinů a přenáší ji z jádra do místa syntézy proteinů. Je navázaná na ribozom a tvoří jakousi šablonu, pomocí které se skládají jednotlivé aminokyseliny do polypeptidového řetězce.

Transferová RNA (tRNA)

tRNA je tvořena malými molekulami, pro něž je typické zatoupení neobvyklých nukleosidů, např. inozinu, pseudouridinu, ad., které vznikají z nukleosidů primárního transkriptu dalšími chemickými ději. tRNA nese antikodonovou sekvenci, která se páruje s kodonem mRNA a tak přináší aminokyseliny na místo proteosyntézy. Aminokyseliny se na tRNA navazují dvoustupňovou reakcí katalyzovanou aminoacyl-tRNA-syntetázou, která je pro jednotlivé aminokyseliny různá.

Ribozomální RNA (rRNA)

rRNA tvoří přibližně polovinu ribozomu, organely zajišťující proteosyntézu. Její syntéza probíhá v jadérku, kde transkripce z DNA vznikají několikanásobně delší prekurzory, které jsou několikanásobně delší než hotová rRNA obsažená v ribosomech. Geny pro rRNA jsou v genomu zastoupeny ve velkém množství kopií, což je dáno velkým počtem ribozomů v buňce.

Další typy RNA

Jako malá jaderná RNA (snRNA) se označují malé molekuly, které jsou součástí tzv. spliceozomů. Ty se podílejí na posttranskripčních úpravách a vystřížení intronů.

Mikro RNA blokuje expresi komplementárních úseků mRNA. [1], [3], [4]

2.2 Syntéza RNA

Syntéza RNA probíhá v jádře z prekurzorů – ribonukleosidtrifosfátů. Templátem je jedno vlákno DNA, řetězce vznikají zcela nové, nepotřebují předpřipravený primer jako DNA. Nové vlákno RNA je komplementární k templátovému vláknu DNA, namísto thyminu však obsahuje uracil. Syntéza RNA probíhá ve směru 5'→3'. Reakci, která navazuje ribonukleotidy na 3'-OH skupinu konce řetězce, katalyzují RNA-polymerázy. Syntéza RNA začíná na specifických místech označovaných jako promotory. [1], [3], [4]

2.3 Transkripce

Proces transkripce probíhá v buněčném jádře, má tři fáze a je katalyzována třemi enzymy - RNA – polymerázami, z nichž každá katalyzuje transkripci specifické skupiny genů.

Pro iniciaci transkripce je u eukaryotických buněk krom příslušné RNA-polymerázy nezbytná přítomnost specifických bílkovin - transkripčních faktorů. Nejprve dojde k interakci enzymu a určitých transkripčních faktorů na promotorovou oblast DNA, což je konzervativní úsek zahrnující kratší úseky DNA označované jako moduly. Takto vytvořený komplex způsobí rozvolnění dvoušroubovice DNA, čímž se získá templátové vlákno. [1], [3], [4]

V této druhé fázi, elongaci, se vlákno RNA prodlužuje navazováním odpovídajících ribonukleotidů. Během tohoto procesu je na nově vznikající vlákno RNA na 5' konci navázána 7-metylguanositová čepička, která zajišťuje nově vzniklému vláknu ochranu před enzymatickou lýzou. [1], [3], [4]

Terminace celého procesu probíhá u jednotlivých RNA-polymeráz různě. V případě RNA-polymerázy II je transkripce zastavena o několik tisíc nukleotidů dále za terminační sekvenci a místo, které bude 3' koncem nové RNA se utvoří endonukleázovým odštěpením. Poté je na tento konec navázán úsek obsahující až 200 zbytků AMP. Tato reakce se označuje jako polyadenylace. [1], [3], [4]

RNA-polymerázy I a III provádějí terminaci transkripce v závislosti na svých terminačních signálech.

V některých případech může také dojít k tzv. editaci RNA, což je proces, během kterého se mění některé sekvence nukleotidů v RNA před jejich translací, nicméně tento děj má význam především při expresi mitochondriálních genů rostlinných buněk nebo prvoků rodu *Trypanosoma*. [1], [3], [4]

2.4 Posttranskripční úpravy

Většina eukaryotických genů obsahuje úseky, které jsou exprimovány, a úseky, které nikoliv. Mluvíme o exonech a intronech. Exony jsou sekvence DNA, které jsou dále exprimovány a které jsou přerušovány nekódujícími sekvencemi – introny. [1]

Význam intronů zatím není zcela jasný. Jejich velikost je rozmanitá, stejně jako jejich počet nebo zastoupení jednotlivých nukleotidů. Předpokládá se, že by mohly hrát úlohu v regulaci genové exprese nebo oddělovat exony jednoho genu kódující různé podjednotky určitého proteinu. V některých případech dochází k nestejné úpravě při sestřihu, což dává možnost vzniku vzájemně příbuzných bílkovin na základě jednoho genu. V tomto případě introny fungují jako „enhancery“ při meiotickém crossing-overu uvnitř genů kódujících bílkoviny. Bylo však také zjištěno, že introny snáze a čteněji podléhají mutacím, které však vůbec

nemusí mít vždy při translaci a následně funkci žádný vliv, i když bylo prokázáno, že mutace v konzervativních úsecích, které napomáhají identifikaci intronu, mohou zodpovídat za některá genetická onemocnění člověka. [1]

Exony i introny jsou z DNA přepisované do RNA, poté dochází k sestřihu transkriptu, jehož výsledkem je mRNA obsahující pouze přepsané exony. Introny jsou velice variabilní, obsahují však konzervativní úseky, které napomáhají rozpoznávat místa sestřihu. [1]

K tomuto sestřihu může dojít třemi mechanismy.

Sestřih endonukleázami a následná DNA ligace

Tímto způsobem jsou upravovány tRNA u kvasinek a probíhá ve dvou krocích. Nejprve endonukleáza navázaná na jadernou membránu katalyzuje odštěpení intronu na obou koncích a poté ligáza zajistí spojení obou částí do jedné definitivní molekuly tRNA. Místo, kde k sestřihu dojde, závisí na 3D struktuře molekul tRNA. [1]

Autokatalytický sestřih

K autokatalytickému sestřihu dochází bez přítomnosti enzymů, reakce je iniciována samotnou molekulou RNA. Tyto děje byly popsány u rRNA vývojově nižších eukaryotních organismů a také u rRNA, tRNA a mRNA autonomních organel vyšších eukaryot. Jednou z variant samosestřihu je reakce s guaninovým nukleotidem nebo nukleosidem, který má volnou 3'-OH skupinu. Nejprve se rozvolní vazba prvního exonu s intronem, poté se fosfodiesterová vazba ze spojení druhého exonu s intronem přenesne na první intron a molekula RNA se spojí, ve třetím kroku je pak intron spojen uvolněnou fosfodiesterovou vazbou do kruhové molekuly.

K tomu, aby k autokatalytické reakci došlo, musí být reagující skupiny dostatečně blízko. Tento způsob sestřihu je tedy závislý na sekundární konformaci molekuly prekurzorové RNA. Zároveň by mohlo dojít ke zpětné reakci, čemuž zabrání rychlé odbourání kruhových intronů.

Jiný způsob samosestřihu se podobá následujícímu, ale nezapojují se do něj spliceozomy. [1]

Spliceozomy

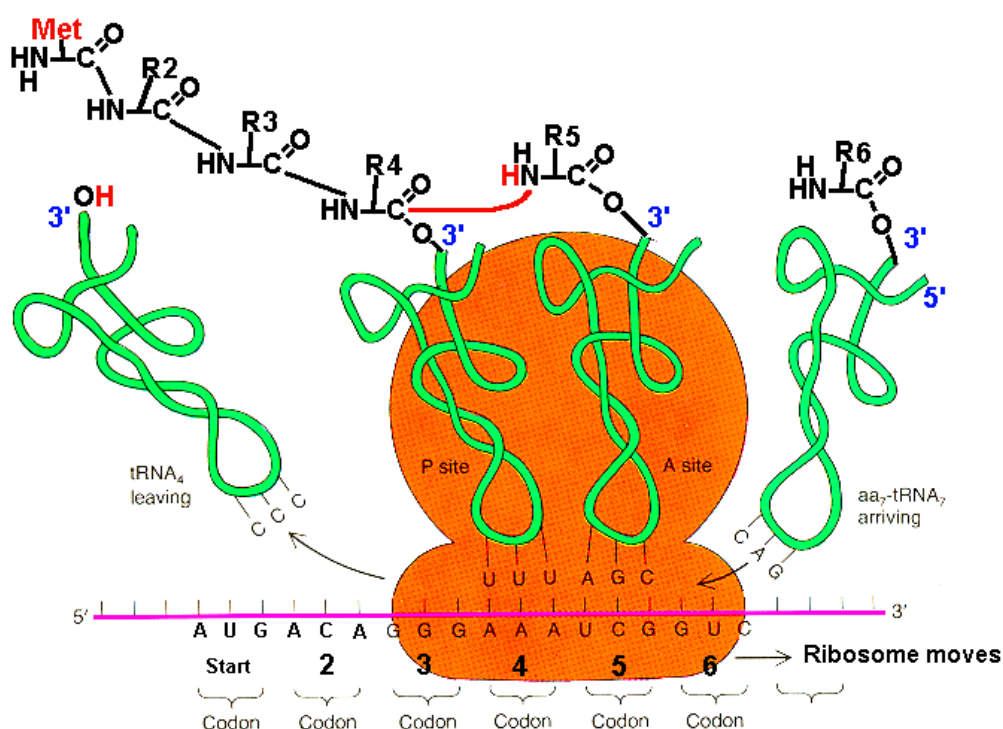
Spliceozomy zajišťují sestřih jaderných pre-mRNA. Jedná se o malé útvary, které obsahují několik malých molekul tzv. snRNA a různé proteiny označované jako snop (small nuclear ribonucleoproteins) a seskupují se do podjednotek U1 – U6, z nichž se jediná U3 sestřihu neúčastní, je součástí jádérka a účastní se formování ribozomů. Mechanismus probíhá ve dvou krocích, některé děje však nejsou zatím zcela prostudované. [1]

2.5 Translace

Translace je proces překlada genetické informace do struktury bílkovin. Tento děj je velice složitý a účastní se ho celá řada makromolekul – RNA, bílkoviny, enzymy. Hlavní úlohu v tomto procesu má ribozom. [1], [3], [4]

2.5.1 Ribozom

Jedná se o organelu, která se může nacházet v mitochondriích nebo chloroplastech, jadérku nebo v cytoplazmě buněk vázaná na endoplazmatické retikulum. Skládá se ze dvou podjednotek – malé a velké, a je tvořena tRNA a bílkovinami. Ribozom nese tři místa pro vazbu tRNA – aminoacylové místo, na které se váže t-RNA nesoucí další aminokyselinu, peptidylové místo pro tRNA nesoucí nově vznikající peptidové vlákno a místo exitu, kde se uvolňuje tRNA, která o svoji aminokyselinu přišla. [1], [3], [4]



Modified from Griffiths et al., AN INTRODUCTION TO GENETIC ANALYSIS, 6th Ed., W.H. Freeman & Co., 1996.

Obrázek 4 Ribozom a translace [16]

2.5.2 Iniciační translace

Translace je iniciována napojením tRNA nesoucí methionin, protože kodon pro methionin je zároveň iniciačním kodonem. Následujícími ději může být methionin ze začátku peptidu odstraněn. tRNA nesoucí methionin vytvoří komplex s příslušným iniciačním faktorem a tento komplex se pak navazuje na druhý komplex tvořený mRNA, malou podjednotkou ribozomu a dalšími iniciačními faktory. V závěru této fáze se iniciační faktory oddělí a

připojuje se ke komplexu i velká podjednotka ribozomu. tRNA nesoucí methionin se v peptidylovém místě naváže svým antikodonem na AUG mRNA, tím je na aminoacylovém místě připraven nový kodon a může začít druhá fáze translace. [1], [3], [4]

2.5.3 Elongace

Na A místo ribozomu se naváže aminoacyl-tRNA podle kodonu na mRNA. Tato aminokyselina vytvoří peptidovou vazbu s karboxylovou skupinou peptidu, který je nesen tRNA navázanou v místě P. Tuto reakci katalyzuje peptidyltransferáza. Pro vytvoření peptidové vazby je nutná energie dodaná rozštěpením molekuly GTP. Poté se ribozom posouvá po mRNA o tři nukleotidy směrem k 3'-konci, rostoucí peptidový řetězec se přesouvá na tRNA v místě A a ta se přesouvá na místo P. tRNA, která se takto uvolnila, se posouvá do místa E. [1], [3], [4]

2.5.4 Terminace

Ukončení translace je signalizováno terminačními kodony – UAA, UAG, UGA. Tyto kodony jsou rozeznány uvolňovacími faktory (RF) a ty mění děj tak, že se na karboxylovou skupinu naváže voda, a tím se peptid uvolní z tRNA v místě P. Poté se ribozom rozpadá na podjednotky, které jsou připraveny pro další translaci. [1], [3], [4]

3 Význam a vývoj intronů

Existují dvě teorie o vzniku intronů. První z nich (introns-late) uvažuje o tom, že introny jsou v evoluci nevýznamné a objevují se v pozdních fázích evoluce a to pouze u eukaryot. Jejich vznik popisuje jako začlenění části kódu do oblastí budoucích sestřihů (proto-splice místa). Druhá (introns-early) považuje introny za nezbytnou podmínku při evoluci genů genovým posunem, kdy vznikaly exony peptidových modulů rekombinací uvnitř intronů. Podle této teorie by introny existovaly dříve, než došlo k rozdělení prokaryot od eukaryot a předpokládá, že prokaryota by do dnešních dob všechny introny ztratila. [7]

Fedorov, Merican a Gilbert provedli rozsáhlou studii, ve které porovnávali introny rostlin, živočichů a hub, vybrali vzorky tak, aby odstranili kopie genů a porovnali pozice intronů u homologních genů. Výsledkem tohoto výzkumu bylo zjištění, že introny v nápadně vysokém procentu, u rostlinných a živočišných genů v 10%, u srovnání hub a rostlin nebo živočichů dokonce 17 – 18 % leží na stejných pozicích, u třiceti nejpodobnějších genů zaznamenali dokonce 44% shodu mezi živočichy a rostlinami a 39 % mezi houbami a rostlinami nebo živočichy. Dále popisují přítomnost „sliding“ pozic, tedy pozic intronů, které sobě odpovídají s posunem do šesti nukleotidů. Tyto „sliding“ introny však mohou vznikat i arteficiálně při strojovém zpracování a zarovnání sekvencí. Část těchto intronů je původních, některé však byly do DNA zařazeny až nedávno. Jejich stejné pozice jsou výsledkem náhody. Zřejmě je většina podobných pozic intronů původních, tedy potvrzující teorii introns-early. [7], [13]

Na okrajích jaderných intronů bývají přítomny dinukleotidové sekvence GT a AG a uvnitř ve vzdálenosti přibližně 30 nukleotidů proti směru transkripce od 3' sestřihového místa intronu, se nachází sekvence TACTAAC. Jiné konzervativní sekvence jsou pak přítomny u genů autonomních organel – mitochondrií a chloroplastů. [1]

4 Metody vyhledávání intronů a exonů u eukaryotických organismů

Metody pro vyhledávání určitých bazí – v našem případě intronů a exonů lze rozdělit podle principu hledání do dvou velkých kategorií.

4.1 Metody vyhledávání intronů a exonů založené na vyhledání určité sekvence

První kategorii tvoří metody, které rozlišují určitou nukleotidovou sekvenci, kterou určitým algoritmem vyhledávají v genomu.

4.1.1 Metoda Katayamy a Tomoshige

Tato metoda je jedním z příkladů metod založených na detekci GT struktur. V první řadě metoda vyhledává kombinaci šesti bazí okolo GT sekvence, zjistí, jak často se vyskytují v GT oblasti hraniční mezi intronem a exonem a jak často se vyskytují v oblasti nehraniční. Poté jsou vybrány kombinace těchto šestic s nejnižším počtem výskytů – tedy ty, které jsou nejspecifičtější – pro hraniční i nehraniční oblasti. Na základě těchto šestic pak lze vyhodnotit, je-li GT sekvence hraniční nebo ne. Výhodou této metody je její rychlost, není však tak přesná jako konvenční metody. [2]

4.1.2 Genetické programování

Princip genetického programování je podobný principu biologické evoluce. Na začátku jsou vytvořeni jedinci – počítačové programy. U každého jedince je stanovena zdatnost a je vybráno několik jedinců s vysokou zdatností, ze kterých se různými postupy získávají jedinci noví. Pro ty je rovněž stanovena zdatnost a proces se opakuje, dokud není algoritmus ukončen. Jedinec s nejvyšší hodnotou zdatnosti je považován za nejlepšího. [5]

4.1.3 Hidden Markov Model

Hidden Markov Model se využívá při zpracování signálu, v algoritmech rozpoznávání řeči a často také v bioinformatice, obecně řeší problém vyhledání určitého motivu. V bioinformatice se využívá hlavně pro porovnávání sekvencí, predikci struktur nebo detekci genů. Pro algoritmus je nutná nejprve fáze vyhodnocení systému – v tomto případě DNA. Na základě tohoto pozorování se zvolí pravděpodobně nejvhodnější Hidden Markov Model. Ve druhé fázi se snažíme determinovat tzv. skryté stavy, tedy stavy, které nejsou přímo zřejmé, ale způsobují něco, co zřejmé je. Na základě těchto pozorování je sestaven konkrétní model, který nejlépe popisuje danou situaci a na jehož základě lze motivy identifikovat. [8]

4.2 Metody vyhledávání intronů a exonů pomocí Fourierovy transformace

Další možností je zpracování sekvence DNA pomocí Fourierovy transformace, kde je nutné převést DNA na signál. Toho dosáhneme jednoduše přiřazením číselných hodnot písmenům v DNA sekvenci.

Fourierova transformace se používá nejčastěji pro vyhledávání exonů. Je to dáno tím, že právě části jaderných DNA eukaryot vykazují určitou periodicitu opakování rovnu délce kodonu, tedy rovnu třem. Zvolíme-li délku okna N rovnu násobku tří, po vykreslení výkonového spektra získáme na hodnotě vzorku $N/3$ peak. Okno se poté posouvá. Je nutné zajistit dostatečně velké okno, aby se periodičita dostatečně projevila, čím delší okno ale použijeme, tím náročnější bude výpočet. [17]

5 Popis metod

Pro zpracování analýz byl využit program Matlab, jehož součástí je bioinformatický toolbox, který obsahuje řadu funkcí pro zpracování a analýzu bioinformatických dat.

Práce obsahuje tři skripty pro analýzu intronů a dva skripty pro předzpracování dat, která do analýz vstupují. Tato úprava je oproti spouštění jedním skriptem výpočetně výrazně rychlejší, odpadá opakované procházení souborů, které je u takto velkého souboru dat časově náročné.

5.1 Získání dat pro analýzu

Vstupní data byla získána z anotovaných sekvencí uložených v databázi NCBI. Pro automatické načítání dat byl použit formát genbank. Data vstupují do analýz již předzpracovaná v souborech **introny.mat** a **boundary.mat**.

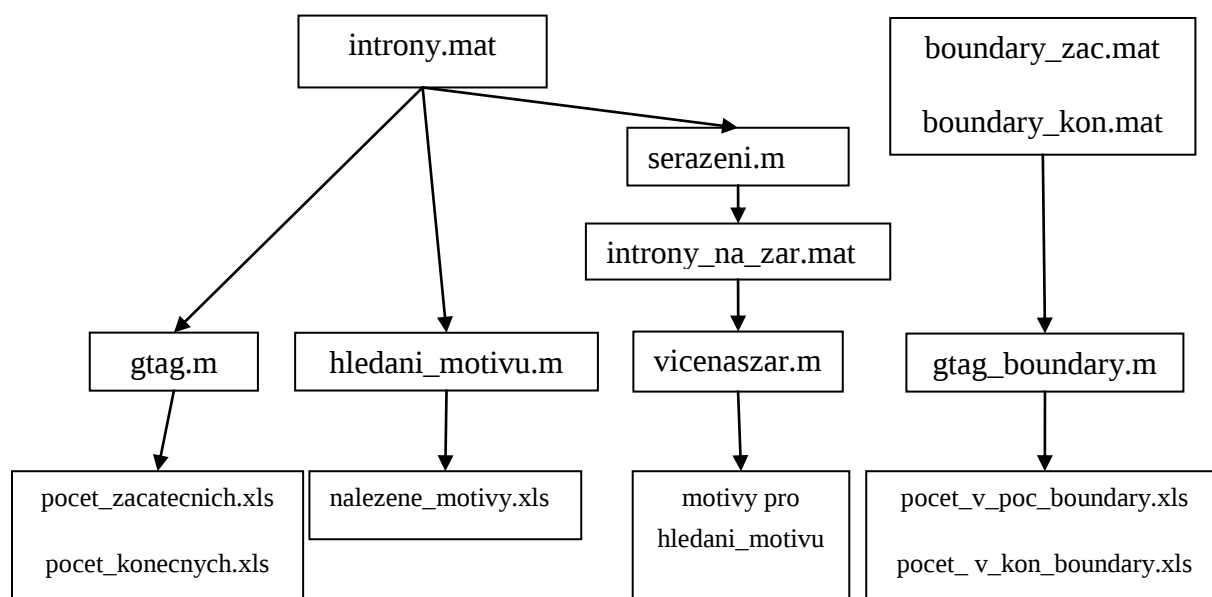
5.1.1 Vyhledavac

Skript **vyhledavac** vyhledává a ukládá introny ze všech genů pro další analýzy. Geny jsou uloženy v adresáři sekvence, který je ve stejné složce jako skript samotný. Postupně jsou procházeny a funkcí **genbankread** jsou ze všech souborů načteny do workspace. Takto získáme strukturu obsahující gen i další informace, nás bude zajímat vlastní sekvence a CDS. Pro vyhledání intronů slouží vlastní funkce **introny**, která introny izoluje podle pozic uložených v CDS. Výsledky jsou zařazeny do matice, která je následně uložena do souboru **introny.mat**.

5.1.2 Vyhl_boundary

Vstupními data pro analýzu hraniční oblasti obsahuje soubor **boundary.mat**. Tento soubor získáme spuštěním skriptu **vyhl_boundary**, který pracuje obdobně jako **vyhledavac**. Projde adresář sekvence, pomocí funkce **getgenbank** načte informace o genech a cds a tato data postoupí funkci **boundary**, která zajistí vyhledání nukleotidů nacházejících se deset pozic před příslušným cds a deset pozic za ním. Zajímají nás zvláště oblasti intron – exon a exon – intron, proto funkce odlišuje tyto úseky podle sudosti nebo lichosti čítače i. Funkce obsahuje také podmínku, která vylučuje krátké oblasti, které před nebo za CDS mají méně než deset nukleotidů. Výsledky uloží do souboru **boundary_zac.mat** a **boundary_kon.mat**.

Návaznost skriptů na vstupní data je popsána na blokovém schématu.



Obrázek 5 Blokové schéma

5.2 Ověřování konzervativnosti na rozhraní intronů a exonů

Pro ověření konzervativnosti dinukleotidových sekvencí AG a GT byl vytvořen skript **gtag**, který zjišťuje počet nukleotidů na začátku a na konci intronu. Vstupem je soubor **introny.mat**. Z tohoto souboru je získána sekvence prvních a posledních deseti nukleotidů. Funkcí **pocitani** stanovujeme počty nukleotidů A, C, G a T a ukládáme je do vektorů. Výsledky jsou uloženy do souborů **pocet_konecnich.xls** a **pocet_zacatecnich.xls**, kde řádky představují počty jednoho nukleotidu na pozicích 1 – 10 a sloupce počet nukleotidů A – T na příslušné pozici.

Analýza hraniční oblasti se spouští skriptem **gtag_boundary**. Vstupem jsou hraniční oblasti vypsané do souboru **boundary.mat**, ten je načten a pak jsou jednotlivé řádky prohledávány a pomocí funkce **pocitani** jsou určeny počty nukleotidů na příslušných pozicích. Výsledky jsou uloženy do souborů **pocet_poc_boundary.xls** pro oblast intron-exon a **pocet_kon_boundary.xls** pro oblast exon-intron.

Pro zobrazení konzervativnosti příslušných oblastí byla použita funkce sekvenčního loga, které popisuje konzervativnost sekvencí nukleotidů v řetězci nukleotidových kyselin nebo aminokyselin v proteinovém řetězci. Je tvořeno ze zarovnaných sekvencí. Pro každou pozici je vytvořen sloupec písmen označujících jednotlivé nukleotidy nebo aminokyseliny. Celková výška tohoto sloupce charakterizuje konzervativnost sekvence na dané pozici. Výška jednotlivých symbolů odpovídá relativní frekvenci daného nukleotidu nebo aminokyseliny na

příslušné pozici. Bioinformatický toolbox v Matlabu obsahuje funkci **seqlogo**, která sekvenční logo ze zarovnaných sekvencí tvoří. [19]

5.3 Vyhledávání konzervativních motivů

Analýza konzervativnosti motivů se spouští skriptem **doexcelu**. Na počátku je nutné vyplnit jednořádkovou matici, která obsahuje ty sekvence, které označíme za motiv a které chceme vyhledávat. Poté je pro každý motiv spuštěna funkce **prohledavani**, která pomocí funkce **hledani_motivu**, **vyhodnoceni** a **vyhodnoceni2** introny analyzuje. Všechny tyto funkce budou popsány níže.

Motivy, jejichž konzervativnost byla ověřována, byly získány vícenásobným zarovnáním některých sekvencí a použitím takto získaných konsenzuálních sekvencí na všechny introny uložené v **introny.mat**. Mezi motivy byl také zařazen motiv TACTAAC, jehož konzervativnost již ověřena byla.[1]

Výsledkem skriptu je tabulka MS Excel **nalezene_motivy**, kde první sloupec obsahuje motiv, který byl vyhledáván, druhý sloupec počet všech intronů, které do analýzy vstoupily, třetí sloupec počet intronů, ve kterých byl motiv nalezen a další sloupce obsahují výsledky funkcí **vyhodnoceni** a **vyhodnoceni2**, tedy pozice, na kterých byl motiv nalezen.

5.3.1 Funkce hledani_motivu

Funkce načítá soubor **introny.mat** a pomocí vyhledává v intronech požadovaný motiv. Výsledkem je matice identifikace, jejíž každý řádek odpovídá jednomu intronu, který daný motiv obsahuje, a sloupce určují pozici, na které se motiv nachází. Dalším výstupem funkce je celkový počet intronů a počet intronů, ve kterých se motiv vyskytuje.

5.3.2 Funkce vyhodnoceni

Funkce jako vstupní parametr požaduje matici identifikace. Intron je funkcí rozdělen po deseti procentech, a pokud je přítomen v dané části vyhledávaný motiv, přiřadí do vektoru histogram na příslušné místo hodnotu 1. Cílem funkce tedy je zhodnotit, zda se daný motiv nachází v některé procentuálně dané oblasti intronů nebo ne, nezajímá nás počet nalezených intronů v dané části, protože opakovaná přítomnost motivu v dlouhých intronech by výsledek výrazně zkreslila.

5.3.3 Funkce vyhodnoceni2

Funkce pracuje na obdobném principu jako výše uvedená funkce **vyhodnoceni**. Rozdíl je, že prohledává prvních a posledních 100 nukleotidů v každém intronu, tuto oblast dělí na deset dílů a počítá, kolik motivů celkem se v dané oblasti nachází. V tomto případě je počet nalezených motivů nutné uvádět, protože informace o počtu nám také říká, jak významný výskyt v dané oblasti je.

6 Analýza genů

Pro analýzu byly využity geny člověka (*Homo sapiens*) a tří dalších různých organismů – myši domácí (*Mus musculus*), kura bankivského (*Gallus gallus*) a včely medonosné (*Apis mellifera*). Taxonomické zařazení organismů uvádí Tabulka 1.

Tabulka 1 Taxonomické zařazení uvedených organismů

	Kmen	Třída	Řád	Čeleď
<i>Homo sapiens</i>	strunatci (<i>Chordata</i>)	savci (<i>Mammalia</i>)	primáti (<i>Primates</i>)	hominidi (<i>Hominidae</i>)
<i>Mus musculus</i>	strunatci (<i>Chordata</i>)	savci (<i>Mammalia</i>)	hlodavci (<i>Rodentia</i>)	myšovití (<i>Muridae</i>)
<i>Gallus Gallus</i>	strunatci (<i>Chordata</i>)	ptáci (<i>Aves</i>)	hrabaví (<i>Galliformes</i>)	bažantovití (<i>Phasianidae</i>)
<i>Apis mellifera</i>	členovci (<i>Arthropoda</i>)	hmyz (<i>Insecta</i>)	blanokřídlí (<i>Hymenoptera</i>)	včelovití (<i>Apidae</i>)

Pro analýzu byly použity geny u člověka ze somatických chromozomů číslo 1, 4, 8 a 10 a pohlavních chromozomů X a Y, u myši ze somatických chromozomů číslo 1, 4, 12 a 16 a pohlavních chromozomů X a Y, u kura ze somatických chromozomů 1 a 28 a pohlavního chromozomu W a u včely z chromozomů LG1 – LG16. [18]

Pro spuštění všech analýz je doporučena 64bit verze programu Matlab a konfigurace počítače s minimálně 4 GB RAM.

6.1 Ověřování konzervativnosti dinukleotidových sekvencí AG a GT

6.1.1 *Homo sapiens*

Celkem bylo zpracováno 803 intronů.

Konzervativnost dinukleotidové sekvence GT na počátku nukleotidu se potvrdila. Nukleotid G byl přítomen na první pozici v 88 % případů, nukleotid T na druhé pozici 85 % případů. Kromě těchto dvou nukleotidů se ve více než padesáti procentech analyzovaných případů na dalších, tzn. na třetím až šestém místě od počátku, nacházely nukleotidy AAG, jak ukazuje Tabulka 2 a Tabulka 3.

Tabulka 2 Absolutní počty jednotlivých nukleotidů na začátku intronů u člověka

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	52	73	445	527	87	176	248	183	161	215
C	28	20	29	91	92	157	152	220	172	187
G	706	28	272	75	531	143	183	145	219	162
T	17	682	57	110	93	327	220	255	251	239

Tabulka 3 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů u člověka

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	6,48	9,09	55,42	65,63	10,83	21,92	30,88	22,79	20,05	26,77
C	3,49	2,49	3,61	11,33	11,46	19,55	18,93	27,40	21,42	23,29
G	87,92	3,49	33,87	9,34	66,13	17,81	22,79	18,06	27,27	20,17
T	2,12	84,93	7,10	13,70	11,58	40,72	27,40	31,76	31,26	29,76

Konzervativnost sekvence AG byla také potvrzena. Procentuálně se nukleotid A vyskytoval na předposlední pozici v 85 % případů a nukleotid G na posledním místě v 84 % případů. Na třetí a páté pozici od konce se vyskytuje stejný nukleotid u více než padesáti procent analyzovaných intronů. Na pozicích šest až deset jsou přítomny stejné nukleotidy ve 40 až 50 % případů. Jednotlivé hodnoty jsou zobrazeny v Tabulka 4 a Tabulka 5.

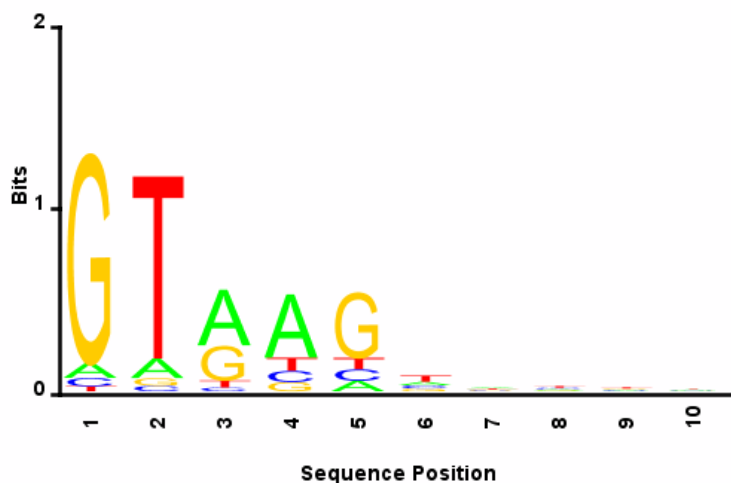
Tabulka 4 Absolutní počty jednotlivých nukleotidů na konci intronů u člověka

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	26	686	100	209	127	97	115	132	95	108
C	76	31	457	208	164	211	236	229	251	222
G	675	23	32	180	80	94	80	81	96	123
T	26	63	214	206	432	401	372	361	361	350

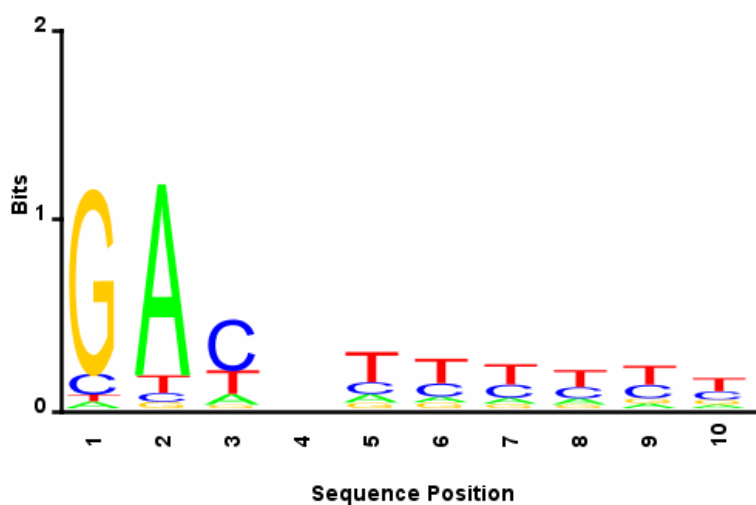
Tabulka 5 Procentuální zastoupení jednotlivých nukleotidů na konci intronů u člověka

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	3,24	85,43	12,45	26,03	15,82	12,08	14,32	16,44	11,83	13,45
C	9,46	3,86	56,91	25,90	20,42	26,28	29,39	28,52	31,26	27,65
G	84,06	2,86	3,99	22,42	9,96	11,71	9,96	10,09	11,96	15,32
T	3,24	7,85	26,65	25,65	53,80	49,94	46,33	44,96	44,96	43,59

Analýza byla potvrzena vytvořením sekvenčního loga, viz Obrázek 6 a Obrázek 7.



Obrázek 6 Sekvenční logo začátku intronů u člověka



Obrázek 7 Sekvenční logo konce intronů u člověka

6.1.2 Mus musculus

Celkově bylo analýzou zpracováno 405 intronů.

Konzervativnost dinukleotidové sekvence GT na počátku nukleotidu se potvrdila. Nukleotid G byl přítomen na první pozici v 84 % případů, nukleotid T na druhé pozici v 77 % případů. Kromě těchto dvou nukleotidů se ve více než padesáti procentech intronů nacházel stejný nukleotid na třetí, čtvrté a páté pozici. Získané hodnoty reprezentuje Tabulka 6 a Tabulka 7.

Tabulka 6 Absolutní počty jednotlivých nukleotidů na začátku intronů u myši

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	28	47	215	216	44	81	128	102	94	118
C	18	22	33	61	49	80	95	93	98	80
G	341	24	98	54	242	85	101	91	104	88
T	18	312	59	74	70	159	81	119	109	119

Tabulka 7 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů u myši

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	6,91	11,60	53,09	53,33	10,86	20,00	31,60	25,19	23,21	29,14
C	4,44	5,43	8,15	15,06	12,10	19,75	23,46	22,96	24,20	19,75
G	84,20	5,93	24,20	13,33	59,75	20,99	24,94	22,47	25,68	21,73
T	4,44	77,04	14,57	18,27	17,28	39,26	20,00	29,38	26,91	29,38

Konzervativnost sekvence AG byla také potvrzena. Procentuálně se nukleotid A vyskytoval na předposlední pozici v 75 % případů a nukleotid G na posledním místě v 74 % případů. Na třetí pozici od konce se vyskytuje stejný nukleotid u více než padesáti procent analyzovaných intronů. Jednotlivé hodnoty jsou zobrazeny v Tabulka 8 a Tabulka 9.

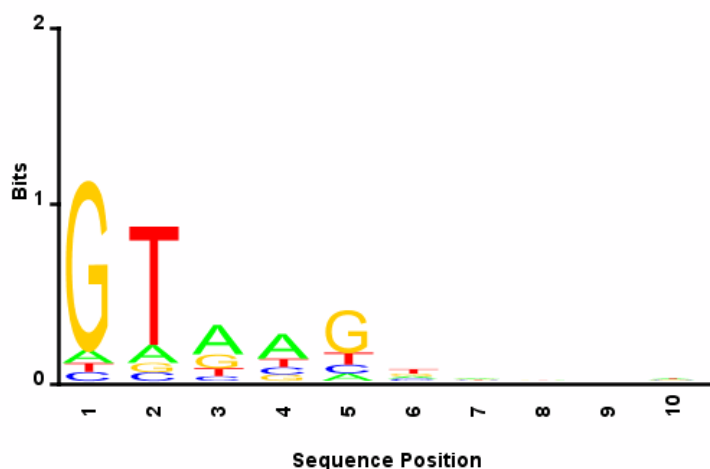
Tabulka 8 Absolutní počty jednotlivých nukleotidů na konci intronů u myši

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	24	311	58	124	61	56	64	48	55	53
C	58	22	219	100	112	140	119	141	113	101
G	307	23	20	92	35	54	48	46	55	67
T	16	49	108	89	197	155	174	170	182	184

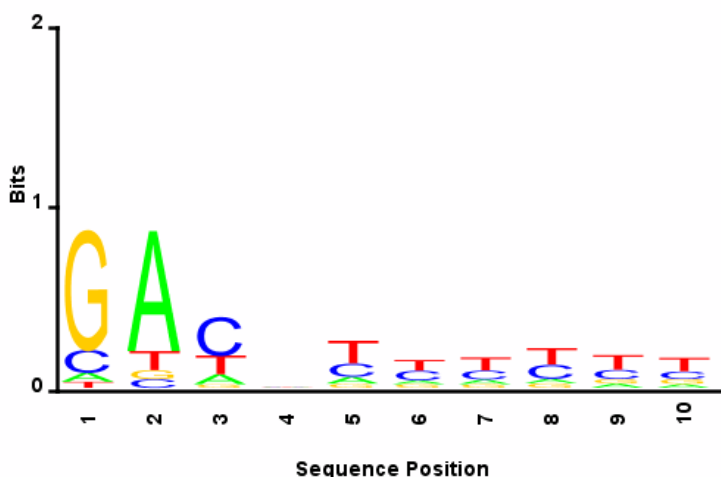
Tabulka 9 Procentuální zastoupení jednotlivých nukleotidů na konci intronů u myši

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	5,78	74,94	13,98	29,88	14,70	13,49	15,42	11,57	13,25	12,77
C	13,98	5,30	52,77	24,10	26,99	33,73	28,67	33,98	27,23	24,34
G	73,98	5,54	4,82	22,17	8,43	13,01	11,57	11,08	13,25	16,14
T	3,86	11,81	26,02	21,45	47,47	37,35	41,93	40,96	43,86	44,34

Analýza byla potvrzena vytvořením sekvenčního loga, viz Obrázek 8 a Obrázek 9.



Obrázek 8 Sekvenční logo začátku intronů u myši



Obrázek 9 Sekvenční logo konce intronů u myši

6.1.3 Gallus gallus

Celkově bylo analýzou zpracováno 3449 intronů.

Konzervativnost dinukleotidové sekvence GT na počátku nukleotidu se potvrdila. Nukleotid G byl přítomen na první pozici v 80 % případů, nukleotid T na druhé pozici ve více než 76 % případů. Kromě těchto dvou nukleotidů se ve více než padesáti procentech intronů nacházel stejný nukleotid na čtvrté a páté pozici (Tabulka 10 a Tabulka 11).

Tabulka 10 Absolutní počty jednotlivých nukleotidů na začátku intronů kura

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	252	384	1722	2012	324	970	1096	602	879	845
C	313	239	398	454	626	560	715	851	860	995
G	2766	207	1062	413	2119	713	935	1115	855	887
T	118	2619	267	570	380	1206	703	881	855	722

Tabulka 11 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů kura

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	7,31	11,13	49,93	58,34	9,39	28,12	31,78	17,45	25,49	24,50
C	9,08	6,93	11,54	13,16	18,15	16,24	20,73	24,67	24,93	28,85
G	80,20	6,00	30,79	11,97	61,44	20,67	27,11	32,33	24,79	25,72
T	3,42	75,94	7,74	16,53	11,02	34,97	20,38	25,54	24,79	20,93

Konzervativnost sekvence AG byla také potvrzena. Procentuálně se nukleotid A vyskytoval na předposlední pozici v 84 % případů a nukleotid G na posledním místě v 86 % případů. Další shoda byla nalezena na třetí a páté pozici od konce, kde se vyskytuje ve více než polovině intronů stejný nukleotid. Jednotlivé hodnoty jsou zobrazeny (Tabulka 12 a Tabulka 13).

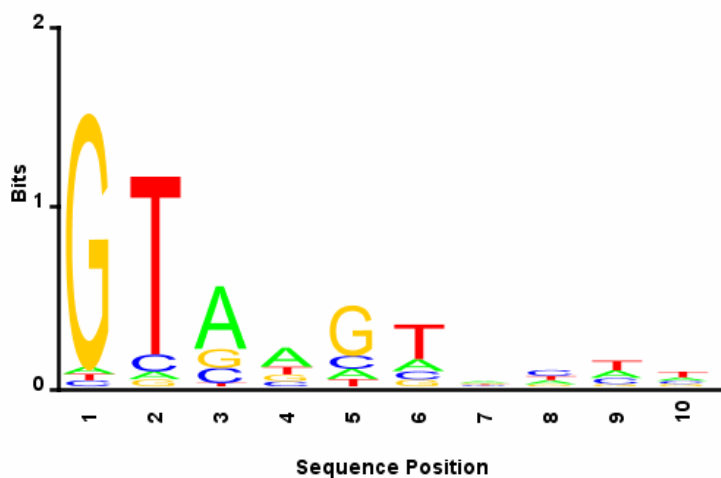
Tabulka 12 Absolutní počty jednotlivých nukleotidů na konci intronů u kura

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	110	2901	226	841	137	250	358	328	434	400
C	223	393	2253	1088	889	1439	1336	1162	1129	1065
G	2961	99	222	1044	501	414	345	620	452	485
T	155	56	748	476	1922	1346	1410	1339	1434	1499

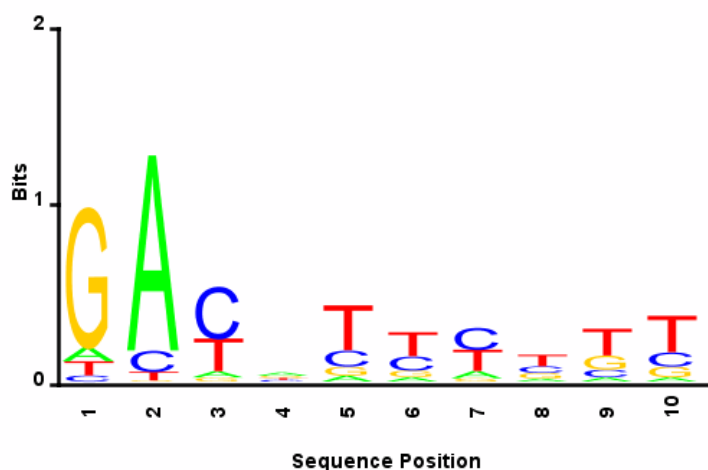
Tabulka 13 Procentuální zastoupení jednotlivých nukleotidů na konci intronů u kura

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	3,19	84,11	6,55	24,38	3,97	7,25	10,38	9,51	12,58	11,60
C	6,47	11,39	65,32	31,55	25,78	41,72	38,74	33,69	32,73	30,88
G	85,85	2,87	6,44	30,27	14,53	12,00	10,00	17,98	13,11	14,06
T	4,49	1,62	21,69	13,80	55,73	39,03	40,88	38,82	41,58	43,46

Analýza byla potvrzena vytvořením sekvenčního loga, viz Obrázek 10 a Obrázek 11.



Obrázek 10 Sekvenční logo začátku intronů u kura



Obrázek 11 Sekvenční logo konce intronů u kura

6.1.4 *Apis mellifera*

Celkově bylo analýzou zpracováno 993 intronů.

Konzervativnost dinukleotidové sekvence GT na počátku nukleotidu se potvrdila. Nukleotid G byl přítomen na první pozici v téměř 79 % případů, nukleotid T na druhé pozici ve více než 80 % případů. Kromě těchto dvou nukleotidů se ve více než padesáti procentech případů na dalších, tzn. na třetím, až šestém místě od počátku nacházely nukleotidy AAGT, jak ukazuje Tabulka 14 a Tabulka 15.

Tabulka 14 Absolutní počty jednotlivých nukleotidů na začátku intronů u včely

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	101	107	613	601	229	277	410	351	358	338
C	39	56	51	93	68	107	127	132	152	187
G	781	32	189	70	506	89	152	130	108	114
T	72	798	140	229	190	520	304	380	375	354

Tabulka 15 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů u včely

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	10,17	10,78	61,73	60,52	23,06	27,90	41,29	35,35	36,05	34,04
C	3,93	5,64	5,14	9,37	6,85	10,78	12,79	13,29	15,31	18,83
G	78,65	3,22	19,03	7,05	50,96	8,96	15,31	13,09	10,88	11,48
T	7,25	80,36	14,10	23,06	19,13	52,37	30,61	38,27	37,76	35,65

Konzervativnost sekvence AG byla také potvrzena. Procentuálně se nukleotid A vyskytoval na předposlední pozici v 85 % případů a nukleotid G na posledním místě v téměř 80 % případů. Na páté a šesté pozici od konce se vyskytuje s vysokou pravděpodobností nukleotid T. Na pozicích tři a sedm až deset jsou přítomny stejné nukleotidy u téměř 50 % intronů (Tabulka 16 a Tabulka 17).

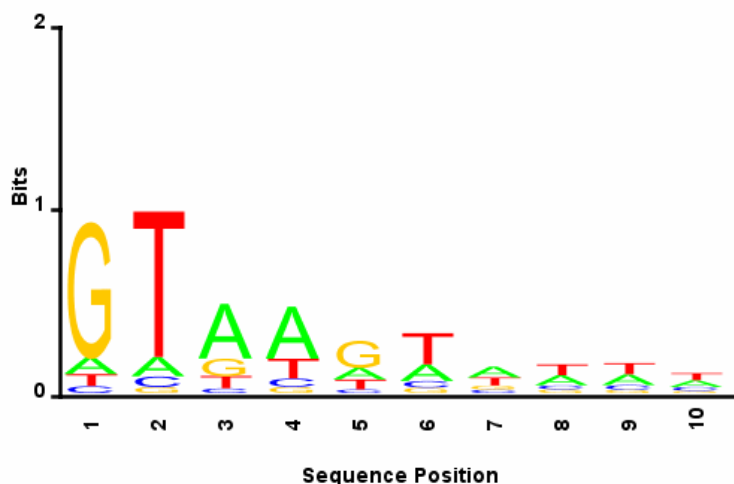
Tabulka 16 Absolutní počty jednotlivých nukleotidů na konci intronů u včely

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	90	845	150	338	156	177	217	237	272	253
C	60	42	490	152	154	123	139	148	118	146
G	790	38	51	173	66	110	158	148	122	103
T	53	68	302	330	617	583	479	460	481	491

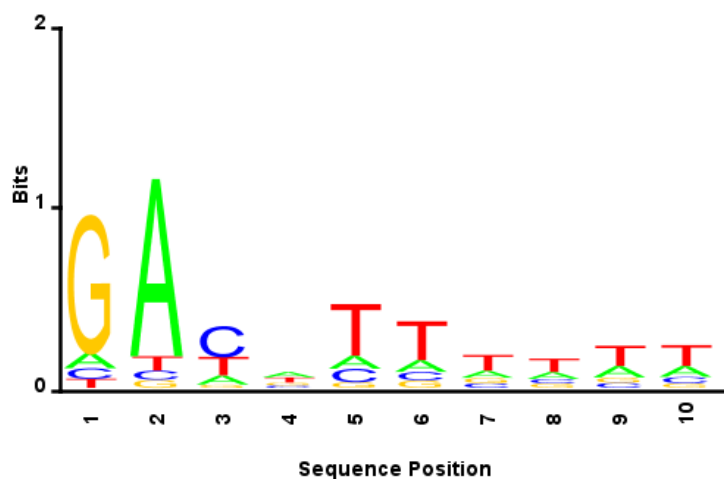
Tabulka 17 Procentuální počty jednotlivých nukleotidů na konci intronů u včely

	pozice nukleotidu vzhledem k CDS									
	1	2	3	4	5	6	7	8	9	10
A	9,06	85,10	15,11	34,04	15,71	17,82	21,85	23,87	27,39	25,48
C	6,04	4,23	49,35	15,31	15,51	12,39	14,00	14,90	11,88	14,70
G	79,56	3,83	5,14	17,42	6,65	11,08	15,91	14,90	12,29	10,37
T	5,34	6,85	30,41	33,23	62,13	58,71	48,24	46,32	48,44	49,45

Analýza byla potvrzena vytvořením sekvenčního loga, viz Obrázek 12 a Obrázek 13.



Obrázek 12 Sekvenční logo začátku intronů u kura



Obrázek 13 Sekvenční logo konce intronů u kura

6.2 Analýza hraniční oblasti intronů – exonů

Vstupem do analýzy hraničních oblastí byl soubor **boundary.mat**. Pro přehlednější zobrazení v tabulkách uvádím pouze hodnoty pro část hraniční oblasti tvořenou začátky a konci exonů. Začátky a konce intronů byly uvedeny výše v analýze GT a AG.

6.2.1 Homo sapiens

Hraniční oblast exonu nevykazuje žádnou konzervativnost na rozdíl od hraniční oblasti intronu, s výjimkou předposlední pozice na konci exonu, kde se v 62% případů vyskytuje nukleotid A (viz Tabulka 18, Tabulka 19, Tabulka 20 a Tabulka 21).

Tabulka 18 Absolutní počty nukleotidů na začátku exonů u člověka

	pozice nukleotidu vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	204	167	156	173	175	172	190	188	158	178	177
C	83	109	108	136	142	145	137	136	138	134	143
G	273	94	186	131	131	177	160	147	160	164	151
T	54	244	164	174	166	120	127	143	158	138	143

Tabulka 19 Procentuální zastoupení nukleotidů na začátku exonů u člověka

	pozice nukleotidu vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	33,22	27,20	25,41	28,18	28,50	28,01	30,94	30,62	25,73	28,99	28,83
C	13,52	17,75	17,59	22,15	23,13	23,62	22,31	22,15	22,48	21,82	23,29
G	44,46	15,31	30,29	21,34	21,34	28,83	26,06	23,94	26,06	26,71	24,59
T	8,79	39,74	26,71	28,34	27,04	19,54	20,68	23,29	25,73	22,48	23,29

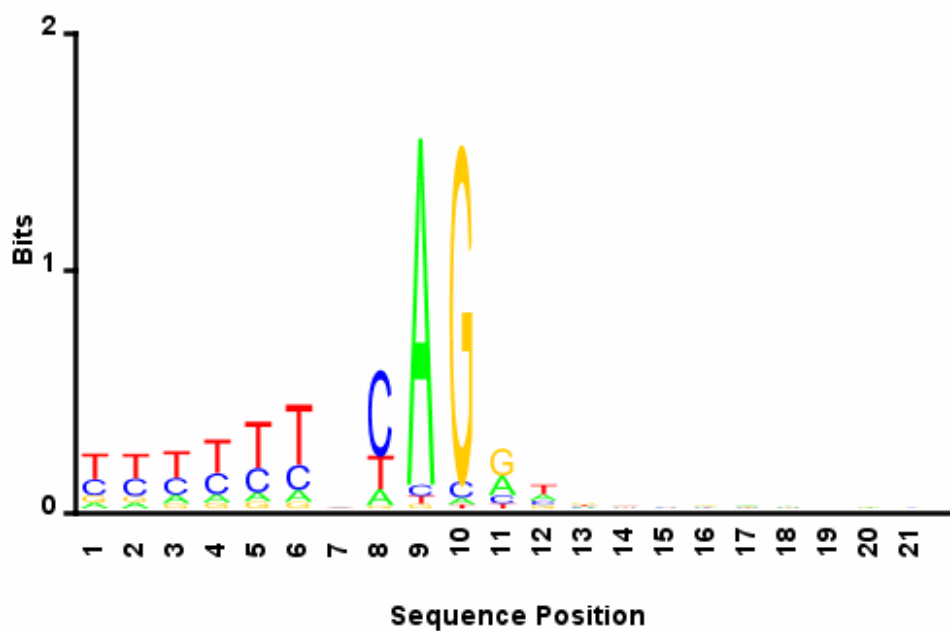
Tabulka 20 Absolutní počty nukleotidů na konci exonů u člověka

	pozice nukleotidu vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	185	177	185	183	184	173	197	192	195	382	117
C	141	158	143	158	133	130	148	154	200	70	19
G	129	147	153	109	147	154	116	139	101	84	442
T	160	133	134	165	151	158	154	130	119	79	37

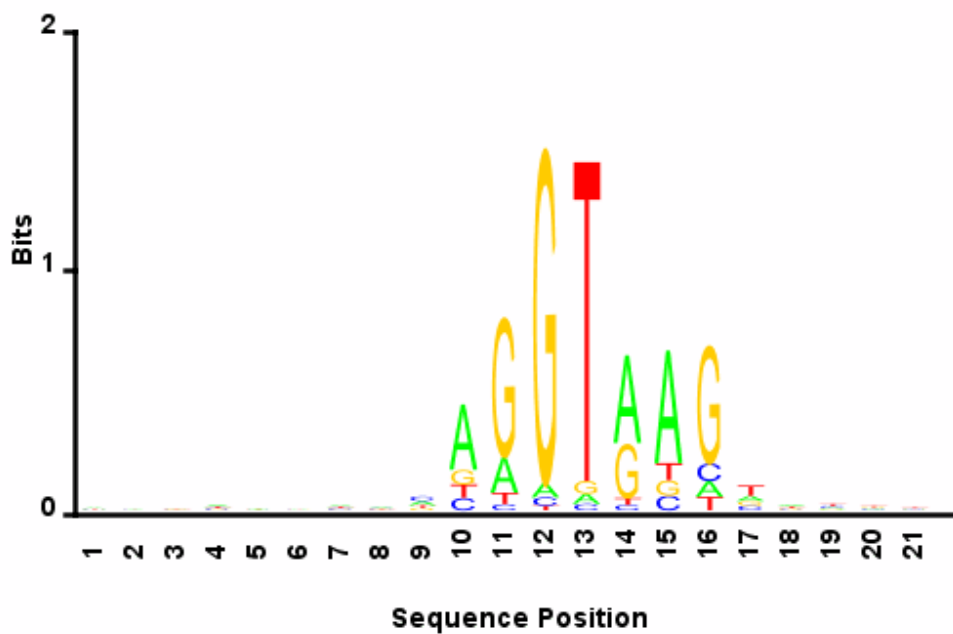
Tabulka 21 Procentuální zastoupení nukleotidů na konci exonů u člověka

	pozice nukleotidu vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	30,08	28,78	30,08	29,76	29,92	28,13	32,03	31,22	31,71	62,11	19,02
C	22,93	25,69	23,25	25,69	21,63	21,14	24,07	25,04	32,52	11,38	3,09
G	20,98	23,90	24,88	17,72	23,90	25,04	18,86	22,60	16,42	13,66	71,87
T	26,02	21,63	21,79	26,83	24,55	25,69	25,04	21,14	19,35	12,85	6,02

Analýza pomocí sekvenčního loga zobrazuje celou hraniční oblast, tzn. deset nukleotidů před a deset nukleotidů za pozicí CDS. Z výsledků je patrné, že konzervativnost vykazuje pouze oblast intronu.



Obrázek 14 Sekvenční logo začátku intronů u člověka



Obrázek 15 Sekvenční logo konce intronů u člověka

6.2.2 Mus musculus

Hraniční oblast exonu nevykazuje žádnou konzervativnost na rozdíl od hraniční oblasti intronu, s výjimkou předposlední pozice na konci exonu, kde se v 65% případech vyskytuje nukleotid A (viz Tabulka 22, Tabulka 23, Tabulka 24 a Tabulka 25).

Tabulka 22 Absolutní počty nukleotidů na začátku exonů u myši

	pozice nukleotidů vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	130	90	98	111	82	104	103	90	75	104	122
C	44	52	76	77	114	77	76	93	87	92	70
G	122	76	102	69	90	78	82	81	94	90	67
T	52	130	72	91	62	89	87	84	92	62	89

Tabulka 23 Procentuální zastoupení nukleotidů na začátku exonů u myši

	pozice nukleotidů vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	37,36	25,86	28,16	31,90	23,56	29,89	29,60	25,86	21,55	29,89	35,06
C	12,64	14,94	21,84	22,13	32,76	22,13	21,84	26,72	25,00	26,44	20,11
G	35,06	21,84	29,31	19,83	25,86	22,41	23,56	23,28	27,01	25,86	19,25
T	14,94	37,36	20,69	26,15	17,82	25,57	25,00	24,14	26,44	17,82	25,57

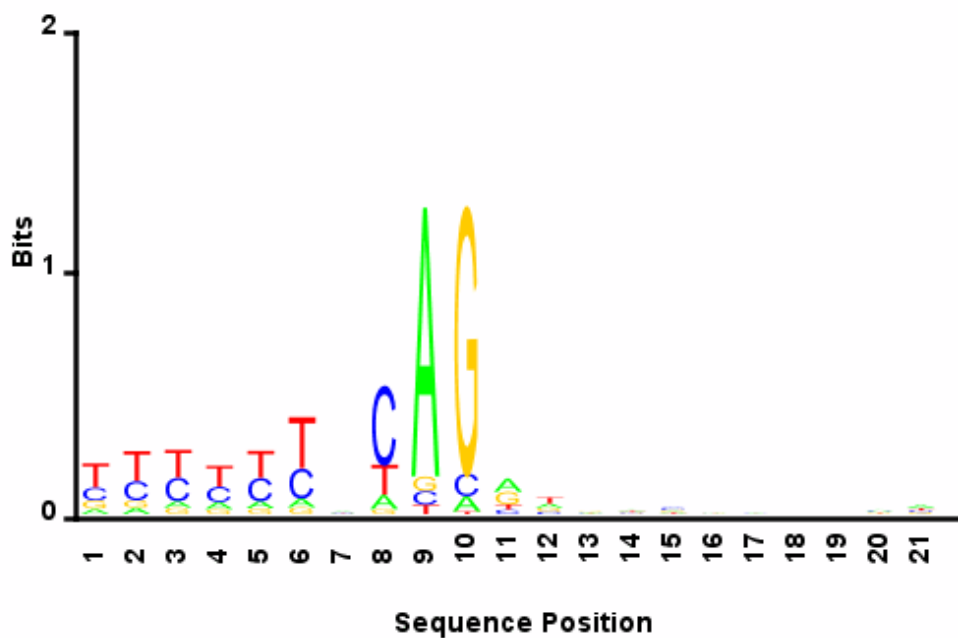
Tabulka 24 Absolutní počty nukleotidů na konci exonů u myši

	pozice nukleotidu vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	102	106	125	92	84	85	93	80	97	225	64
C	87	87	67	85	95	86	95	105	93	29	11
G	74	82	70	84	66	79	57	94	56	52	255
T	85	73	86	87	103	98	103	69	102	42	18

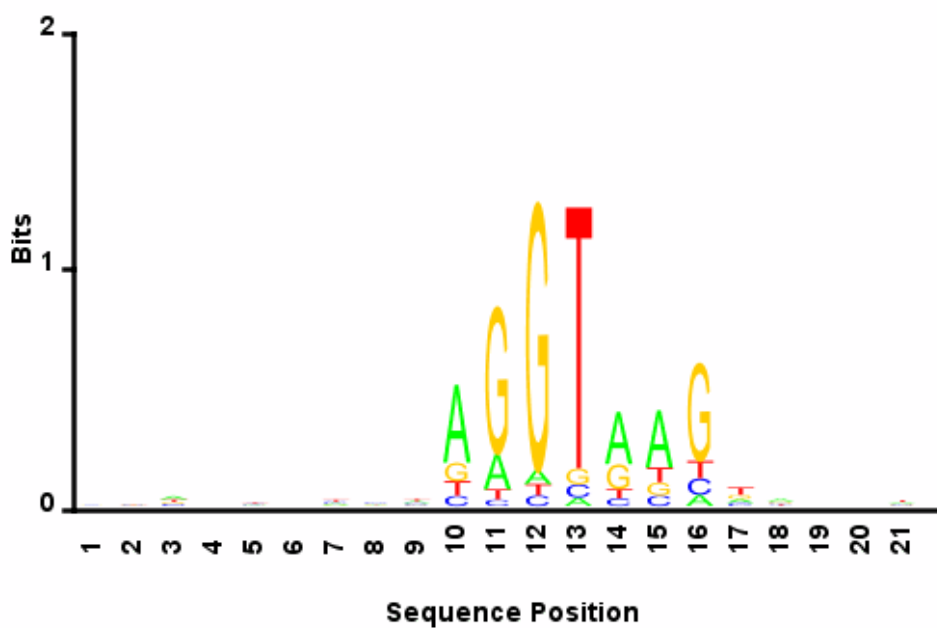
Tabulka 25 Procentuální zastoupení nukleotidů na začátku exonů u myši

	pozice nukleotidu vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	29,31	30,46	35,92	26,44	24,14	24,43	26,72	22,99	27,87	64,66	18,39
C	25,00	25,00	19,25	24,43	27,30	24,71	27,30	30,17	26,72	8,33	3,16
G	21,26	23,56	20,11	24,14	18,97	22,70	16,38	27,01	16,09	14,94	73,28
T	24,43	20,98	24,71	25,00	29,60	28,16	29,60	19,83	29,31	12,07	5,17

Analýza pomocí sekvenčního loga zobrazuje celou hraniční oblast, tzn. deset nukleotidů před a deset nukleotidů za pozicí CDS. Z výsledků je patrné, že konzervativnost vykazuje pouze oblast intronu.



Obrázek 16 Sekvenční logo začátku intronů u myši



Obrázek 17 Sekvenční logo konce intronů u myši

6.2.3 Gallus gallus

Hraniční oblast exonu nevykazuje žádnou konzervativnost na rozdíl od hraniční oblasti intronu, s výjimkou předposlední pozice na konci exonu, kde se v 61% případech vyskytuje nukleotid A (Tabulka 26, Tabulka 27, Tabulka 28 a Tabulka 29).

Tabulka 26 Absolutní počty nukleotidů na začátku exonů u kura

	pozice nukleotidu vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	174	117	114	131	157	130	139	150	117	124	145
C	80	92	86	122	131	113	132	112	119	124	112
G	199	75	179	136	92	147	152	116	156	156	130
T	64	233	138	128	137	127	94	139	125	113	130

Tabulka 27 Procentuální zastoupení nukleotidů na začátku exonů u kura

	pozice nukleotidu vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	33,66	22,63	22,05	25,34	30,37	25,15	26,89	29,01	22,63	23,98	28,05
C	15,47	17,79	16,63	23,60	25,34	21,86	25,53	21,66	23,02	23,98	21,66
G	38,49	14,51	34,62	26,31	17,79	28,43	29,40	22,44	30,17	30,17	25,15
T	12,38	45,07	26,69	24,76	26,50	24,56	18,18	26,89	24,18	21,86	25,15

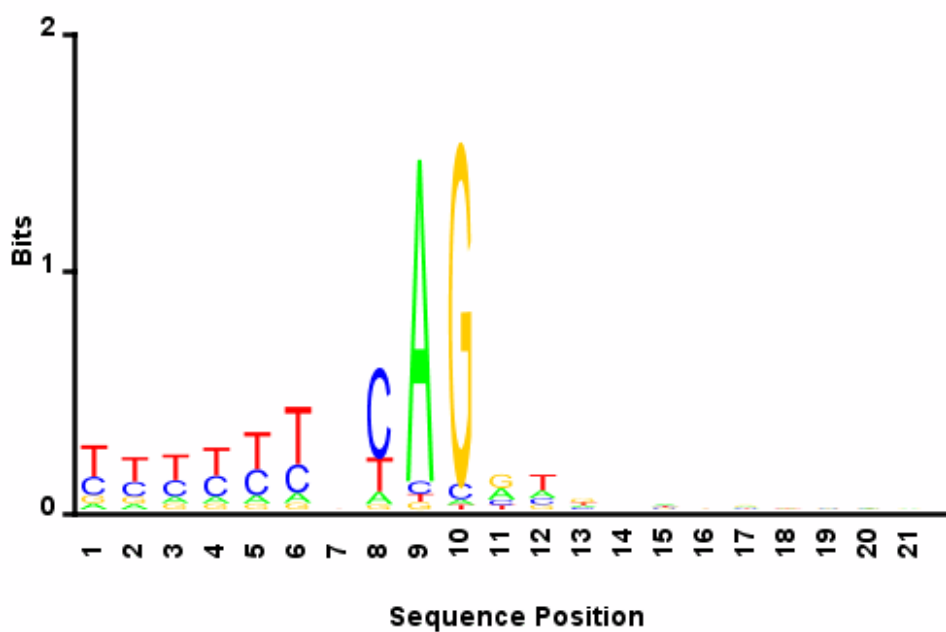
Tabulka 28 Absolutní počty nukleotidů na konci exonů u kura

	pozice nukleotidu vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	164	152	161	145	135	152	165	155	175	314	89
C	136	123	130	120	133	133	124	133	139	51	10
G	103	132	120	117	133	119	115	123	95	87	379
T	114	110	106	135	116	113	113	106	108	65	39

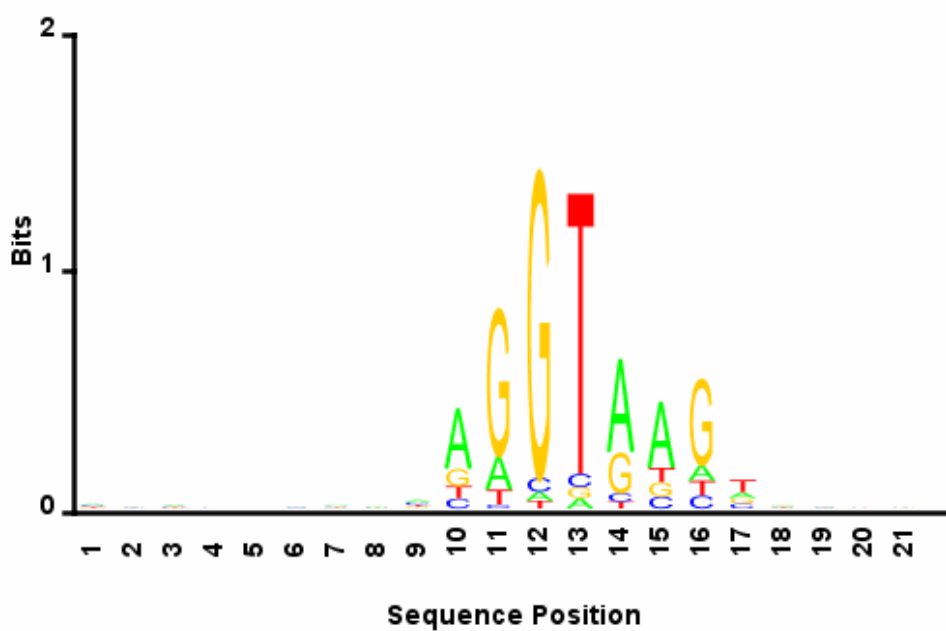
Tabulka 29 Procentuální zastoupení nukleotidů na konci exonů u kura

	pozice nukleotidu vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	31,72	29,40	31,14	28,05	26,11	29,40	31,91	29,98	33,85	60,74	17,21
C	26,31	23,79	25,15	23,21	25,73	25,73	23,98	25,73	26,89	9,86	1,93
G	19,92	25,53	23,21	22,63	25,73	23,02	22,24	23,79	18,38	16,83	73,31
T	22,05	21,28	20,50	26,11	22,44	21,86	21,86	20,50	20,89	12,57	7,54

Analýza pomocí sekvenčního loga zobrazuje celou hraniční oblast, tzn. deset nukleotidů před a deset nukleotidů za pozicí CDS. Z výsledků je patrné, že konzervativnost vykazuje pouze oblast intronu.



Obrázek 18 Sekvenční logo hraniční oblasti na začátku exonu u kura



Obrázek 19 Sekvenční logo hraniční oblasti na konci exonu u kura

6.2.4 Apis mellifera

Hraniční oblast na začátku exonu nevykazuje na žádné pozici nadpoloviční shodu. Hraniční oblast na konci exonu vykazuje konzervativnost v 60% na poslední a předposlední pozici (viz Tabulka 30, Tabulka 31, Tabulka 32 a Tabulka 33).

Tabulka 30 Absolutní počty nukleotidů na začátku exonů u včely

	pozice nukleotidu vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	365	196	237	229	279	275	265	264	252	271	276
C	100	122	121	190	193	171	180	202	164	160	167
G	297	122	281	212	181	208	244	177	237	211	205
T	100	422	223	231	209	208	173	219	209	220	214

Tabulka 31 Procentuální zastoupení nukleotidů na začátku exonů u včely

	pozice nukleotidu vzhledem k CDS										
	CDS	1	2	3	4	5	6	7	8	9	10
A	42,34	22,74	27,49	26,57	32,37	31,90	30,74	30,63	29,23	31,44	32,02
C	11,60	14,15	14,04	22,04	22,39	19,84	20,88	23,43	19,03	18,56	19,37
G	34,45	14,15	32,60	24,59	21,00	24,13	28,31	20,53	27,49	24,48	23,78
T	11,60	48,96	25,87	26,80	24,25	24,13	20,07	25,41	24,25	25,52	24,83

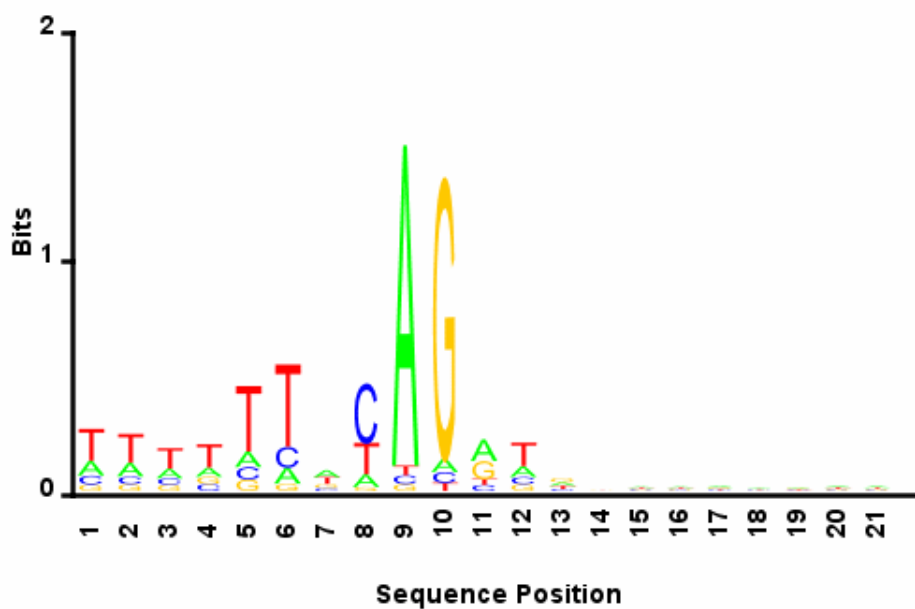
Tabulka 32 Absolutní počty nukleotidů na konci exonů u včely

	pozice nukleotidu vzhledem k CDS vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	276	256	256	279	252	249	262	305	290	525	231
C	192	169	165	198	172	186	194	164	203	97	51
G	154	191	196	160	166	160	145	182	125	120	518
T	240	246	245	225	272	267	261	211	244	120	62

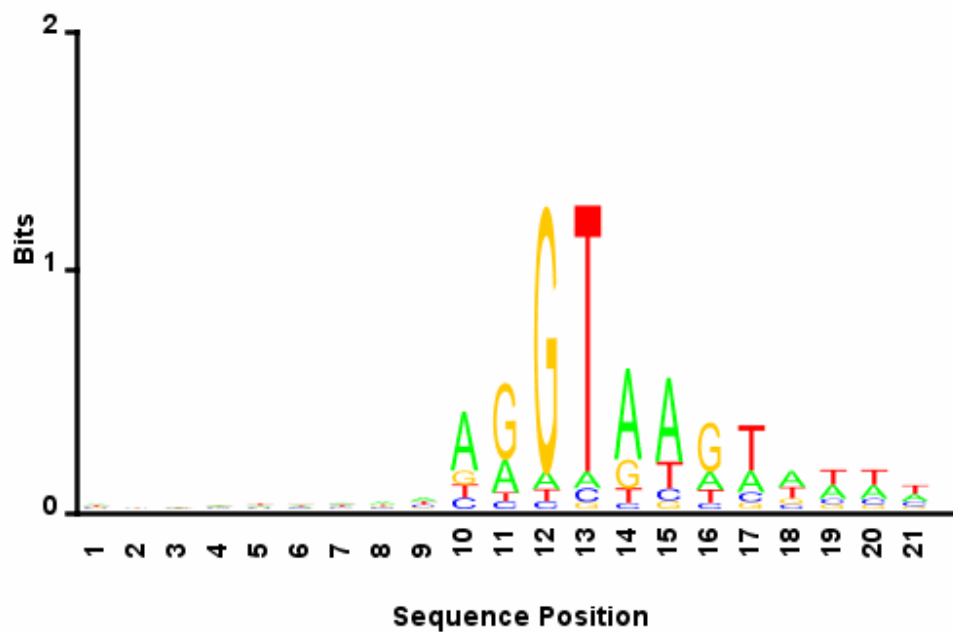
Tabulka 33 Procentuální zastoupení nukleotidů na konci exonů u včely

	pozice nukleotidu vzhledem k CDS vzhledem k CDS										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	CDS
A	32,02	29,70	29,70	32,37	29,23	28,89	30,39	35,38	33,64	60,90	26,80
C	22,27	19,61	19,14	22,97	19,95	21,58	22,51	19,03	23,55	11,25	5,92
G	17,87	22,16	22,74	18,56	19,26	18,56	16,82	21,11	14,50	13,92	60,09
T	27,84	28,54	28,42	26,10	31,55	30,97	30,28	24,48	28,31	13,92	7,19

Analýza pomocí sekvenčního loga zobrazuje celou hraniční oblast, tzn. deset nukleotidů před a deset nukleotidů za pozicí CDS. Z výsledků je patrné, že konzervativnost vykazuje pouze oblast intronu a v menší míře poslední dva nukleotidy v exonu.



Obrázek 20 Sekvenční logo hraniční oblasti na začátku exonu u včely



Obrázek 21 Sekvenční logo hraniční oblasti na konci exonu u včely

6.3 Vyhledávání konzervativních motivů

Celkově bylo u organismů zkoumáno 190 motivů. Tyto motivy byly vybrány na základě vícenásobného zarovnání sekvencí po jejich seřazení podle délky. Pro každý organismus je vložena tabulka s nejčastěji se vyskytujícími motivy pro daný organismus, počtem výskytů a jeho procentuálním vyjádřením. Zobrazeny jsou výsledky, které se vyskytovaly alespoň v 75 % intronů daného organismu.

Dále byla snaha omezit místo výskytu daného motivu v určité části intronu. Při analýze se ukázalo, že se nedá specifikovat místo výskytu ani u jednoho organismu ani pro jeden motiv.

6.3.1 Homo sapiens

Do analýzy vstupovalo 803 intronů. Tabulka 34 ukazuje přehled nejčastěji se vyskytujících motivů společně s údajem o počtu intronů, ve kterých byl celkem nalezen v absolutních a relativních hodnotách.

Tabulka 34 Nejčastěji nalezené motivy u člověka

	motiv	introny, ve kterých byl motiv vyhledán	
		počet	% z celkového počtu
1	tgcag	698	86,92
2	tttgt	698	86,92
3	ttctt	695	86,55
4	tcctt	691	86,05
5	aattt	679	84,56
6	tttaa	678	84,43
7	aaatt	674	83,94
8	tttag	670	83,44
9	agttt	666	82,94
10	gaaat	666	82,94
11	gtctt	659	82,07
12	tgtga	659	82,07
13	tcttg	658	81,94
14	agaat	655	81,57
15	cctga	652	81,2
16	gctgt	652	81,2
17	tgaat	652	81,2
18	acttt	651	81,07
19	ttgta	651	81,07

	motiv	introny, ve kterých byl motiv vyhledán	
		počet	% z celkového počtu
20	acatt	649	80,82
21	tccat	649	80,82
22	cttta	648	80,7
23	ttaat	647	80,57
24	ctgaa	644	80,20
25	agcaa	642	79,95
26	catgg	642	79,95
27	atcct	635	79,08
28	tagaa	631	78,58
29	tatgt	631	78,58
30	atcat	629	78,33
31	tcaca	629	78,33
32	tttac	625	77,83
33	cttgt	620	77,21
34	tatttt	620	77,21
35	aagtt	617	76,84
36	agagc	614	76,46
37	ttattt	610	75,97
38	tacaa	603	75,09

Box TACTAAC se u člověka podařilo najít ve 215 intronech z celkových 803, což odpovídá jeho přítomnosti v 27% případů.

6.3.2 *Mus musculus*

Do analýzy vstupovalo 405 intronů. Tabulka 35 ukazuje přehled nejčastěji se vyskytujících motivů společně s údajem o počtu intronů, ve kterých byl celkem nalezen v absolutních a relativních hodnotách.

Tabulka 35 Nejčastěji nalezené motivy u myši

	motiv	introny, ve kterých byl motiv vyhledán	
		počet	% z celkového počtu intronů
1	ttctt	356	87,9
2	tgtga	344	84,94
3	tttgt	342	84,44
4	tcttg	338	83,46
5	cctga	335	82,72
6	agttt	330	81,48
7	acatt	329	81,23
8	ctgaa	329	81,23
9	cttgt	329	81,23
10	tatgt	328	80,99
11	catgg	326	80,49
12	aattt	325	80,25
13	tccat	325	80,25
14	tgaat	325	80,25
15	atcct	324	80,00
16	tgacg	324	80,00
17	tttaa	324	80,00
18	agtcc	323	79,75
19	tcaca	322	79,51
20	agaat	321	79,26
21	cttta	321	79,26
22	aaatt	317	78,27
23	gaaat	317	78,27
24	gtctt	316	78,02
25	tagaa	315	77,78
26	ttag	315	77,78
27	agcaa	314	77,53
28	ttgta	312	77,04
29	aagtt	311	76,79
30	gaact	311	76,79
31	tttac	311	76,79
32	atcat	310	76,54
33	tcaag	309	76,3
34	gctgt	308	76,05
35	ttaat	308	76,05
36	ttata	305	75,31

Box TACTAAC se u myši podařilo najít ve 118 intronech z celkových 405, což odpovídá jeho přítomnosti v 29% případů.

6.3.3 *Gallus gallus*

Do analýzy vstupovalo 3516 intronů. Tabulka 36 ukazuje přehled nejčastěji se vyskytujících motivů společně s údajem o počtu intronů, ve kterých byl celkem nalezen v absolutních a relativních hodnotách.

Tabulka 36 Nejčastěji nalezené motivy u kura

	motiv	introny, kde byl motiv vyhledán	
		počet	% z celkového počtu intronů
1	tgcag	2716	77,25
2	tcctt	2656	75,54

Box TACTAAC se u kura podařilo najít v 438 intronech z celkových 3516, což odpovídá jeho přítomnosti v 12% případů.

6.3.4 Apis mellifera

Do analýzy vstupovalo 993 intronů. Tabulka 37 ukazuje přehled nejčastěji se vyskytujících motivů společně s údajem o počtu intronů, ve kterých byl celkem nalezen v absolutních a relativních hodnotách.

Tabulka 37 Nejčastěji nalezené motivy u včely

	motiv	introny, kde byl motiv vyhledán	
		počet	% z celkového počtu intronů
1	aattt	894	90,03
2	aaatt	882	88,82
3	tttaa	863	86,91
4	ttaat	862	86,81
5	ttata	851	85,70
6	tatttt	808	81,37
7	ttattt	776	78,15
8	agat	774	77,95
9	aaatat	762	76,74

Box TACTAAC se u kura podařilo najít v 227 intronech z celkových 993, což odpovídá jeho přítomnosti v 29 % případů.

Závěr

Bakalářská práce se zaměřila na analýzu variability intronů. Introny jsou negenové úseky DNA, o jejichž původu, významu a struktuře se mnoho neví. Zatím je zřejmé, že se jedná o velice různorodé oblasti DNA, které tvoří větší část DNA. Literatura popisuje některé oblasti, které se jeví konzervativně a které by snad, v případě mutací, mohly hrát roli při vzniku některých onemocněních.

Pro vlastní analýzu byly využity geny člověka a tří dalších organismů, z nichž byly na základě údajů zjištěných z databáze NCBI extrahovány introny. Celkem bylo analyzováno 5 650 intronů z 31 chromozomů.

Práce se v první řadě zabývala ověřením konzervativnosti počátku a konce intronů. Literatura udává, že introny jsou nejčastěji uvozeny sekvencí GT a ukončeny sekvencí AG. Analýza potvrdila, že tyto dinukleotidové sekvence konzervativní jsou u všech zkoumaných organismů. Nukleotidy se nacházely na odpovídajících místech v sedmdesáti až téměř devadesáti procentech případů. U člověka, kura a včely se na počátku intronu po GT nachází AAG, u myši na třetí pozici shoda nebyla zjištěna, na čtvrté a páté se pak vyskytuje AG přibližně stejně často jako u ostatních. Dále se u všech zkoumaných organismů přibližně stejně často vyskytuje na konci intronu před sekvencí AG ještě C, a to přibližně v polovině intronů.

Analýza hraniční oblasti neposkytla příliš nových informací. Ověřila platnost konzervativnosti GT a AG sekvencí, konzervativnost byla vysledována pouze v intronech a na předposlední pozici exonů, kde se shoda pohybovala okolo 60 % u všech organismů. Toto zjištění by odpovídalo přítomnosti terminačních kodonů.

Další část byla věnovaná vyhledávání a ověřování konzervativnosti motivů. Pro získání motivů jsme vícenásobně zarovnali některé introny zkoumaných organismů seřazených podle délky a takto získané motivy jsme zkusili najít ve všech intronech obsažených v souboru **introny.mat**. Postupně jsme zkoumali potenciální konzervativnost 190 motivů o délce mezi pěti a deseti nukleotidy. Nejmenší množství početně zastoupených motivů jsme našli u kura bankivského, což bylo zřejmě dáno tím, že právě z genů kura jsme získali největší množství intronů, jejichž počet řádově ostatní převyšuje. Z prvních pěti nejčastěji se vyskytujících motivů u člověka najdeme dva ve vysokém procentu i u myši a kura – jedná se o motivy TGCAG a TTCTT. Tyto motivy se nalézají u těchto tří organismů alespoň v 70% intronů.

Dalšími analýzami jsme chtěli najít podobnost v pozici vyhledaných sekvencí uvnitř intronů. Zvolili jsme dva postupy – rozdělení intronu do úseků o deseti procentech délky a zhodnocení prvních a posledních 100 nukleotidů v každém intronu. Ani jedna cesta však nevedla k lokalizaci a upřesnění výskytu některého z motivů.

Přítomnost konzervativního boxu TACTAAC se nám podařit neprokázalo, procento výskytu nepřekročilo 30 %.

Je pravděpodobné, že se vzrůstajícím počtem analyzovaných intronů by se výsledky mohly začít více blížit předpokládaným hodnotám, je také možné, že by se také dala vysledovat určitá pravidla, charakterizující geny, ve kterých se konzervativní úseky nacházejí.

Z provedených analýz je zcela zřejmé, že introny jsou velice variabilní oblasti genomu. Tuto různorodost vykazují v mnoha parametrech – v délce jednotlivých intronů, v počtu intronů přítomných v jednom genu nebo v zastoupení nukleotidů v jednotlivých intronech. Zdá se, že jediným opravdu konzervativním úsekem budou počáteční a koncové oblasti. Je možné, že podrobnějším a efektivnějším zkoumáním většího počtu intronů se podaří najít nové úseky podobné mezi více introny a že se díky nim podaří specifikovat význam intronů. To však dalece přesahuje možnosti této práce.

Seznam použité literatury

- [1] SNUSTAD, D a Michael J SIMMONS. Genetika. 5th ed. Překlad Jiřina Relichová. Brno: Masarykova univerzita, 2009, xxi, 871 s. ISBN 978-802-1048-522.
- [2] KATAYAMA a YUJIRO TOMOSHIGE. Base Selection Method for Recognition of Exon-Intron Boundaries. MAMECTIS'08 Proceedings of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems. 2008, s. 283-288. ISSN 1790-2769. Dostupné z: <http://www.wseas.us/e-library/conferences/2008/corfu/mnw/mnw44.pdf>.
- [3] LODISH, Harvey, Arnold BERK, Paul MATSUDAIRA, Chris A. KAISER, Monty KRIEGER, Matthew P. SCOTT, Lawrence ZIPURSKY a James DARNELL. 6th ed. New York: W.H. Freeman and Company, c2008, xxxvii, 1150 s. ISBN 9781429203142.
- [4] ALBERTS, Bruce. Molecular biology of the cell. 5th ed. New York, N.Y.: Garland science, c2008, xxxiii, 1268 s., [89] s. příl. ISBN 9780815341055.
- [5] VUKUSIC, Ivana, Sushma Nagaraja GRELLSCHEID a Thomas WIEHE. Applying genetic programming to the prediction of alternative mRNA splice variants. Science Direct: Genomics. 2007, s. 471-479. Dostupné z: <http://brendelgroup.org/WikiVB/data/media/home/research/ymzhang/gp.pdf> GP-based methods for extracting exons from dna semente
- [6] OHTA, Takehiro, Ikuo YOSHIHARA, Kunihiro YAMAMORI a Moritoshi YASUNAGA. GP-Based Method for Extracting Exons from DNA Sequence. 2003. Dostupné z: <http://ir.lib.miyazaki-u.ac.jp/dspace/bitstream/10458/280/3/KJ00002426418.pdf>
- [7] ROGOZIN, Ivan B., Liran CARMEL, Miklos CSUROS a Eugene V. KOONIN. Origin and evolution of spliceosomal introns. 2012. DOI: 10.1186/1745-6150-7-11. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3488318/>
- [8] BOYLE, Roger D. Hidden Markov Models [online]. [cit. 2012-12-13]. Dostupné z: http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/main.html

- [9] GELFMAN, Sahar, David BURSTEIN, Osnat PENN, Anna SAVCHENKO, Maayan AMIT, Schraga SCHWARTZ, Tal PUPKO a Gil AST. Changes in exon–intron structure during vertebrate evolution affect the splicing pattern of exons. 2011. DOI: 10.1101/gr.119834.110. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3246205/>
- [10] IWATA, Hiroaki a Osamu GOTOH. Comparative analysis of information contents relevant to recognition of introns in many species. 2011. DOI: 10.1186/1471-2164-12-45. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3033335/>
- [11] STABELL, Fredrik B., Nicolas J. TOURASSE a Anne-Brit KOLSTØ. A conserved 3' extension in unusual group II introns is important for efficient second-step splicing. 2009. DOI: 10.1093/nar/gkp186. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2691827/>
- [12] IRIMIA, Manuel a Scott William ROY. Spliceosomal introns as tools for genomic and evolutionary analysis. 2008. DOI: 10.1093/nar/gkn012. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2275149/>
- [13] FEDOROV, Alexei, Amir Feisal MERICAN a Walter GILBERT. Large-scale comparison of intron positions among animal, plant, and fungal genes. 2002. DOI: 10.1073/pnas.242624899. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC138576/>
- [14] COLLEGE OF ARTS AND SCIENCES, The University of Tokyo. A comprehensive approach to Live Science: Nucleic Acids [online]. 2011. vyd. [cit. 2012-12-13]. Dostupné z: http://csls-text3.c.u-tokyo.ac.jp/active/06_05.html
- [15] TWYMAN, Richard. Gene expression. [online]. [cit. 2012-12-13]. Dostupné z: http://genome.wellcome.ac.uk/doc_WTD020757.html
- [16] Chapter 5. Genetic Code, Translation, Splicing. KAP BIOLOGY DEPT KENYON COLLEGE. [online]. [cit. 2012-12-13]. Dostupné z: <http://biology.kenyon.edu/courses/biol114/Chap05/Chapter05.html>
- [17] VAIDYANATHAN, P. P. Genomics and Proteomics: A Signal Processor's Tour. 2004. Dostupné z: <http://authors.library.caltech.edu/9665/1/VAIieecsm04.pdf>

[18] BioLib.cz [online]. © 1999-2013 [cit. 2013-05-30]. Dostupné z:
<http://www.biolib.cz/cz/main/>

[19] CROOKS, Gavin E., Gary HON, John-Marc CHADONIA a Steven BRENNER.
WebLogo: A Sequence Logo Generator. *Genome Research*, 14:1188-1190, (2004) Dostupné
z: <http://weblogo.berkeley.edu/Crooks-2004-GR-WebLogo.pdf>

[20] THE MATHWORKS, Inc. MathWorks [online]. © 1994-2013 [cit. 2013-05-30].
Dostupné z: <http://www.mathworks.com/solutions/>

Seznam použitých zkratk

DNA	deoxyribonukleová kyselina
RNA	ribonukleová kyselina
A	adenin
T	thymin
G	guanin
C	cytosin
NCBI	National Center for Biotechnology Information

Seznam obrázků

Obrázek 1 Primární struktura DNA a RNA [14].....	10
Obrázek 2 Dvouvláknová DNA [14]	11
Obrázek 3 Genová exprese [15]	13
Obrázek 4 Ribozom a translace [16]	17
Obrázek 5 Blokové schéma.....	23
Obrázek 6 Sekvenční logo začátku intronů u člověka	27
Obrázek 7 Sekvenční logo konce intronů u člověka	27
Obrázek 8 Sekvenční logo začátku intronů u myši	29
Obrázek 9 Sekvenční logo konce intronů u myši.....	29
Obrázek 10 Sekvenční logo začátku intronů u kura.....	31
Obrázek 11 Sekvenční logo konce intronů u kura	31
Obrázek 12 Sekvenční logo začátku intronů u kura.....	33
Obrázek 13 Sekvenční logo konce intronů u kura	33
Obrázek 14 Sekvenční logo začátku intronů u člověka	35
Obrázek 15 Sekvenční logo konce intronů u člověka	35
Obrázek 16 Sekvenční logo začátku intronů u myši	37
Obrázek 17 Sekvenční logo konce intronů u myši.....	37
Obrázek 18 Sekvenční logo hraniční oblasti na začátku exonu u kura	39
Obrázek 19 Sekvenční logo hraniční oblasti na konci exonu u kura	39
Obrázek 20 Sekvenční logo hraniční oblasti na začátku exonu u včely	41
Obrázek 21 Sekvenční logo hraniční oblasti na konci exonu u včely.....	41

Seznam tabulek

Tabulka 1 Taxonomické zařazení uvedených organismů	25
Tabulka 2 Absolutní počty jednotlivých nukleotidů na začátku intronů u člověka	26
Tabulka 3 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů u člověka	26
Tabulka 4 Absolutní počty jednotlivých nukleotidů na konci intronů u člověka	26
Tabulka 5 Procentuální zastoupení jednotlivých nukleotidů na konci intronů u člověka.....	26
Tabulka 6 Absolutní počty jednotlivých nukleotidů na začátku intronů u myši.....	28
Tabulka 7 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů u myši	28
Tabulka 8 Absolutní počty jednotlivých nukleotidů na konci intronů u myši	28
Tabulka 9 Procentuální zastoupení jednotlivých nukleotidů na konci intronů u myši	28
Tabulka 10 Absolutní počty jednotlivých nukleotidů na začátku intronů kura	30
Tabulka 11 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů kura.....	30
Tabulka 12 Absolutní počty jednotlivých nukleotidů na konci intronů u kura.....	30
Tabulka 13 Procentuální zastoupení jednotlivých nukleotidů na konci intronů u kura	30
Tabulka 14 Absolutní počty jednotlivých nukleotidů na začátku intronů u včely.....	32
Tabulka 15 Procentuální zastoupení jednotlivých nukleotidů na začátku intronů u včely	32
Tabulka 16 Absolutní počty jednotlivých nukleotidů na konci intronů u včely	32
Tabulka 17 Procentuální počty jednotlivých nukleotidů na na konci intronů u včely	32
Tabulka 18 Absolutní počty nukleotidů na začátku exonů u člověka.....	34
Tabulka 19 Procentuální zastoupení nukleotidů na začátku exonů u člověka	34
Tabulka 20 Absolutní počty nukleotidů na konci exonů u člověka	34
Tabulka 21 Procentuální zastoupení nukleotidů na konci exonů u člověka	34
Tabulka 22 Absolutní počty nukleotidů na začátku exonů u myši	36
Tabulka 23 Procentuální zastoupení nukleotidů na začátku exonů u myši.....	36
Tabulka 24 Absolutní počty nukleotidů na konci exonů u myši.....	36
Tabulka 25 Procentuální zastoupení nukleotidů na začátku exonů u myši.....	36
Tabulka 26 Absolutní počty nukleotidů na začátku exonů u kura	38
Tabulka 27 Procentuální zastoupení nukleotidů na začátku exonů u kura.....	38
Tabulka 28 Absolutní počty nukleotidů na konci exonů u kura	38
Tabulka 29 Procentuální zastoupení nukleotidů na konci exonů u kura.....	38
Tabulka 30 Absolutní počty nukleotidů na začátku exonů u včely.....	40
Tabulka 31 Procentuální zastoupení nukleotidů na začátku exonů u včely	40
Tabulka 32 Absolutní počty nukleotidů na konci exonů u včely	40
Tabulka 33 Procentuální zastoupení nukleotidů na konci exonů u včely	40
Tabulka 34 Nejčastěji nalezené motivy u člověka	42
Tabulka 35 Nejčastěji nalezené motivy u myši.....	43
Tabulka 36 Nejčastěji nalezené motivy u kura	44
Tabulka 37 Nejčastěji nalezené motivy u včely	44