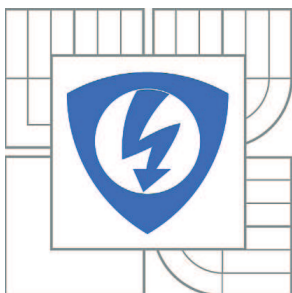


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF BIOMEDICAL ENGINEERING

VYHLEDÁVÁNÍ CPG OSTRŮVKŮ V GENOMU EUKARYOTICKÝCH ORGANISMŮ

IDENTIFYING CPG ISLANDS IN GENOMES OF EUKARYOTES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

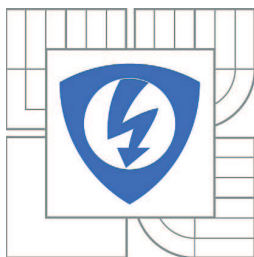
KATEŘINA PŘIKRYLOVÁ

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. DENISA MADĚRÁNKOVÁ

BRNO 2012



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav biomedicínského inženýrství

Bakalářská práce

bakalářský studijní obor

Biomedicínská technika a bioinformatika

Studentka: Kateřina Přikrylová

ID: 124999

Ročník: 3

Akademický rok: 2011/2012

NÁZEV TÉMATU:

Vyhledávání CpG ostrůvků v genomu eukaryotických organismů

POKYNY PRO VYPRACOVÁNÍ:

1) Zpracujte literární rešerši na téma CpG ostrůvků v genomech eukaryotických organismů. 2) Zaměřte se také na existující algoritmy pro vyhledávání CpG ostrůvků a porovnejte jejich možnosti a parametry. 3) Vybraný algoritmus realizujte v prostředí Matlab. 4) Funkčnost realizovaného algoritmu vyzkoušejte na vybrané části genomu člověka a šimpanze. 5) Pomocí realizovaného algoritmu proveďte analýzu CpG ostrůvků v částech genomů alespoň tří eukaryotických organismů. 6) Výsledky zhodnoťte a porovnejte s výsledky vybrané volně přístupné aplikace pro vyhledávání CpG ostrůvků.

DOPORUČENÁ LITERATURA:

[1] ANTEQUERA, F. Structure, function and evolution of CpG island promoters. Cell. Mol. Life Sci. 2003, vol. 60, p. 1647-1658.

[2] GARDINER-GARDEN, M., Frommer, M. CpG islands in vertebrate genomes. J. Mol. Biol. 1987, vol. 196., p. 261-282.

Termín zadání: 6.2.2012

Termín odevzdání: 25.5.2012

Vedoucí práce: Ing. Denisa Maděránková

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 zabývá vyhledáváním CpG ostrůvků v DNA. V teoretické části je popsána struktura a složení deoxyribonukleové kyseliny, chromosomů a genů. Jsou zde uvedeny základní informace o lidském a šimpanzím genomu a vysvětlení CpG ostrůvků. Praktickou část tvoří program, který vyhledává úseky CpG ostrůvků v sekvencích, zobrazí je a uloží. Součástí práce je analýza genů společných u Homo sapiens, Pan troglodytes, Mus musculus a Bos taurus.

KLÍČOVÁ SLOVA

DNA, gen, genom, sekvence, cytosin, guanin, CpG ostrůvky, člověk, šimpanz.

ABSTRACT

Bachelor thesis deals with searching CpG islands in DNA. The theoretical part describes the structure and composition of deoxyribonucleic acid, chromosomes and genes. There is basic information about human and chimpanzees genome and specification of CpG islands. Practical part consists of a program that searches for CpG islands in sequences, it displays and saves it. This work includes analysis of genes which are common for Homo sapiens, Pan troglodytes, Mus musculus and Bos taurus.

KEYWORDS

DNA, gene, genome, sequence, cytosine, guanine, CpG islands, human, chimpanzee.

PŘIKRYLOVÁ, K. Vyhledávání CpG ostrůvků v genomu eukaryotických organismů. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2012. 52 s. Vedoucí bakalářské práce Ing. Denisa Maděránková.

Prohlášení

Prohlašuji, že svou bakalářskou práci na téma „Vyhledávání CpG ostrůvků v genomu eukaryotických organismů“ 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 bakalářské práce dále prohlašuji, že v souvislosti s vytvořením tohoto projektu jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne 25. května 2012

.....
podpis autora

Poděkování

Děkuji vedoucímu bakalářské práce Ing. Denise Maděránkové za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé bakalářské práce.

V Brně dne 25. května 2012

.....
podpis autora

Obsah

1	ÚVOD	9
2	TEORETICKÁ ČÁST.....	10
2.1	DNA.....	10
2.1.1	Primární struktura DNA	11
2.1.2	Párování bází	11
2.1.3	Sekundární struktura DNA	12
2.2	GENETICKÁ INFORMACE a GEN	13
2.2.1	Strukturní geny	15
2.2.2	Geny pro RNA.....	15
2.2.3	Regulační oblasti	15
2.2.4	Překrývající se geny.....	16
2.3	GENOM.....	16
2.3.1	Chromosomy	16
2.3.2	Lidský genom	18
2.3.3	Šimpanzí genom	19
2.4	CpG ostrůvky	20
2.4.1	Původ a evoluce CpG ostrůvků	20
2.4.2	CpG ostrůvky a CpG poor	21
3	PRAKTICKÁ ČÁST.....	23
3.1	Program VYHLEDAVAC.m.....	23
3.1.1	Načtení sekvence	24
3.1.2	Vyhledávání CpG ostrůvků	24
3.1.3	Grafické zobrazení a uložení	26
3.2	Analýza CpG ostrůvků.....	27
3.2.1	Analýza programem Vyhledavac.m a internetovým vyhledávačem http://www.ebi.ac.uk/Tools/emboss/cpgplot/	27
3.2.2	PTPN1	28
3.2.3	PML.....	33
3.2.4	MEFV	36

3.3	Shrnutí výsledků analýzy	39
4	ZÁVĚR.....	41
5	ZDROJE.....	43
6	PŘÍLOHY.....	45

SEZNAM OBRÁZKŮ

Obr. 1: Část řetězce DNA [9]	10
Obr. 2: Párování bází Adenin – Thymin [10]	12
Obr. 3: Párování bází Guanin – Cytosin [10].....	12
Obr. 4: Sekundární struktura DNA [10]	13
Obr. 5: Schéma funkčních jednotek genetické informace [3]	14
Obr. 6: Schéma složeného strukturního genu [10].....	15
Obr. 7: Schéma spiralizace chromosomu [10]	17
Obr. 8: Dvě skupiny promotorů u člověka a myši. Červené obdélníky a šipky symbolizují první exon a transkripční zahajovací část [11].....	22
Obr. 9: Uživatelské prostředí programu Vyhledavac.m	23
Obr. 10: Předdefinované vstupní parametry	25
Obr. 11: Příklad vykreslení CpG ostrůvků	26
Obr. 12: Ukázka vykreslení nukleotidů.....	27
Obr. 13: PTPN1 u Homo sapiens v programu Vyhledavac.m	28
Obr. 14: PTPN1 u Homo sapiens v internetovém vyhledávači	29
Obr. 15: PTPN1 u Pan troglodytes v programu Vyhledavac.m.....	29
Obr. 16: PTPN1 u Pan troglodytes v internetovém vyhledávači	30
Obr. 17: PTPN1 u Bos taurus v programu Vyhledavac.m.....	30
Obr. 18: PTPN1 u Bos taurus v internetovém vyhledávači.....	31
Obr. 19: PTPN1 u Mus musculus v programu Vyhledavac.m	32
Obr. 20: PTPN1 u Mus musculus v internetovém vyhledávači	32
Obr. 21: PML u Homo sapiens v programu Vyhledavac.m.....	33
Obr. 22: PML u Homo sapiens v internetovém vyhledávači	33
Obr. 23: PML u Pan troglodytes v programu Vyhledavac.m.....	34
Obr. 24: PML u Pan troglodytes v internetovém vyhledávači.....	34
Obr. 25: PML u Bos taurus v programu Vyhledavac.m	35
Obr. 26: PML u Bos taurus v internetovém vyhledávači	35
Obr. 27: PML u Mus musculus v programu Vyhledavac.m.....	36
Obr. 28: PML u Mus musculus v internetovém vyhledávači	36
Obr. 29: MEFV u Homo sapiens v programu Vyhledavac.m	37
Obr. 30: MEFV u Homo sapiens v internetovém vyhledávači	37
Obr. 31: MEFV u Pan troglodytes v programu Vyhledavac.m.....	38
Obr. 32: MEFV u Pan troglodytes v internetovém vyhledávači	38
Obr. 33: MEFV u Bos taurus v programu Vyhledavac.m	38
Obr. 34: MEFV u Bos taurus v internetovém vyhledávači.....	39
Obr. 35: Obs/Exp v programu Vyhledavac.m.....	40

Obr. 36: Obs/Exp v internetovém vyhledávači	40
Obr. 37: Uživatelské prostředí programu	46
Obr. 38: Správné načtení sekvence	46
Obr. 39: Dialogové okno při nenalezení CpG ostrůvku	47
Obr. 40: Vykreslení grafu a zapsání pozic do listboxu.....	47
Obr. 41: Zobrazení nalezeného úseku	48

1 ÚVOD

Člověk se jakožto živý organizmus z hlediska genetiky nijak neliší od ostatních organismů na naší planetě a platí pro něj úplně stejné biologické a genetické principy. Lidská a šimpanzí DNA jsou si velice podobné, v mnoha úsecích totožné. Mnoho genů máme stejných. A přece se lišíme. Srovnávání DNA člověka a šimpanze je důležité z několika důvodů. Zásadní důležitost má hledání takových rozdílů v dědičné informaci člověka a lidoopů, které by vysvětlily podstatu našeho „lidství“. Říká se, že odlišnost člověka od ostatních živočichů je podmíněna naší schopností abstrakce. [1]

Cílem zadané práce je vyhledat v genomu eukaryotických organismů CpG ostrůvky a ty pak vyhledávat v genomu šimpanze a člověka. Existuje mnoho algoritmů, které se pro tuto problematiku používají, proto součástí práce je jejich vyzkoušení, aplikace jednoho z nich v programovém prostředí Matlab a jejich vzájemné porovnání mezi sebou.

V první části se věnuji především teoretickému úvodu do problematiky práce, která je nutná pro pochopení pozdějších principů. Zahrnuje poznatky o podstatě DNA, její struktuře, kondenzace do vyšších struktur. Zabývá se geny, jejich stavbou, umístěním a funkcí. Další kapitoly jsou věnovány lidskému a šimpanzímu genomu, včetně projektu HUGO, jehož cílem bylo lidský genom osekvenovat. V závěru teoretické části je zmíněna problematika a důležitost CpG ostrůvků v savčím genomu.

Praktická část obsahuje rozbor programu Vyhledavac.m a analýzu CpG ostrůvků pro porovnání funkce vytvořeného programu a internetového. Jedna z podkapitol se zabývá vnitřním zpracováním skriptů hlavního programu, vysvětlení na jakém základě detekuje pozice CpG ostrůvků, jaké podmínky lze uživatelsky stanovit a jaký je programový výstup. Vyhledávač umí načítat sekvence ve FASTA formátu, upravit pro zpracování a má možnost výstup uložit do tabulkového souboru MS Excel. V krátkých podkapitolách jsou krátce rozepsány geny, které byly pro genovou analýzu vybrány. Další podkapitola je věnována analýze CpG ostrůvků u čtyř eukaryotických organismů a to: Homo sapiens, Pan troglodytes, Mus musculus a Bos taurus. Všechny jsou zpracovány vytvořeným programem a porovnány s internetovým volně přístupným vyhledávačem CpG ostrůvků. V závěru analýzy je shrnutí výsledků a hodnocení v jak velké míře se od sebe vyhledávače ve výsledcích rozcházejí.

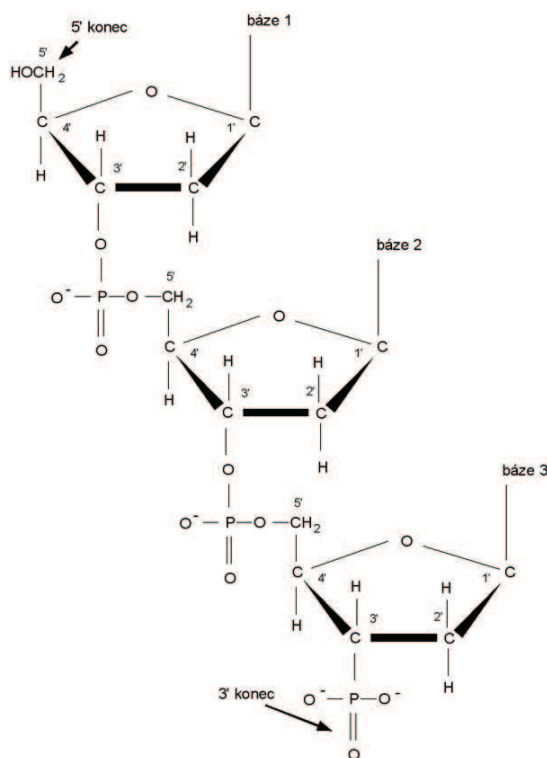
Práce je doplněna o přílohu obsahující uživatelský manuál k vytvořenému programu.

2 TEORETICKÁ ČÁST

2.1 DNA

DNA je součástí nukleových kyselin, které tvoří sice relativně malé procento hmotnosti buňky, ale svým významem v kódování genetické informace a v její expresi představují zcela nezbytný typ biopolymeru naprosto všech živých soustav. Nukleové kyseliny se skládají z monomerů, nukleotidů. Nukleotidy se spojují a navazují navzájem v polynukleotidový řetězec (vlákno). Zastoupením každého nukleotidu v polynukleotidovém řetězci, určíme jeho primární strukturu. [3]

Řetězce všech nukleových kyselin jsou složeny z pravidelně se opakující sekvence fosfát-pentosa, k nimž jsou navázány dusíkaté báze. Dusíkaté báze jsou dvojího druhu, purinové a pyrimidinové. DNA se skládá ze sacharidové složky, deoxyribosy a purinových bází, adeninu a guaninu, a pyrimidinových bází, cytosinu a thyminu.



Obr. 1: Část řetězce DNA [9]

Nukleosidy, které jsou podstatnou součástí DNA, jsou sloučeniny tvořené dusíkatou bází a monosacharidem, spojené N-glykosidovou vazbou. Deriváty purinu, adenin a guanin, nebo pyrimidinu, uracil, cytosin a thymin, tvoří báze. Jejich cukernou složkou jsou ve většině případů ribosa nebo deoxyribosa. [2] Spojením báze s deoxyribózou, vznikne deoxyribonukleosid.

Nukleotidy jsou deriváty nukleosidů, v nichž se na deoxyribosu (ribosu), váže esterovou vazbou kyselina fosforečná.

Každý nukleotid v polynukleotidovém řetězci je vázán esterickou vazbou mezi fosfátovou skupinou jednoho nukleotidu (vázanou na 3. uhlík pentózy) a pátým uhlíkem pentózy druhého nukleotidu. Díky tomuto uspořádání se v řetězci střídají kyselina fosforečná a pentóza. Báze od osy řetězce ustupují. Na jednom konci řetězce je fosfátová skupina (tzv. 5' konec) a na druhém pentóza (tzv. 3' konec). [3] Tato struktura polynukleotidu se obvykle označuje jako páteř polynukleotidu nebo též pentózafosfátová kostra.

Celkový počet pospojovaných nukleotidů, vzájemný poměr čtyř různých nukleotidů a jejich sekvence v polynukleotidovém řetězci charakterizuje jednotlivé druhy nukleových kyselin. [3]

Molekuly DNA jsou důležité proto, že jsou nositeli genetické informace, tzn. obsahují informaci o syntéze všech proteinů v organismu. [2] Určité části DNA nesou informaci o syntéze bílkovin, označují se jakou strukturní geny.

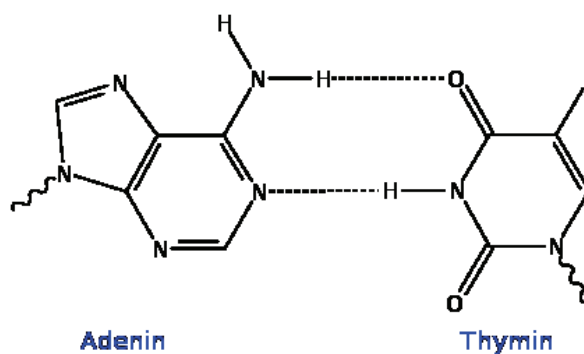
2.1.1 Primární struktura DNA

Primární strukturou DNA je myšleno seřazení (sekvence) nukleotidů, z nichž se tato molekula sestává. DNA je tvořena v převážné většině případů dvěma polynukleotidovými řetězci (dvouvláknová, či dvoupentlicová molekula), přičemž oba řetězce mají opačnou polaritu a probíhají vedle sebe. Výjimečně se u některých virů nalézá molekula DNA tvořená pouze jedním řetězcem, případně spojením tří řetězců DNA vzniká třířetězcová DNA neboli triplex a mezi čtyřmi DNA řetězci, čtyřřetězcová DNA neboli kvadruplex. U dvouřetězcové DNA probíhají oba řetězce antiparalelně, v jednom řetězci jsou báze uspořádány tak, že na jednom konci molekule je 3' konec jednoho a 5' konec druhého řetězce a ve druhém přesně opačně (ve směru 5' -> 3'). Sekvence nukleotidů v obou řetězcích (vláknech) je na sobě závislá. Oba řetězce jsou navzájem komplementární. Objevuje se zde jev, který označujeme jako párování bází. [2], [3]

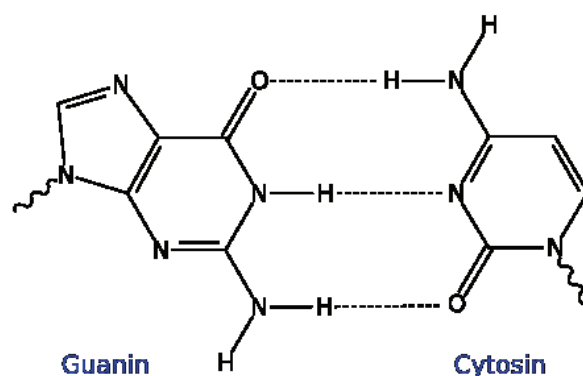
2.1.2 Párování bází

Párováním bází se rozumí spojení dvou bází vodíkovými vazbami. Vodíkové vazby vznikají mezi adeninem v aminoformě a thyminem v ketoformě (dvě vodíkové vazby) a mezi cytosinem v aminoformě a guaninem v ketoformě (tři vodíkové vazby). Obecně platí, že naproti purinové bázi leží vždy báze pyrimidinová. Díky tomu je dáno, že jeden řetězec určuje vždy sekvenci bází v druhém řetězci. Z principu párování bází je zřejmé, že počet adeninu a thyminu v molekule DNA musí být stejné ($A = T$) a stejně tak zastoupení cytosinu a guaninu ($C = G$). Každá molekula DNA má určitou důležitou charakteristiku, poměr $A+T/C+G$. Tento poměr je vždy pro daný druh DNA konstantní. Při bližším

pohledu na strukturu molekuly DNA zjistíme, že se různě střídají čtyři dvojice (páry) bází: A-T, T-A, G-C, C-G. Při popisu sekvence nukleotidů s různými bázemi (zpravidla se hovoří zkráceně o sekvenci bází) vyjadřujeme sekvenci uvedených párů bází. Teoretický počet různých sekvencí je pak 4^n , kde n je celkový počet párů. [3]



Obr. 2: Párování bází Adenin – Thymin [10]



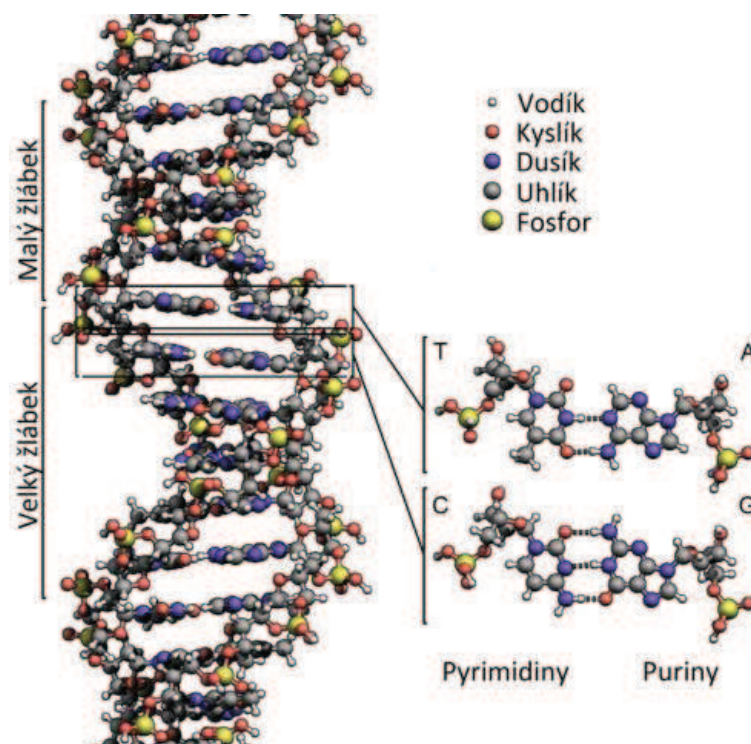
Obr. 3: Párování bází Guanin – Cytosin [10]

2.1.3 Sekundární struktura DNA

Nejčastější podoba sekundární struktury DNA je dvoušroubovice, která má své určité charakteristické rysy. Každá dvoušroubovice se skládá ze dvou polydeoxyribonukleotidových řetězců šroubovicovitě obtáčejících společnou osu neboli osu dvoušroubovice. Oba řetězce musí být vzájemně komplementární, jak vyplývá z principu o párování bází (Obr. 2 a 3). Komplementární řetězce musí být ale také antiparalelní, tzn. liší se směrem fosfodiesterové vazby. S tímto jevem se nesetkáváme jenom u DNA, ale také u všech dvouřetězcových molekul nukleových kyselin. Odborně je označován jako antiparalelizmus, orientace komplementárních polynukleotidových řetězců ve dvouřetězcových molekulách nukleových kyselin, která je charakteristická směrem k fosfodiesterových vazeb $3' \rightarrow 5'$ na jednom řetězci a $5' \rightarrow 3'$ na řetězci druhém. [5]

Vzhledem k tomu, že páry bází od osy šroubovice odstupují v určité vzdálenosti, spojnice atomů C1 komplementárních nukleotidů není na osu dvoušroubovice kolmá ani jí neprochází. Zapřičiňuje to nehladkost dvoušroubovice, ale vytváří dva žlábký různé šířky a hloubky. Poznáváme dva, menší a větší. Větší žlábek má rozměry 1,2 nm, menší 0,6 nm. Větší žlábek je hlubší než menší. Oba žlábký jsou charakteristické přítomností atomů schopných tvořit vodíkové vazby s proteiny, větší žlábek ve větší míře než žlábek menší. [5]

Základní struktura dvoušroubovice je udržována a stabilizována vnitromolekulárními silami (podobně jako u bílkovin) a vzniká automaticky (představuje stav s nejmenší možnou energií). Nejdůležitější roli v udržování stálosti sekundární struktury mají vodíkové můstky mezi komplementárními bázemi a hydrofobní reakce. [3]



Obr. 4: Sekundární struktura DNA [10]

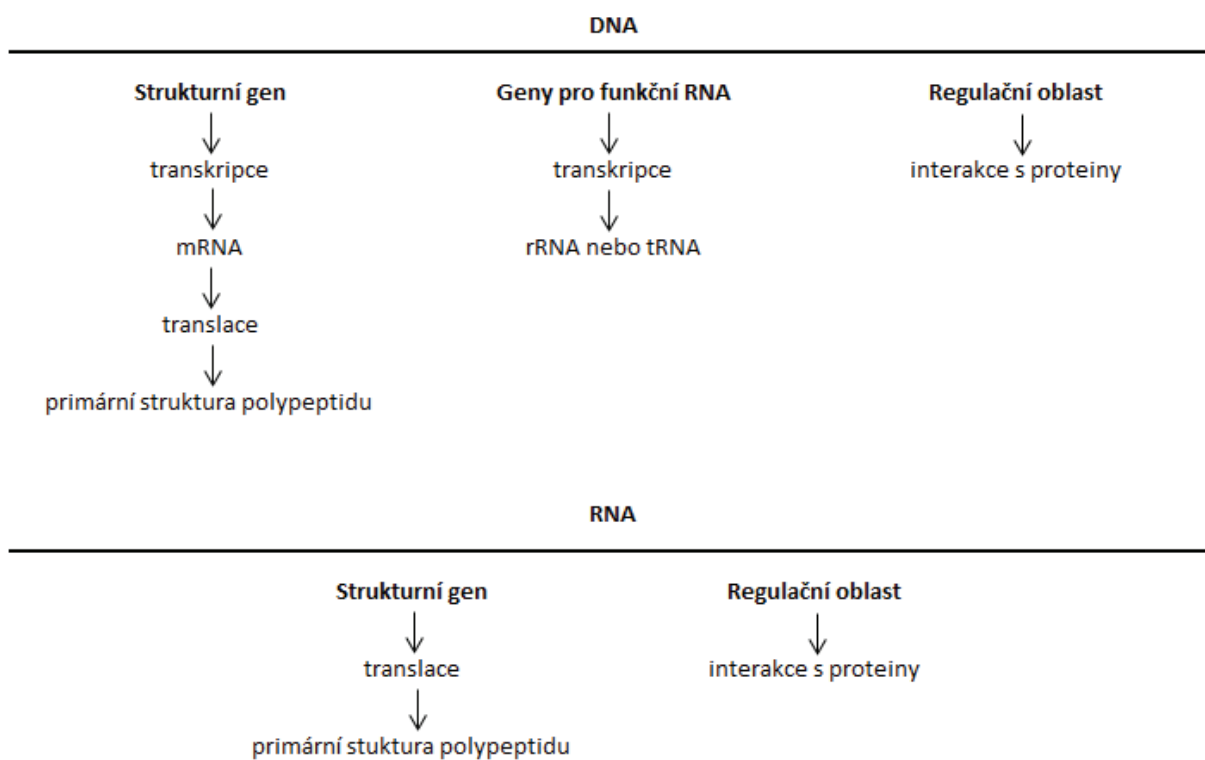
2.2 GENETICKÁ INFORMACE a GEN

Ve fyzice se informace obecně chápe jako míra pořádku. V tomto smyslu lze použít tohoto pojmu pro jakoukoli strukturu nebo systém. Jako míra pořádku vyjadřuje informace instrukce potřebné k vytvoření určité organizace, jakou jsou např. biologické makromolekuly, DNA, RNA a proteiny v buňce. [5]

Genetická informace se zapisuje do univerzálního genetického kódu, který je chápán jako informace, která je obsažena v primární struktuře nukleové kyseliny. V nukleových

kyselinách je kódována pomocí čtyř nukleotidů a to v našem případě v DNA adeninem (A), thyminem (T), guaninem (G) a cytosinem (C). Jako genetickou ji uvádíme proto, že se dědí z rodičů na potomstvo u eukaryotických a prokaryotických organismů prostřednictvím DNA. [3], [5]

Pojem gen se poprvé použil v souvislosti pro označení jednotky dědičnosti W. Johanssenem v roce 1909. Nějakou dobu trvalo, než byl přesný pojem pro gen definován. V současné době jej lze definovat jako úsek polynukleotidového řetězce, který obsahuje genetickou informaci pro strukturu polypeptidu jako translačního produktu (gen strukturní) nebo informaci pro strukturu RNA, která nepodléhá translaci (geny pro rRNA a tRNA). [3] Tato definice platí pro geny DNA virů, prokaryot a eukaryot.

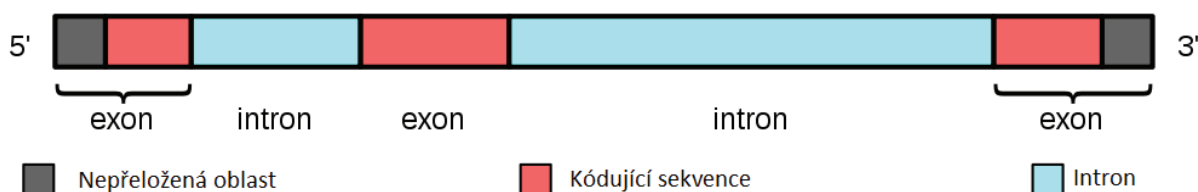


Obr. 5: Schéma funkčních jednotek genetické informace [3]

Gen, který nese informaci o primární struktuře jednoho polypeptidového řetězce, nazýváme jako strukturní. Jinak řečeno kóduje primární strukturu polypeptidu (proteinu) jako translačního produktu. Tento polypeptid může tvořit funkční molekulu bílkoviny s enzymovou, strukturní nebo signální funkcí. U DNA se strukturní gen přepisuje do primárního transkriptu, promítne se do mediátorové RNA, která se následně na ribozomu kóduje do primární struktury proteinu (polypeptidu). [5] Existují dva druhy strukturních genů. Složené strukturní geny a jednoduché strukturní geny.

2.2.1 Strukturní geny

Pro složené strukturní geny je charakteristická jejich skladba. Označujeme je jinak také geny přerušené introny, z čeho vyplývá, že se skládá z intronů a exonů. Navíc jeho primární transkript podléhá posttranskripční úpravě sestřihem. Z celého složeného genu se kompletně všechny introny i exony přepíše do jedné molekuly primárního transkriptu, odkud se následně vyštěpí přepisy intronů a zbylé přepisy exonů se spojí dohromady. Spojením všech naštěpených částí se vytvoří mRNA, která se překládá na ribozomu. Tato úprava, přepis z primárního transkriptu a štěpení přepisů intronů a exonů, včetně jejich následného spojení se odborně označuje jako posttranskripční úprava sestřihem nebo zkráceně prostě sestřih. Ve zjednodušeném podání jsou introny brány jako nekódující úsek, zatímco exony jsou důležitou součástí pro vytvoření mRNA. [5] Složené strukturní geny jsou přímo charakteristické pro eukaryotické organismy.



Obr. 6: Schéma složeného strukturního genu [10]

Jednoduchý gen je oproti složenému opravdu jednoduchý. Označuje se jako gen nepřerušovaný introny, což svým způsobem naprosto vystihuje jeho složení. Jednoduchý gen není složený ze sekvencí, které by měly charakter intronů nebo exonů. Kompletně celý se přepíše do primárního transkriptu, který nepodléhá posttranskripční úpravě sestřihem. Je charakteristický pro prokaryotické organismy.

2.2.2 Geny pro RNA

Geny pro funkční RNA jsou důležité pro strukturu molekul RNA, speciálně pro ty, které nepodléhají translaci, rRNA, tRNA a případně další druhy RNA podílející se na regulaci exprese genetické informace nebo mají jiné funkce. Ve většině případů se několik genů pro tRNA a rRNA přepíše do jedné molekuly primárního transkriptu, který se dále štěpí na jednotlivé funkční typy RNA. Výjimku tvoří mediátorová RNA, která je určena přímo k translaci. [3], [5]

2.2.3 Regulační oblasti

Geny jako regulační oblast jsou částí určitého úseku DNA řetězce, který nese informaci pro strukturu takových molekul RNA, které plní regulační funkci, která má za úkol rozeznávat specifickým proteinem signalizujícím zahájení nebo zastavení transkripce

(u DNA) nebo translace (RNA viry). Každá regulační oblast je určována pomocí specifické sekvence, která určuje, že se na ni bude vázat určitý protein.

2.2.4 Překrývající se geny

Každá DNA nese mnoho různých informací pro translační produkty. Může se ale stát, že určitý úsek může obsahovat informace pro více produktů. V takovém případě mluvíme o tzv. překrývajících se genech. Jsou to strukturní geny, jejichž počáteční iniciační kodony leží na jiných částech úseku DNA, ale jejich koncový terminační kodon je stejný. Takto můžou vzniknout například dva nebo tři translační produkty. Překrývající se úseky představují různé geny, neboť poskytují dva různé translační produkty. [5]

2.3 GENOM

Pojem genom můžeme definovat jako soubor všem molekul DNA v buňce nebo buněčném organismu. V tom případě zahrnuje DNA genovou i negenovou, případně soubor všech genů strukturních a genů pro RNA.

2.3.1 Chromosomy

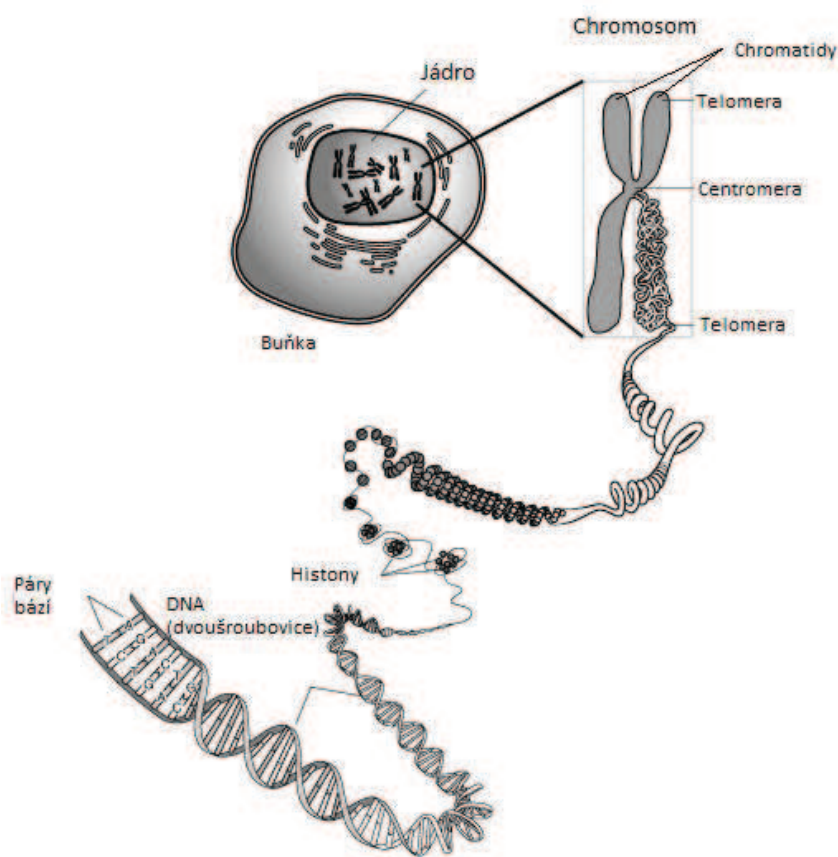
Největší počet genů obsahují chromosomy. Genom eukaryotických organismů se skládá z více jaderných chromosomů a chromosomů mimojaderných. Jaderné chromosomy se označují též jako jaderný genom, mimojaderné jako plazmon. Struktura prokaryotických a eukaryotických chromosomů se hodně liší. V této práci se budeme věnovat pouze chromosomům eukaryotických organismů. [3]

Eukaryotické chromosomy můžeme vždy nalézt v jádře eukaryotických buněk. Jejich vzhled, pokud jej pozorujeme světelným mikroskopem, se v různých fázích dělení buňky mění. Přesto jejich chemická a základní molekulární organizace zůstává po celou dobu naprosto stejná. Eukaryotické chromosomy obsahují DNA a bílkoviny bazického a kyselého charakteru. U těch, kde probíhá transkripce, můžeme přechodně nalézt mRNA, rRNA a tRNA, které jsou v tu chvíli právě syntetizovány. [3]

Každý chromosom obsahuje jednu dlouhou lineární molekulu DNA. Její délka ale dosahuje až několik desítek centimetrů, např. u lidských chromosomů 1,7 – 8,5 cm, a obsahuje řádově 10^8 párů bází, u člověka 50×10^6 – 250×10^6 . [3], [6]

Každá molekula tvořící eukaryotický chromosom musí obsahovat tři typy specializovaných nukleotidových sekvencí, které zajišťují replikaci chromosomů a jejich rozdělení do dceřiných buněk: replikační počátek, centromeru a dvě telomery. Replikační počátek zajišťuje zahájení replikace chromozomální DNA, u velkých chromosomů se nachází až ve stovkách kopií. Centromera je místo, kde se vytváří kinetochor, který zajišťuje během mitózy a meiózy segregaci chromosomů do dceřiných buněk, navíc je velice bohatá na páry AT. Telomery jsou umístěné na koncích chromosomu. Telomery řídí

při replikaci dokončení syntézy dceřiných DNA řetězců na 3' koncích matricových DNA řetězců chromosomu. Jejich hlavní funkce spočívá v kompenzování ztráty koncových nukleotidů při replikaci chromosomu. [3], [6]



Obr. 7: Schéma spiralizace chromosomu [10]

Chromosomová DNA vzhledem k délce svých řetězců a malému prostoru, který je chromosomům v jádře poskytnut, musí mít složitou a několikastupňovou substrukturu, DNA je strukturálně kondenzována. Základní jednotkou substruktury je nukleosom, diskovité tělísko (průměr zhruba 11 nm) tvořené oktamerním komplexem histonů, každý z nich je zastoupen dvakrát. Histony jsou bílkoviny bazického typu, bílkoviny s hojným zastoupením bazických aminokyselin lyzinu a argininu. Díky jejich pozitivnímu náboji se relativně pevně vážou na DNA. Kolem diskovitěho tělíska s histony se ovíjí část molekuly DNA o délce 146 párů bází. DNA řetězec pokračuje dál na další nukleosonová jádra (histonový oktamer), ale až po dalších 60 párech bází, které nazýváme spojovací DNA. V součtu tedy vychází na jeden nukleosom zhruba 200 párů bází DNA. [3], [6]

Dalším stupněm ve struktuře uspořádání je kondenzace „řetízků“ nukleosomů do tzv. 30 nm chromatinových vláken. Velice důležitou roli hraje histon H1, který se váže na histonové jádro jednoho nukleosomu a jednak na začátek spojovací DNA druhého nukleosomu. [3], [6]

30 nm vlákna zaujímají v prostoru ještě vyšší stupeň uspořádání. Vytvářejí jakési smyčky o 20 000 až 80 000 párech bází. Typický lidský chromosom je tvořen 2 600 smyčkami. [3]

V průběhu anafáze mitózy se chromosomy zkracují a kondenzují. Kondenzací rozumíme vlastně těsné nahloučení smyček navzájem. Díky tomu je můžeme dobře ve světelném mikroskopu pozorovat jako tělíska (mitotické chromosomy). V metafázi je struktura chromosomu charakteristická, dvě dceřiné chromatidy jsou spojeny zvláštní strukturou – centromerou, která tvoří hranici mezi rameny chromosomu. Díky různorodému umístění centromery lze chromosomy dělit na akrocentrické, centromera je blíže jednomu z konců chromosomu, submetacentrické, centromera umístěna dále od konce a metacentrické, kdy je centromera zhruba uprostřed chromosomu a vytváří tak přibližně dvě stejná ramena. [3]

2.3.2 Lidský genom

Člověk jako biologický druh se ve svých základních principech neodlišuje příliš velkou měrou od ostatních savců. Platí pro něj všechny poznatky molekulární a obecné genetiky, které byly pozorovány i na jiných organismech. Přesto u něj pozorujeme řadu znaků, morfologických i fyziologických, které jej dělí od vyšších savců a primátů. Mezi tyto znaky patří především rozvoj druhé signální soustavy se schopností abstraktního myšlení, s neobyčejnou kognitivní schopností a vším, co se od těchto vlastní odvíjí. [3], [4]

Lidská genetika je odkázaná na metody pozorování, protože vzhledem k etickým kodexům není možný experimentální přístup. Jedině v ojedinělých nechtěných případech lze pozorovat např. účinky ionizujícího záření na genom člověka a jeho expresi při pokusném či válečném použití atomových zbraní, případně při haváriích atomových zařízení.

U člověka je intenzivně studován patologický výskyt znaků, u řady chorob jsou známy jejich patogeneze od úrovně molekulární až po klinické příznaky, včetně genové determinace a jejího ovlivnění okolím. Tento a mnoho dalších důvodů vedlo k myšlence dešifrovat kompletní lidskou genetickou informaci.

Byl vytvořen projekt, pro určení kompletní sekvence DNA lidského genomu. Zahájili jej v roce 1990 a nese anglickou zkratku HUGO, Human Genom Organization. Cílem bylo úplné zmapování standardního lidského jaderného genomu na úrovni chromosomových map, lokalizaci genových lokusů, i určení nukleotidových sekvencí DNA. [3]

Projektu se účastnilo více jak 200 světových špičkových pracovišť, kde největší koncentrace výzkumné kapacity byla v Národních ústavech zdraví v USA. Během několika let se společnost Wellcome Trust (Velká Británie) stala hlavním partnerem projektu, další pocházeli z Japonska, Francie, Německa, Číny a jiných zemí. Projekt byl

naplánovaný na 15 let, přesto díky velkým pokrokům v technologii byl ukončen o dva roky dříve. [7]

Hlavní náplň projektu tvořilo identifikování celého lidského genomu, zhruba 20 000 – 25 000 genů, stanovit sekvence tří miliard párů bází, které tvoří lidskou DNA, uložit tyto data pro analýzu. Aby se získaná data mohla dobře analyzovat, bylo potřeba vylepšit technologii pro datové analýzy tak velkých souborů dat a následně použít nové technologie v soukromém sektoru. Jako jeden z posledních cílů bylo třeba vyřešit etické, právní a sociální otázky (ELSI), které mohly vzejít z programu HUGO. [7]

I když je program již několik let dokončen, analýza dat potrvá ještě mnoho let. Důležitým krokem bylo rozhodnutí vlády přenést technologie do soukromého sektoru. Licencováním technologií v soukromých společnostech a odměňováním grantů pro inovativní výzkum, se program urychlil. [7]

Přestože je mapování lidské DNA dokončeno, projekt se zabývá sekvenováním DNA i jiných organismů – myši (*Mus musculus*), ovocné mušky (*Drosophila melanogaster*), červa oblého (*Ceanorhabditis elegans*), kvasinky (*Saccharomyces cerevisiae*) a řady bakterií (např. *Escherichia coli* a *Bacillus subtilis*). Analýza genomů těchto experimentálně významných organismů umožní vyhnout se ověřování funkcí identifikovaných genů na lidech a poskytne srovnání struktury strukturální genů, které jsou evolučně méně ovlivněné než sekvence ostatní. [3]

Hlavním přínosem z poznání kompletní sekvence DNA lidského genomu je v nalezení specifického rozmístění strukturních a regulačních genů, které sekvence mutací jsou charakteristické pro určité patologické stavy, dále při porovnávání s jinými genomy různých organismů a vyhledání specifických „lidských“ oblastí. Přesto rozluštěním lidského genomu nezískáme odpověď na to, co je to standardní lidský genom. [3]

2.3.3 Šimpanzí genom

Šimpanzí genom byl dokončen roku 2005, dva roky poté, co byl rozluštěn genom lidský. Všichni doufali, že rozluštění šimpanzího genomu konečně poskytne přesné důkazy o podobnosti DNA jako dokladu toho, že lidoopi a člověk měli společného předka. Často je tato podobnost uváděna jako důkaz evolučního původu člověka.

Do roku 2005 byly pouze dohady, že lidská DNA je až z 99 % podobná šimpanzí, ale tato skutečnost je velice zavádějící. Přesto šimpanzí genom nebyl sekvenován na stejné úrovni jako genom lidský a když byl dokončen, tak se skládal zpočátku z velkého počtu malých neorientovaných a náhodných útržků. Pro sestavení celkového obrazu DNA do souvislého dílu, které by představovaly dlouhé úseky chromozomů, byli vědci nuceni použít lidský genom jako vodítko či rám k zakotvení a orientaci šimpanzích sekvencí.

Takže k poskládání jinak náhodného šimpanzího genomu byla využita evoluční domněnka o údajném přechodu opic v člověka. [8]

V současné chvíli nebylo ještě provedeno zcela vyvážené srovnání celého genomu mezi šimpanzem a člověkem. Přesto byly provedeny studie, které snižují číslo podobnosti genomů z 99% na pouhých 86%. Ale i tak existují velké bloky sekvenčních anomálií mezi šimpanzem a člověkem, které jsou přímo nesrovnatelné. [8]

Kvůli omezení v technologii sekvenování DNA, nemají ještě vědci v současnosti kompletní genomické sekvence pro lidi ani šimpanze. V sekvenci, kterou zatím mají, je třeba provést mnohem rozsáhlejší analýzu.

2.4 CpG ostrůvky

CpG ostrůvky jsou krátký úsek DNA, kde je zastoupení CG bází mnohem větší než v jiných částech DNA. CpG se značí proto, že spojení cytosinu a guaninu je tvořeno pomocí fosfodiesterové vazby, která se značí písmenem „p“ mezi „C“ a „G“. Tyto unikátní sekvence jsou přibližně 1kb dlouhé a zasahují do promotorové oblasti všech lidských genů z 60 – 70 %. [14]

CpG ostrůvky jsou často umístěny okolo promotorů genů, které jsou nezbytné pro hlavní funkci buňky nebo v jiných genech, které v buňce podléhají expresi. V těchto částech CG sekvence nejsou methylované. Oproti tomu, CG sekvence v inaktivních genech jsou obvykle methylované aby zadržely svou expresi. Při nahodilé deaminaci se methylovaný cytosin může převést na thymin. Na rozdíl od cytosino – uracilové mutace, která je účinně opravena, cytosino – thyminová mutace může být opravená pouze mechanismem, který v poškozeném řetězci DNA obnoví původní pořadí nukleotidů, což je v tomto případě velice neúčinné. Z toho důvodu se průběhem času methylované CG sekvence převádí na TG sekvence, což vysvětluje deficit CG sekvence v inaktivních genech. [14]

2.4.1 Původ a evoluce CpG ostrůvků

CpG ostrůvky nejsou ve skutečnosti odolné methylaci, jak se původně jevilo. Vyplývá to z tumorových buněk nebo X-inaktivovaných genů. Vystává proto otázka, jak řídit methylaci tak, aby nevznikaly mutace. [13]

Nabízí se domněnka, že navazování transkripčních faktorů (TF) na CpG je může učinit hůře přístupnější pro DNA – methyltransferáze, která je podporována hned několika řadami důkazů. Jako první, výsledkem de novo methylace CpG je mutace na Sp1 (jeden z transkripčních faktorů) vyžadovaného pro vyjádření myšího genu *Aprt*. Rovněž

demethylace CpG je propojena s prvním a druhým exonem na *p16* genu u lidských rakovinových buněk močového měchýře. [13]

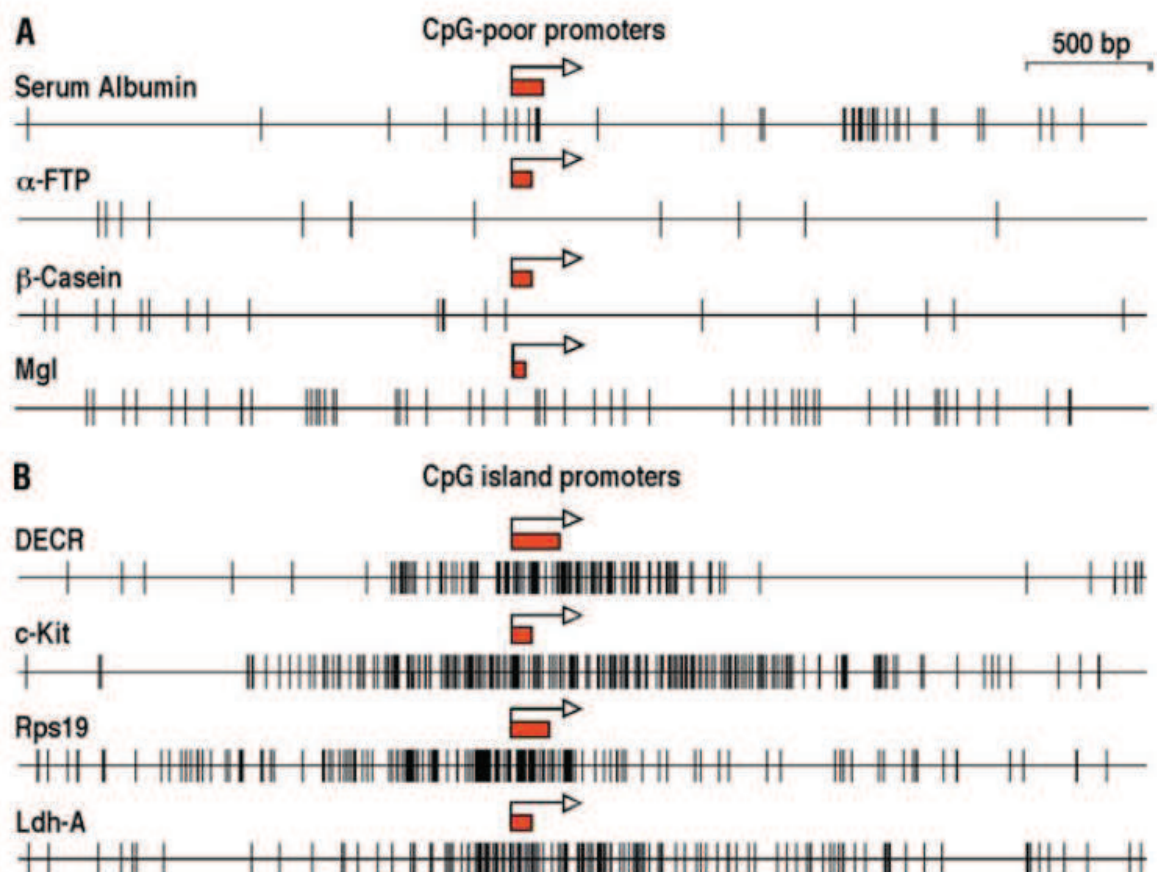
Ačkoliv transkripční faktory mohou přispívat k odolnosti CpG ostrůvku proti de novo methylaci, nemůžeme hledat jejich původ v savčím genomu. Tato myšlenka vede k otázce, proč CpG ostrůvky vlastně existují. Není to pouze řečnická otázka, protože více jak 40 % všech lidských promotorů jsou CpG poor, což znamená, že genový přepis může fungovat i bez přínosu CpG ostrůvků. [13]

2.4.2 CpG ostrůvky a CpG poor

Navzdory rozmanitosti promotorů, geny transkribovány RNA polymerázou II mohou být tříděny do dvou různých a vzájemně se vylučujících skupin. V první skupině, frekvence CpG je stejná jako v ostatních částech genomu, kde je přibližně jeden na každých 100 nukleotidů. Tady můžeme naleznout geny, jejichž exprese je omezena na limitovaný počet buněčných typů. Protikladem této skupiny je druhá, kde na 5' konci genů je seskupení zhruba 1kb dlouhé, kde je četnost CpG přibližně 10 krát vyšší než v jiných částech genomu. Tato seskupení nazýváme jako CpG ostrůvky a takový zřetelný shluk CG lze velice snadno vizuálně poznat i z jednoduchého CpG schéma. [11]

Důsledným spojením CpG ostrůvků s promotorovými oblastmi mnoha genů, přímo nabádají k jejich eventuálnímu použití coby markeru k lokalizování genů v genetické sekvenci. Dokonce i dvacet let po jejich objevení jsou CpG ostrůvky stále spolehlivý znak pro promotorovou predikci v savčím genomu. [11], [12]

Co je tak speciálního na CpG ve srovnání s dalšími 15 dinukleotidy v DNA? CpG ostrůvky jsou oblasti, kde se odehrává methylace a z více jak 80 % těchto oblastí je methylováno na pozici 5' u lidí a myší. Poněkud paradoxní je, že CpG ostrůvky zůstávají nemethylované (v převážné většině případů), zatímco u většiny ostatních CpG, osamocených a rozmístěných po celé délce genomu, methylace proběhne. Dodejme k methylaci lidské i myší, CpG ostrůvky mají obsah C + G přibližně 67 % a 64 %, zatímco průměrně v genomu jich nalezneme kolem 41 % a 42 %. [11]

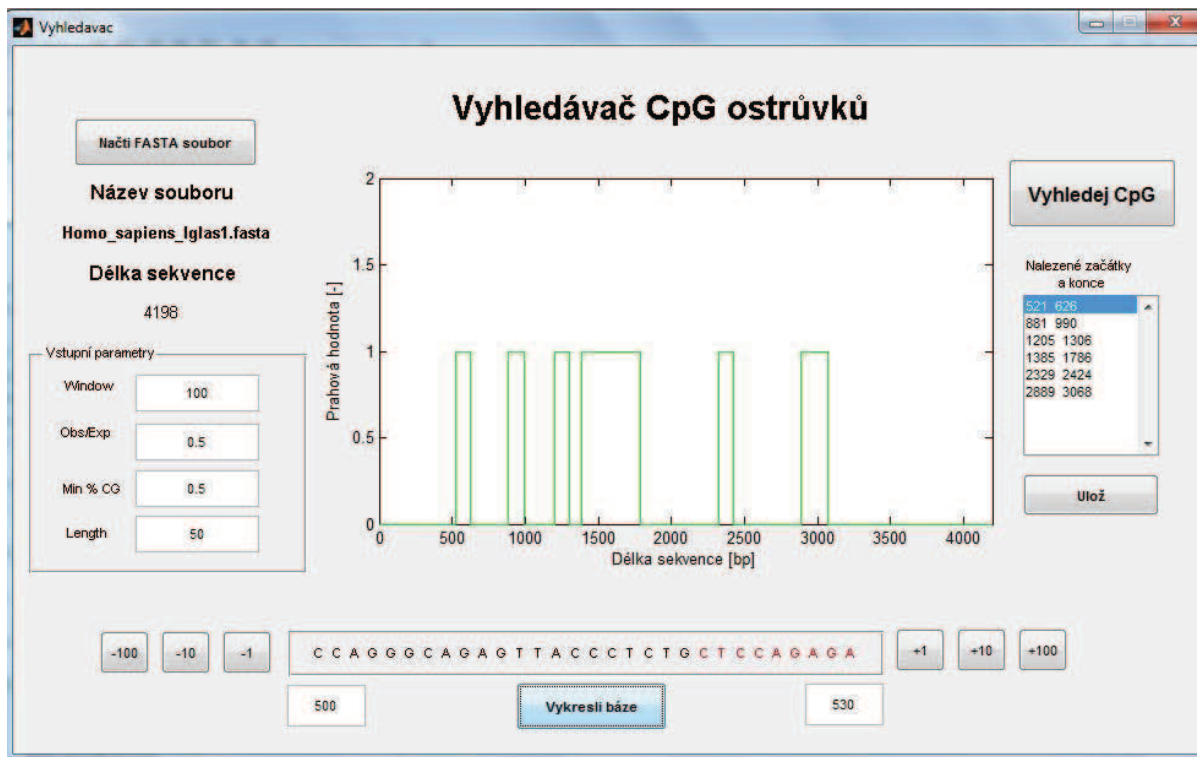


Obr. 8: Dvě skupiny promotorů u člověka a myši. Červené obdélníky a šipky symbolizují první exon a transkripční zahajovací část [11]

3 PRAKTICKÁ ČÁST

3.1 Program VYHLEDAVAC.m

Program Vyhledavac.m byl vytvořen v programovém prostředí MATLAB 7.9.0.529 R2009b. Uživatelské prostředí bylo navrženo v grafickém editoru GUIDE. Program slouží ke genové analýze CpG ostrůvků. Je konstruován pro zpracovávání genetických souborů ve formátu FASTA. Hlavní funkcí programu je ve vybrané sekvenci vyhledat části bohaté na dimery CG a GC a podle uživatelem zadaných podmínek je následně detekovat jako CpG ostrůvek. Program je rozdělen do tří jednoduchých přehledných částí. Načtení sekvence, vyhledání a uložení dat. Pro účely dalšího zpracování dat, které uživatel získá, program vypíše délku sekvence, začátky a konce CpG ostrůvků, vykreslí graf nalezených úseků a přesně zvýrazní báze nacházející se v CpG ostrůvku. Nalezené úseky, začátky a konce lze uložit ve formátu pro použití v tabulkovém editoru MS Excel, s názvem jaký si uživatel sám zvolí. Poslední grafická část programu slouží k interpretaci sekvence v grafické podobě, kdy se po celé její délce lze posunovat a podle barevného rozlišení analyzovat báze označené jako CpG ostrůvek. Po sekvenci se lze posunovat jednoduše posuvnými tlačítky, případně pomocí přesně stanovených rozmezí zadaných uživatelem. Podrobný popis programu a jeho uživatelského rozhraní jsou přesněji popsány v následujících kapitolách. Manuál k programu je zařazen v příloze.



Obr. 9: Uživatelské prostředí programu Vyhledavac.m

3.1.1 Načtení sekvence

Aby program mohl pracovat, potřebuje vstupní data. Program Vyhledavac.m zpracovává sekvence ve formátu FASTA.

Načtení sekvence spustíme v programu stisknutím tlačítka `Nacti`, kterým se spustí funkce `Nacti_Callback`. Do ní je vnořena externí funkce s názvem `NactiFasta`. Jejími výstupy jsou název sekvence a sekvence samotná.

Při spuštění funkce vyskočí výběrové okno, s již přednastaveným souborovým formátem *.fasta*, kde si uživatel zvolí sekvenci, kterou chce analyzovat. Po vybrání a nahrání, ve funkci dojde k uložení názvu souboru, cesty k němu a soubor se otevře. Hned na začátku je definována prázdná proměnná připravená pro načtení hlavičky sekvence a pro sekvenci.

V prvním řádku souboru je definována hlavička sekvence, která se načte do připravené proměnné `hlavicky`. V cyklu probíhá čtení souboru po řádcích. Každý řádek se zapíše do vektoru a každý následující se zařadí za něj. Ve výsledku je ve vektoru celá sekvence. Po načtení celé sekvence se čtení ukončí a soubor zavře. Tím je uzavřena část načítání a následuje část vyhledávání CpG ostrůvků.

Po nahrání sekvence se vypíše její název do kolonky `NazevSekvence`, proběhne výpočet délky sekvence a zobrazení její hodnoty v okně `vypisdelka`.

3.1.2 Vyhledávání CpG ostrůvků

Vyhledávání CpG ostrůvků započne po stisknutí tlačítka `CpGsearch`, čímž se zavolá funkce `CpGsearch_Callback`. Do této funkce je implementována další, tvořící jádro celého programu, funkce s názvem `CpGCount.m`. Pro její správné spuštění potřebuje vstupní data, kterými jsou sekvence a parametry, zadané v textbozech uživatelem. Při nevyplnění textboxů vyskočí upozornění, které vybízí k jejich doplnění, jinak program nebude správně fungovat. Parametry jsou zadávány v datovém typu *string*, proto je nutné je ihned převést do formátu *double*. Poté je můžeme všechny načíst do funkce společně se sekvencí, kterou jsme získali nahráním z externí funkce `NactiFasta`.

Funkce `CpGsearch` vyhledává v sekvenci CpG ostrůvky na základě uživatelem zadaných podmínek. Hned na začátku je nutno definovat proměnné, do kterých se budou zapisovat pozice nalezených CpG ostrůvků.

Vstupní parametry	
Window	100
Obs/Exp	0.5
Min % CG	0.5
Length	50

Obr. 10: Předdefinované vstupní parametry

Pro analýzu je potřeba převést sekvenci v písmenném formátu do formátu číselného. Znak pro cytosin a guanin jsou nahrazeny číslem jedna a adenin s thyminem nulou. Během převádění se do proměnných `mapaC` a `mapaG` načítají nalezené báze C a G.

Jakmile je sekvence převedena na čísla, přijde na řadu vyhledávání CG a GC dimerů. V okně, jehož velikost určil uživatel na počátku, se pohybuje jedno malé okénko o defaultní velikosti deset. Pokud je v malém okénku sedm a více jedniček, do proměnné `velka_pozice` vyznačující pozici CpG přidá jedničku.

Pro detekci CpG ostrůvků je třeba spočítat dvě kritéria. Poměr cytosinu a guaninu a výpočet pozorované/očekávané (`Obs/Exp`). Princip výpočtu poměru je jednoduchý. Pomocí nachystaných proměnných `mapaC` a `mapaG` jej lze snadno vypočítat. Suma všech nalezených C plus suma všech nalezených G vydělené velikostí okna. V tom stejném cyklu probíhá i výpočet `Obs/Exp` dle uvedeného vzorce:

$$Obs/Exp = \frac{Počet\ CpG}{Počet\ C * Počet\ G} * N, \quad (1)$$

kde *Počet CpG* jsou nalezené dimery CG a GC ve vektoru `velka_pozice`, *Počet C* počet nalezených C, *Počet G* počet nalezených G a *N* je celkový počet bází v sekvenci. Algoritmus je převzat od Gardiner-Garden. [12]

Kalkulace probíhají v okně, které se po sekvenci posunuje po jednom prvku. Ke konci sekvence dochází k situaci, kdy je okno větší, než délka zbývajících sekvence, proto je zařazena podmínka, kdy se postupně zmenšuje okno pro výpočet. Kvůli narůstající koncové chybě je poslednímu prvku vektoru přiřazena hodnota blížící se k nule.

Závěrečnou a nejdůležitější částí funkce je stanovení velikosti a pozic začátků a konců CpG ostrůvků. Pro definování začátku ostrůvku je předpoklad splnění podmínky,

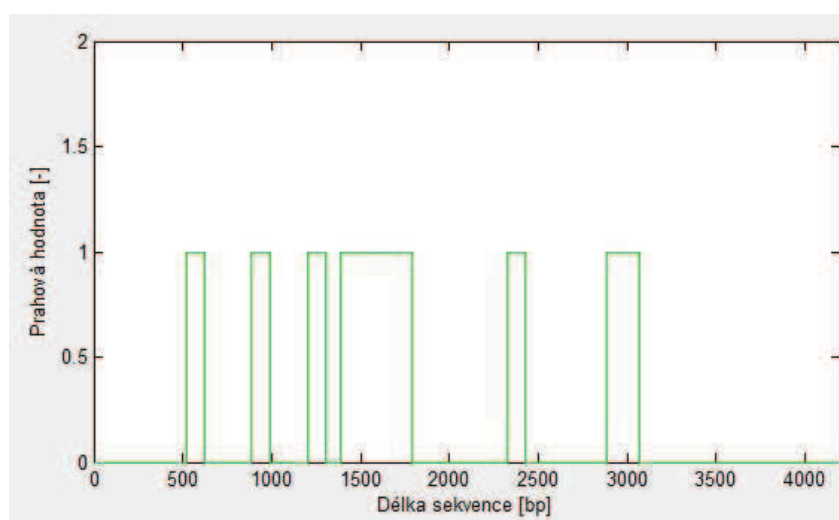
kde velikost poměru CG a Obs/Exp jsou zároveň větší nebo rovny parametrům, které určil uživatel. Jakmile je jedna z podmínek nesplněna, zaznamená se konec ostrůvku. Další z parametrů, který je pro stanovení podstatný, je minimální vzdálenost ostrůvků od sebe. Pokud máme zaznamenané dva ostrůvky a jejich vzájemná vzdálenost je menší jak třicet párů bází, sloučí se dohromady v jeden. Další z podmínek, která je definovaná uživatelem, je minimální velikost ostrůvku. V případě, že je detekovaný ostrůvek menší než zadaná velikost, smaže se. Vykreslují se pouze ty ostrůvky, které splňují veškeré zadané podmínky.

Program při stanovování ostrůvku v první iteraci nemusí kontrolovat, zda se před ním nachází souřadnice již existujícího ostrůvku, přesto v dalších krocích již toto stanovení musí existovat. Pokud v okně nejsou podmínky splněny, program zkontroluje, zda v pozici před ním nebyl detekován ostrůvek. Jestliže ano, zaznačí se konec ostrůvku. Hned vzápětí je provedeno vyhodnocení délky ostrůvku a jeho porovnání s minimální délkou požadovanou uživatelem. Jestli je hodnota detekovaného ostrůvku menší, jsou souřadnice vynulovány a detekce pokračuje.

Výstupem z funkce jsou tři vektory, vektory `start` se zaznamenanými začátky úseků ostrůvků, `stop` obsahující hodnoty konců ostrůvků a `cpgisland` se souřadnicemi všech ostrůvků, které splnily podmínky vložené uživatelem.

3.1.3 Grafické zobrazení a uložení

Vykreslení rozložení CpG ostrůvku na sekvenci probíhá v hlavní funkci `Vyhledavac.m`. Rozložení CpG ostrůvků se promítá v grafu a v listboxu po pravé straně grafu se vypíší nalezené začátky a konce ostrůvků. Ty lze pomocí tlačítka `Ulož` uložit do souboru, ve formátu pro zpracování v tabulkovém editoru MS Excel, pro další použití.



Obr. 11: Příklad vykreslení CpG ostrůvků

Dalším uživatelským nástrojem je vykreslení nukleotidů a přebarvení bází na základě stanovení CpG ostrůvků. Když se na dané části nachází ostrůvek, nukleotid se přebarví na červenou. K vypsání bází je vyhrazen panel s rozsahem zobrazení 30 nukleotidů. Po celé sekvenci se lze posunovat buď přepisováním pozic, nebo stisknutím tlačítek po stranách panelu. Posun je možný o jeden, deset a sto nukleotidů na obě strany. Pokud sekvence dosáhne svého minima nebo maxima a uživatel i přesto bude chtít posouvat po sekvenci, není to možné a sekvence zůstává zobrazena v pozici prvních třiceti nebo posledních třiceti nukleotidů.

Princip vykreslení je jednoduchý. Při splnění vstupních podmínek, zda jsou zadaná správná čísla (nenulová, s rozsahem 30 nebo menší apod.) vstupují proměnné do cyklu, kde se podle uživatelem zvolených pozic, na nich provádí analýza, jestli se v nich nachází CpG ostrůvek. Pokud ano, na pozici se barva písmena změní na červenou a zařadí se za předchozí do proměnné `znak{i}`. Ta se po skončení cyklu vypíše do panelu jako přebarvené nebo nepřebarvené nukleotidy.



Obr. 12: Ukázka vykreslení nukleotidů

3.2 Analýza CpG ostrůvků

Pro zjištění správné funkce programu je třeba jej srovnat s volně přístupným internetovým vyhledávačem. Geny pro analýzu jsou vybrány z databáze GenBank. Aby byla analýza kompletní, tak programem byly porovnávány geny Homo sapiens, Pan troglodytes, Bos taurus a Mus musculus. Vybrány byly: gen kódující protein tyrosin fosfatáza, non-receptor typ 1 (PTPN1), gen leukemie myelocytů (PML) a gen Středozevní horečky (MEFV). Geny od Homo sapiens pochází z verze Build 37.3 a chromosomu 15, 16 a 20, Pan troglodytes z verze Build 3.1 a chromosomu 15, 16 a 20, Mus musculus z verze Build 37.2 a chromosomu 2, 9 a 16 a Bos taurus z verze Build 6.1 a chromosomu 13, 21 a 25. Záměrně byly zvoleny geny s poměrně velkým počtem bází, aby se zjistilo, zda program rychle zvládne i dlouhou sekvenci.

3.2.1 Analýza programem Vyhledavac.m a internetovým vyhledávačem <http://www.ebi.ac.uk/Tools/emboss/cpgplot/>

Pro analýzu CpG ostrůvku byly aplikovány parametry, které korespondovaly s možnostmi, které má internetový vyhledávač. U obou jsou zvoleny stejné, aby bylo možné výsledky porovnat a zhodnotit. Internetový vyhledávač

<http://www.ebi.ac.uk/Tools/emboss/cpgplot/> (dále zmiňován jako internetový vyhledávač) byl zvolen pro svoje přehledné uživatelské prostředí a pro jednoduchý výstup v podobě seznamu nalezených ostrůvků, jejich délky a grafem jejich umístění na sekvenci.

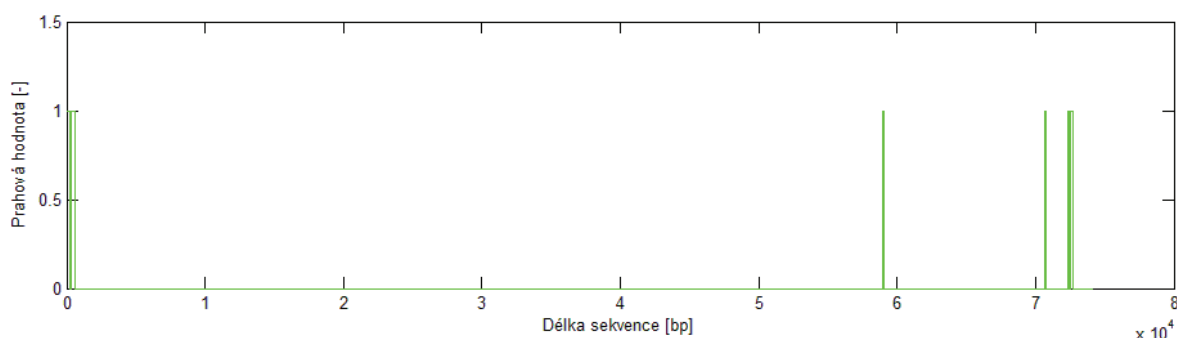
Aby byly znatelné rozdíly mezi vyhledávači, jsou grafy obou vyhledávačů řazeny za sebou. Výstup z internetového vyhledávače je oříznut pouze na nalezené CpG (ve výstupu jsou ve dvou dalších grafech zobrazeny hodnoty Obs/Exp a %CG).

Tabulka 1: Parametry stanovené pro analýzu

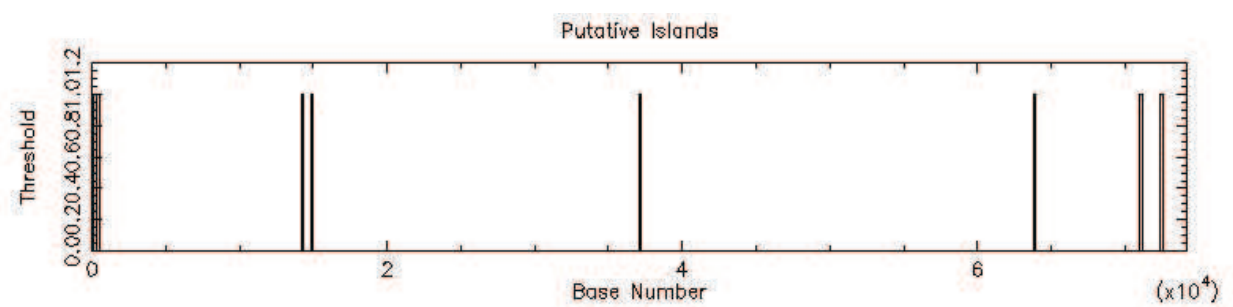
Parametr	Hodnota
Window	100
Obs/Exp	0.6
Min % CG	0.5
Length	100

3.2.2 PTPN1

Protein kódovaný tímto genem je jeden z počátečních členů protein-tyrosinofosfátázového kmene (PTP), který byl izolován a identifikován na základě své enzymatické aktivity a aminokyselinové sekvence. PTP katalyzuje hydrolýzu fosfátových monoesterů speciálně tyrosinových zbytků. PTP jsou známy pro své signální molekuly, které usměrňují množství buněčných procesů zahrnující buněčný růst, diferenciaci, mitotický cyklus a nádorotvorné přeměny. Tento PTP je pozorován jako negativní regulátor insulinové signalizace defosforylujícím fosfotyrosinovým zbytkem receptoru insulinové kinázy, stejně jako JAK2 a TYK2 kinázy, které mají svou roli v této PTP při kontrole buněčného růstu a odpovědnosti buňky za interferonovou stimulaci. [15]



Obr. 13: PTPN1 u Homo sapiens v programu Vyhledavac.m

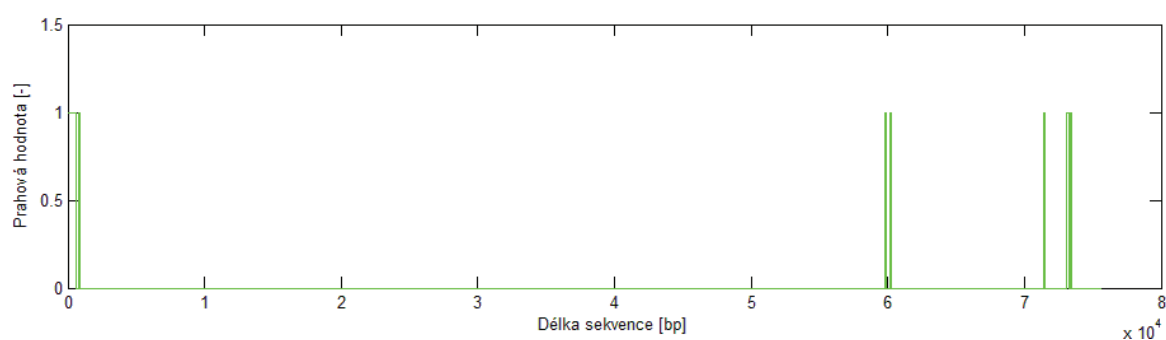


Obr. 14: PTPN1 u Homo sapiens v internetovém vyhledávači

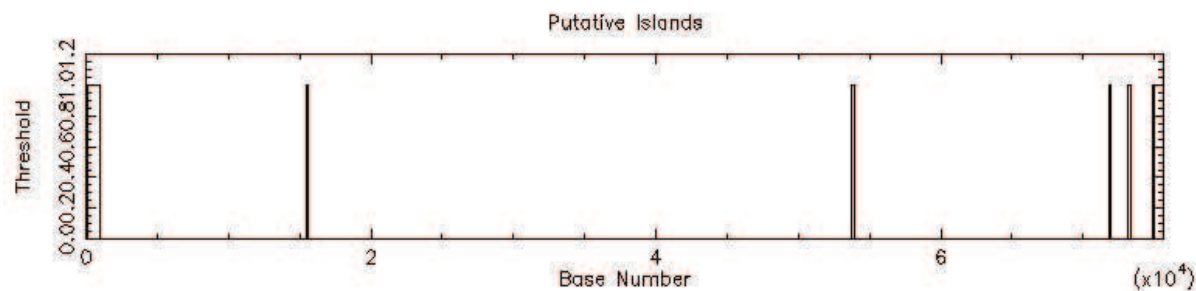
Tabulka 2: Porovnání vyhledávačů u genu PTPN1 v Homo sapiens

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
48	296	1	246
347	564	325	586
14255	14362	58943	59082
14878	14998	72313	72454
37103	37216	72565	72694
63851	63953		
71024	71178		
72370	72530		

Z tabulky 2 je patrné, že se vyhledávače scházejí v prvních dvou vyhledaných ostrůvcích (± 20 bází), v ostatních nalezených se liší až na poslední, předposlední u vytvořeného programu, kde je také částečná shoda.



Obr. 15: PTPN1 u Pan troglodytes v programu Vyhledavac.m

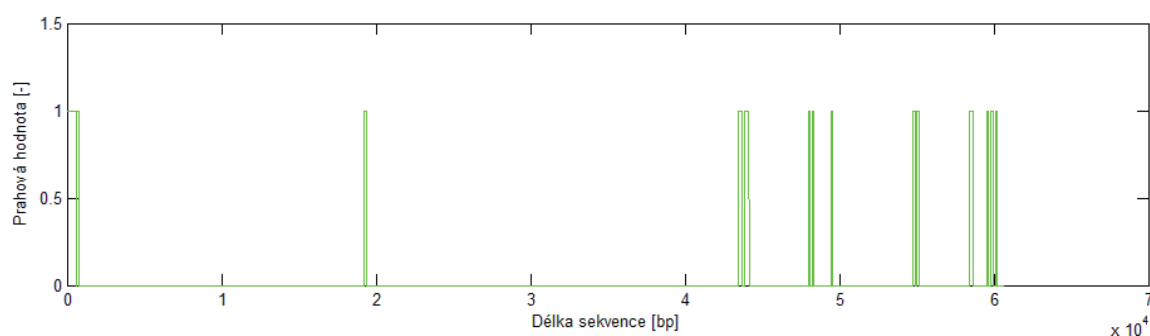


Obr. 16: PTPN1 u *Pan troglodytes* v internetovém vyhledávači

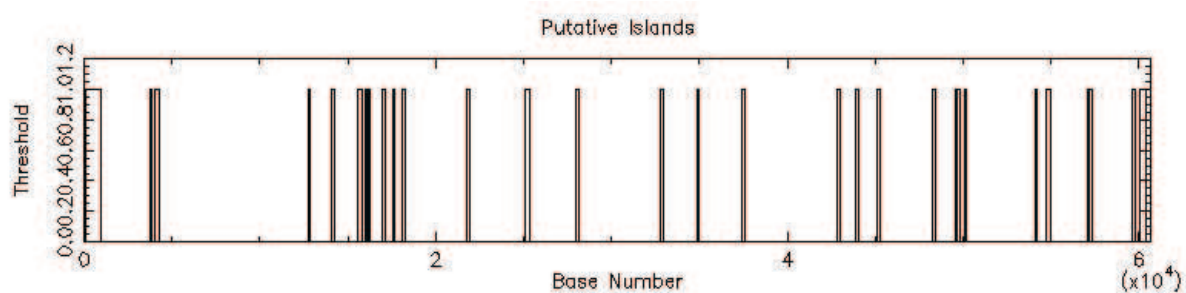
Tabulka 3: Porovnání vyhledávačů u genu PTPN1 v *Pan troglodytes*

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
84	924	1	606
15468	15569	781	886
53702	53981	59746	59850
71789	71943	60175	60282
73137	73291	71405	71523
74832	74935	73081	73226
		73325	73462

V případě sekvenční genu u *Pan troglodytes* se shoda nachází v prvních částech a u předposledního ostrůvku. Ve středních částech buď ostrůvky nalezeny nejsou, a pokud ano, odlišují se o velký počet bází (až 1000 bp).



Obr. 17: PTPN1 u *Bos taurus* v programu Vyhledavac.m

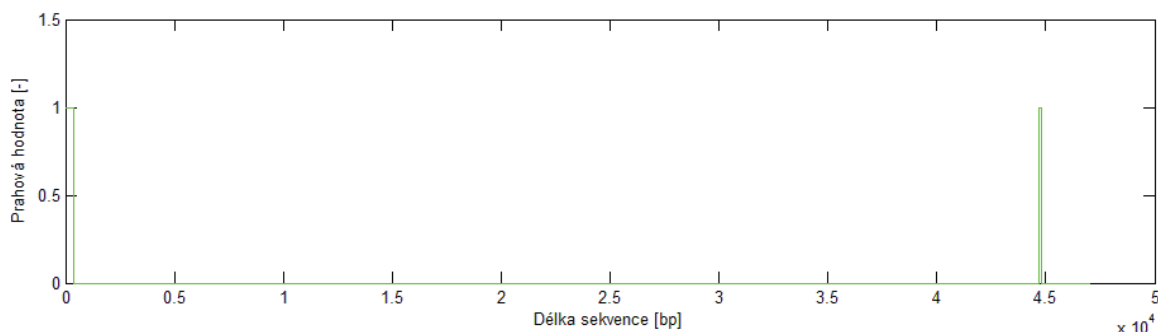


Obr. 18: PTPN1 u Bos taurus v internetovém vyhledávači

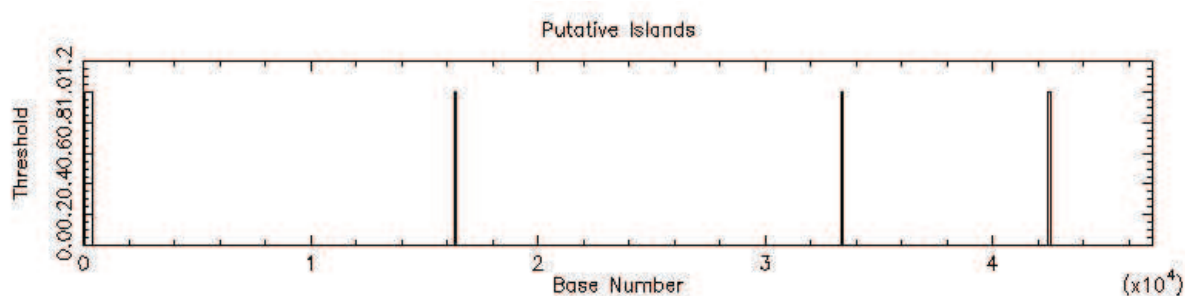
Tabulka 4: Porovnání vyhledávačů u genu PTPN1 v Bos taurus

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
47	920	1	602
3764	3867	641	778
4055	4290	19221	19350
12745	12868	43477	43666
14090	14267	43881	44146
15580	15778	47973	48082
15972	16122	48217	48346
16158	16265	49441	49578
16991	17129	54789	54926
17567	17674	55025	55182
18104	18264	58449	58650
21790	21936	59545	59646
25074	25328	59833	59646
27963	28170	60085	60202
32781	32948		
34854	34959		
37441	37542		
42820	42974		
43911	44011		
45074	45238		
48242	48401		
49539	49657		
49684	49827		
50050	50198		
54078	54186		
54755	54994		
57047	57156		
57176	57339		
59652	59816		
60092	60420		

Zde lze rozeznat rozlišovací schopnost obou programů. Internetový vyhledávač našel mnoho částí bohatých na CG části, které označil jako CpG ostrůvky. Vlastní program tolik ostrůvků nenalezl, ale v koncové části se vyhledávače shodují s rozdílem pouze několika desítek bází.



Obr. 19: PTPN1 u *Mus musculus* v programu Vyhledavac.m



Obr. 20: PTPN1 u *Mus musculus* v internetovém vyhledávači

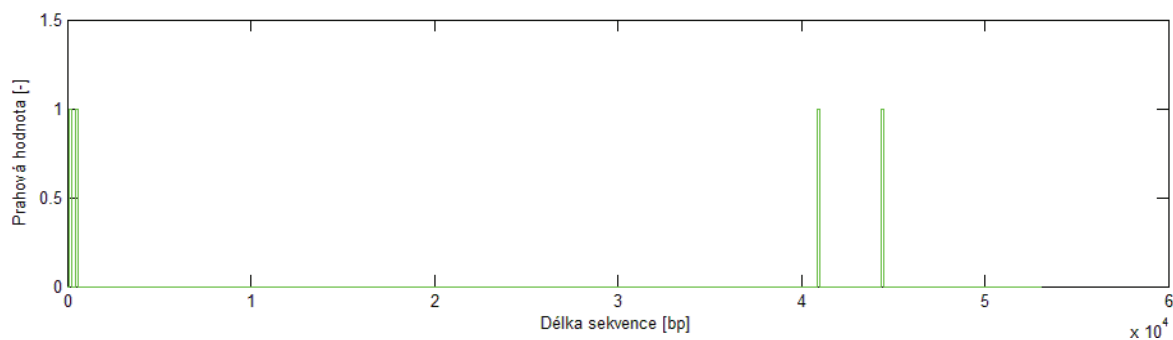
Tabulka 5: Porovnání vyhledávačů u genu PTPN1 v *Mus musculus*

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
47	419	1	388
16311	16414	44686	44817
33329	33429		
42445	42568		

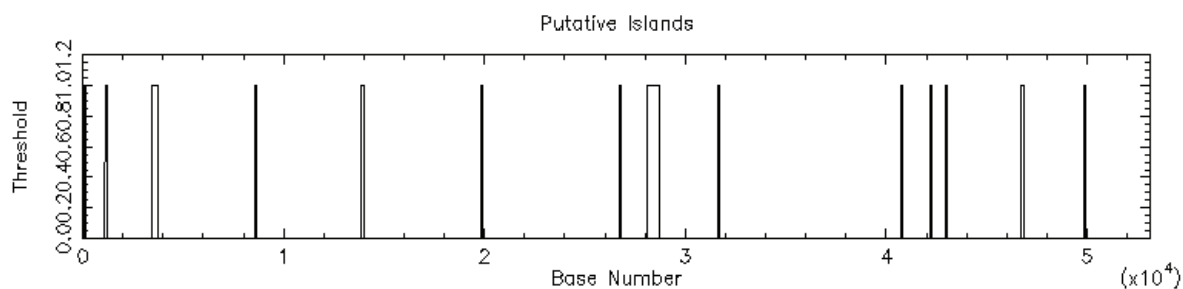
Gen u *Mus musculus* není příliš bohatý na CpG ostrůvky. Vytvořený program našel pouze dva, oproti internetovému, který našel čtyři. Nejsou příliš rozdílné v detekci prvního ostrůvku, ale zbytek se již mezi sebou neshodují.

3.2.3 PML

Protein kódovaný tímto genem je členem TRIM kmene. Tento fosfoprotein se lokalizuje na jádra, jako funkce transkripčního faktoru a tumorového supresoru. Projevuje se v souvislosti s buněčným cyklem a reguluje odezvu p53 na onkogenické signály. [17]



Obr. 21: PML u Homo sapiens v programu Vyhledavac.m

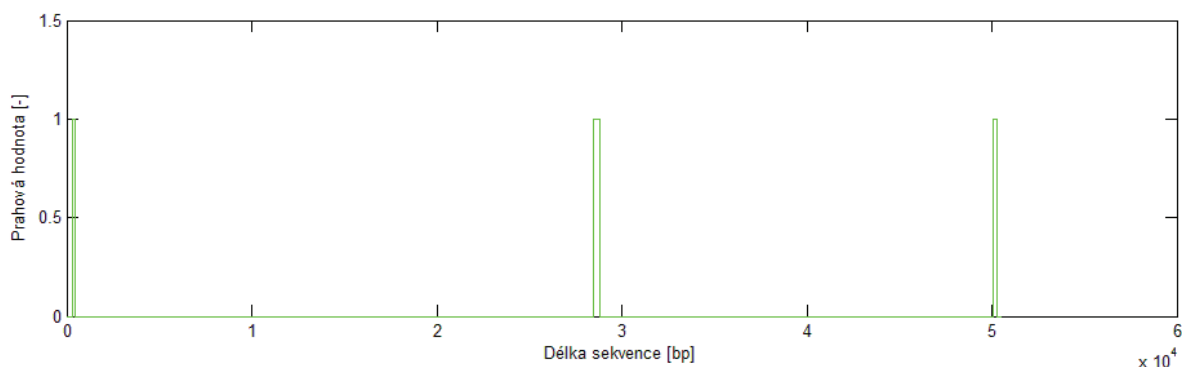


Obr. 22: PML u Homo sapiens v internetovém vyhledávači

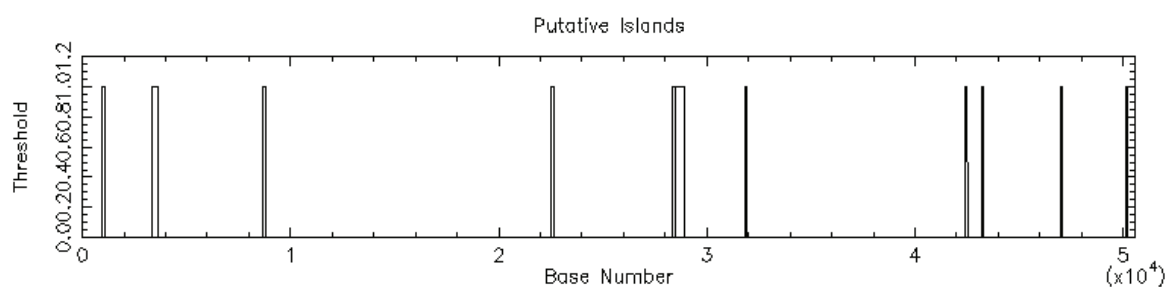
Tabulka 6: Porovnání vyhledávačů u genu PML v Homo sapiens

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
56	164	77	242
1106	1221	441	566
3477	3773	40889	40994
8564	8669	44389	44490
13872	13985		
19808	19934		
26719	26818		
28122	28725		
31586	31702		
40700	40832		
42175	42279		
42936	43043		
46721	46822		
49854	49953		

U genu PML je rozdíl mezi vyhledávacími markantní. Vyhledavac.m našel pouze čtyři ostrůvky, oproti velkému množství, které vyhledal internetový vyhledávač. Velikost nalezených ostrůvků není větší než 100 bp.



Obr. 23: PML u *Pan troglodytes* v programu Vyhledavac.m

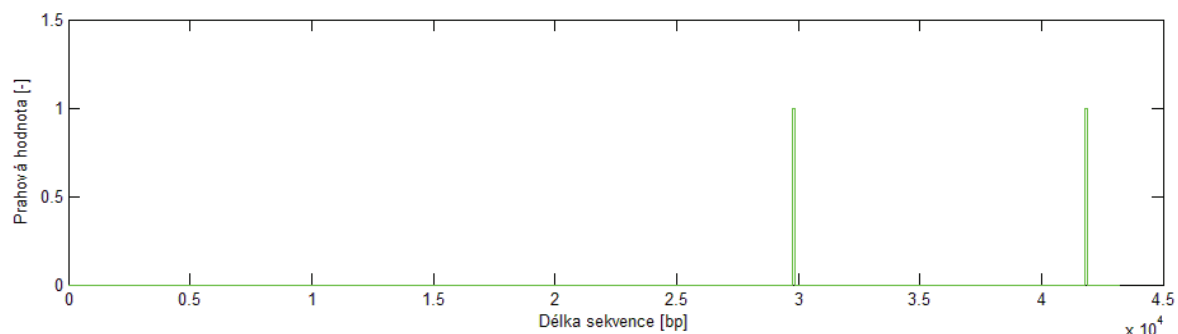


Obr. 24: PML u *Pan troglodytes* v internetovém vyhledávači

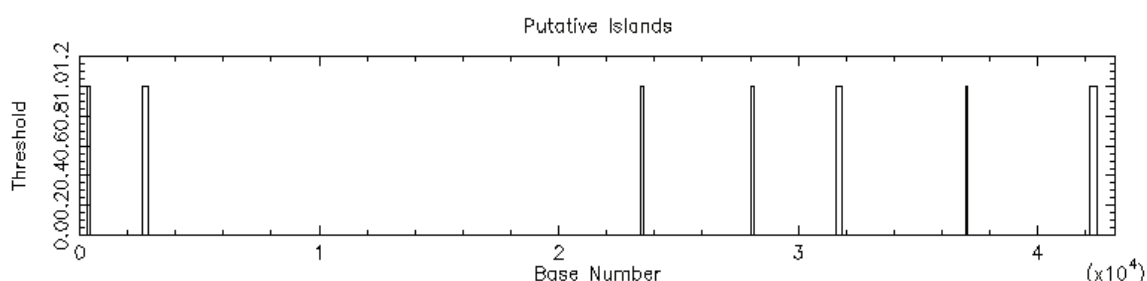
Tabulka 7: Porovnání vyhledávačů u genu PML v *Pan troglodytes*

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
963	1078	301	422
3344	3625	28465	28834
8694	8793	50092	50310
22516	22654		
28333	28461		
28467	28942		
31808	31914		
42395	42518		
43186	43303		
46966	47067		
50099	50198		

U šimpanze vytvořený program vyhledal pouze tři ostrůvky, kde poslední z nich koresponduje s posledním ostrůvkem (jeho počáteční pozice), nalezeným vyhledávačem z internetu a ostrůvek na pozici 28467 (u internetového) má shodný počátek s vytvořeným, start na pozici 28465.



Obr. 25: PML u Bos taurus v programu Vyhledavac.m

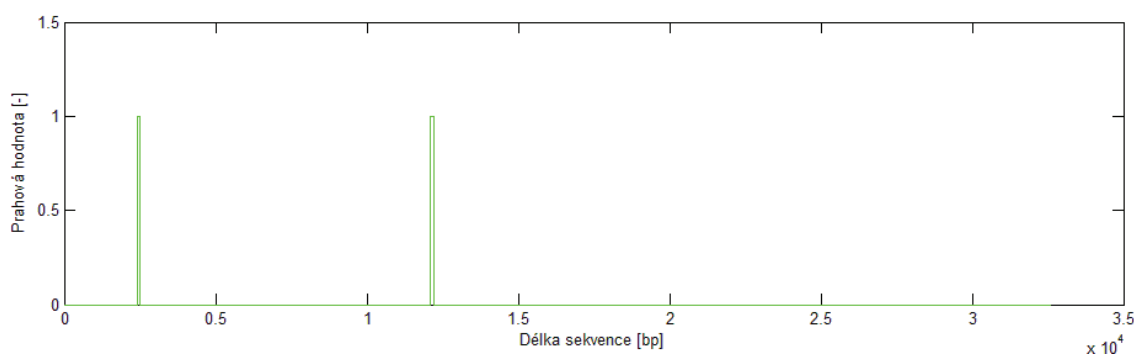


Obr. 26: PML u Bos taurus v internetovém vyhledávači

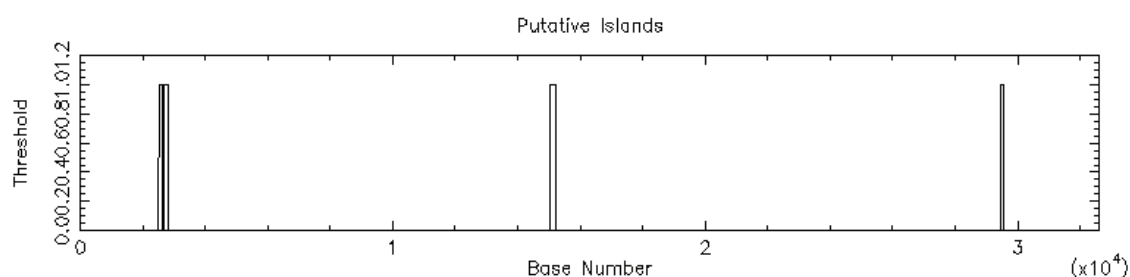
Tabulka 8: Porovnání vyhledávačů u genu PML v Bos taurus

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
330	452	29781	29886
2603	2882	41784	41898
23390	23540		
28007	28171		
31575	31833		
36976	37076		
42167	42470		

Vyhledávače se v tomto případě velice liší. Nejen počtem nalezených ostrůvků, také jejich nalezenými pozicemi. Převážná část ostrůvků není větší jak 100 bp.



Obr. 27: PML u *Mus musculus* v programu Vyhledavac.m



Obr. 28: PML u *Mus musculus* v internetovém vyhledávači

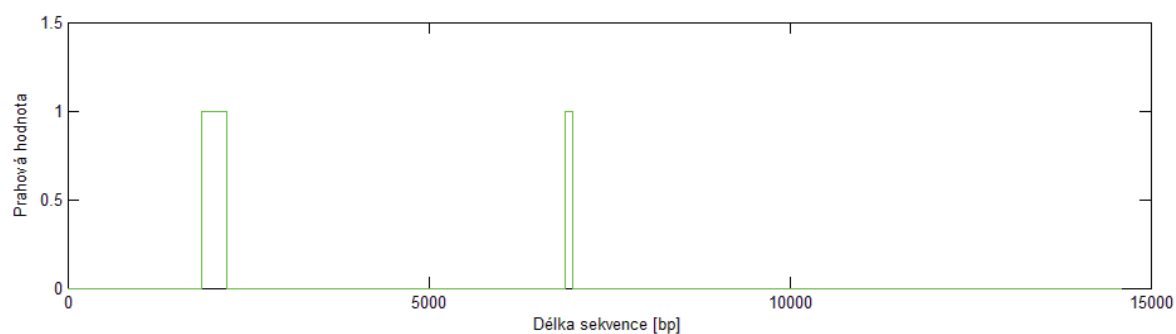
Tabulka 9: Porovnání vyhledávačů u genu PML v *Mus musculus*

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
2512	2648	2413	2518
2692	2796	12101	12210
15034	15211		
29457	29560		

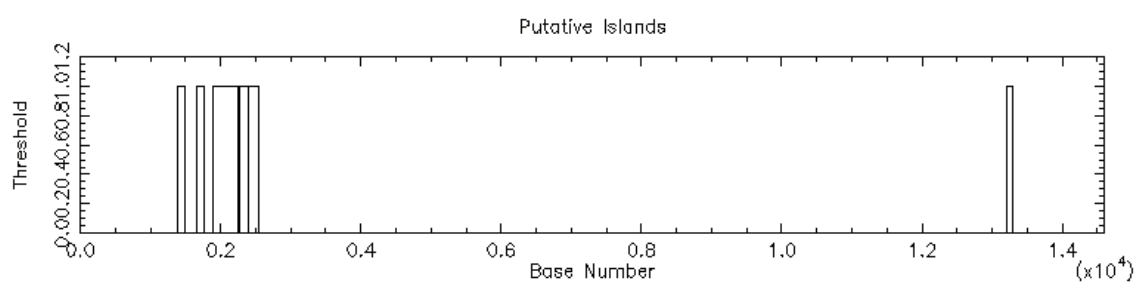
Nalezené úseky v genu u myši se liší řádově o sto bází v prvním nalezeném úseku. Počet vyhledaných ostrůvků není tak značný jako v předchozích případech, přesto druhý ostrůvek nalezený programem má naprosto jinou pozici než některé z detekovaných u internetového vyhledávače.

3.2.4 MEFV

Tento gen kóduje protein známý též jako pyrin nebo marenostrin, důležitý modulátor vrozené obranyschopnosti. Mutace tohoto genu jsou spojeny se Středozevní horečkou (dědičný opakující se horečkový syndrom). [16]



Obr. 29: MEFV u Homo sapiens v programu Vyhledavac.m

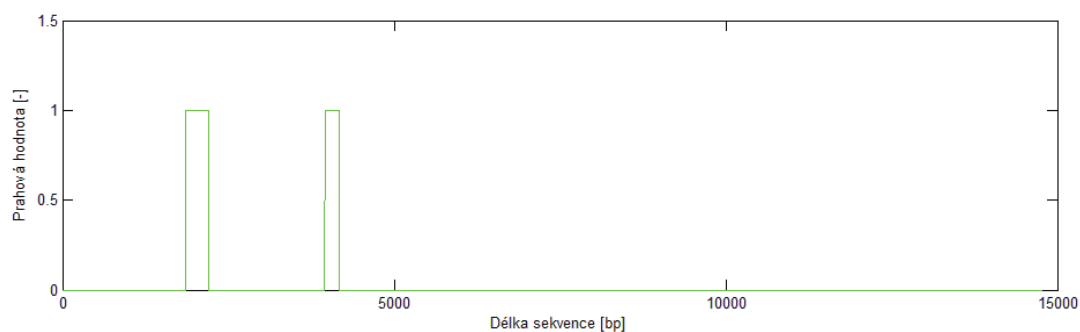


Obr. 30: MEFV u Homo sapiens v internetovém vyhledávači

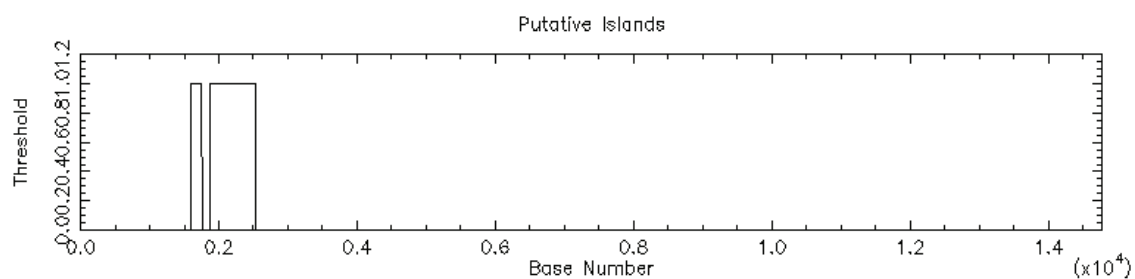
Tabulka 10: Porovnání vyhledávačů u genu MEFV v Homo sapiens

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
1390	1490	1861	2202
1655	1759	6889	6990
1892	2258		
2263	2387		
2404	2542		
13201	13300		

Gen MEFV je mnohem kratší, než předchozí dva geny, aby se ověřila funkčnost programu i při malých velikostech sekvencí. Program našel dva ostrůvky, z nichž první odpovídá jednomu z detekovaných u internetového vyhledávače s malým rozdílem několika desítek bází.



Obr. 31: MEFV u Pan troglodytes v programu Vyhledavac.m

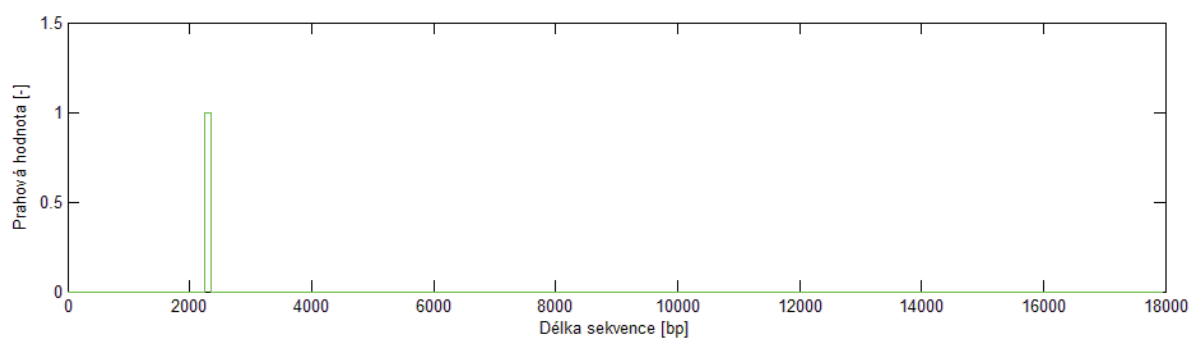


Obr. 32: MEFV u Pan troglodytes v internetovém vyhledávači

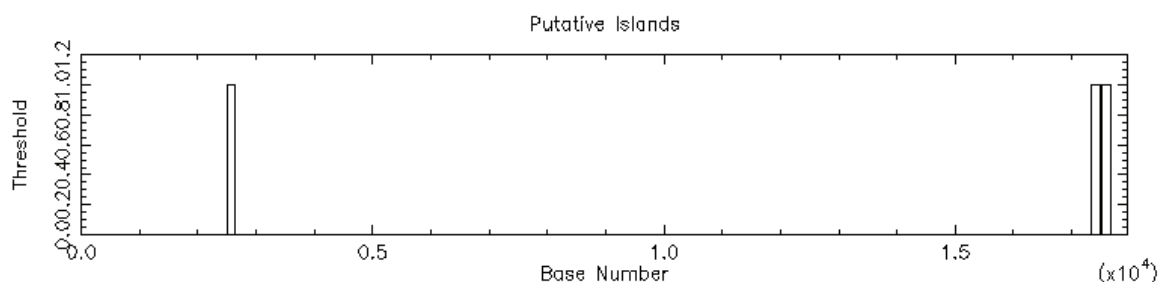
Tabulka 11: Porovnání vyhledávačů u genu MEFV v Pan troglodytes

Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
1600	1756	1856	2196
1869	2535	3955	4172

U šimpanze byly nalezeny pouze dva ostrůvky v obou použitých vyhledávačích. Jejich pozice se vzájemně lišily, internetový našel ostrůvky mezi 1600 až 2535, oproti vlastnímu programu, který detekoval ostrůvky ve větším rozmezí, pozice posledního je 4172.



Obr. 33: MEFV u Bos taurus v programu Vyhledavac.m



Obr. 34: MEFV u Bos taurus v internetovém vyhledávači

Tabulka 12: Porovnání vyhledávačů u genu MEFV v Bos taurus

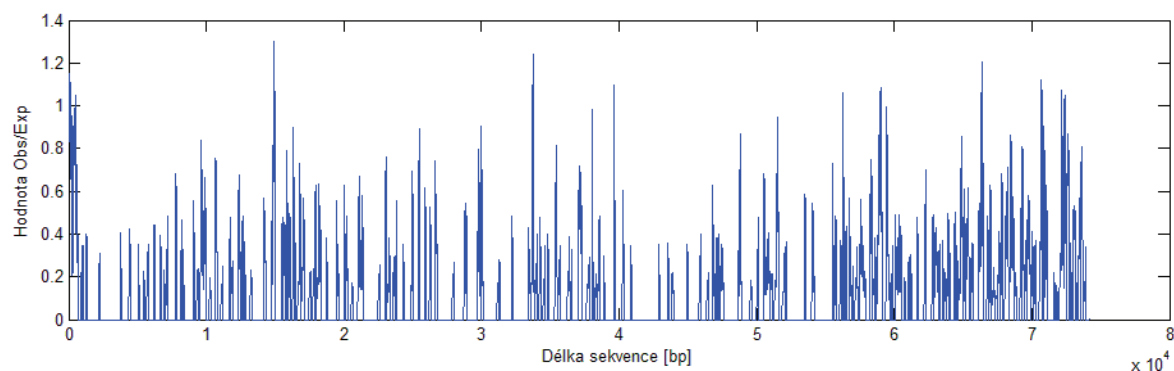
Internetový vyhledávač		Vyhledavac.m	
Začátky	Konce	Začátky	Konce
2517	2626	2253	2354
17335	17494		
17522	17679		

Zde vytvořený program našel jediný ostrůvek, zatímco internetový vyhledal tři. Ani jedna pozice vlastního vyhledávače nekoresponduje s detekovanými v internetovém vyhledávači.

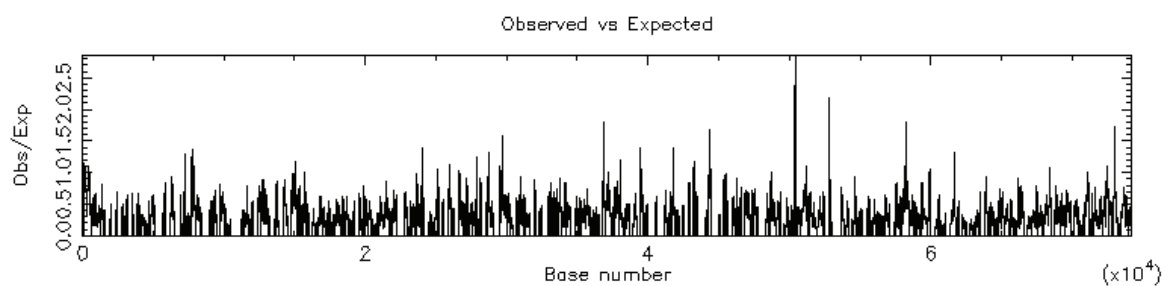
Byl použit i gen MEFV Mus musculus, kde ale žádný ostrůvek detekován nebyl a to jak v programu Vyhledavac.m ani v internetové aplikaci.

3.3 Shrnutí výsledků analýzy

Již při zběžném pohledu na dva výstupy dvou různých programů je zřejmé, že se od sebe určitým způsobem liší. Ve většině případů našel internetový vyhledávač více CpG ostrůvků, než vytvořený program. Často se shodují v počátečních a koncových oblastech, zatímco středové oblasti jsou různé. Přestože oba programy pracují s výpočtem Obs/Exp založeným na stejném principu (viz rovnice 1.), hlavní rozdíl je v počtu detekovaných ostrůvků, které jsou pro výpočet důležité. Na obrázku 35 a 36 je vidět, jak velký rozdíl je v hodnotách spočítaných Obs/Exp.



Obr. 35: Obs/Exp v programu Vyhledavac.m



Obr. 36: Obs/Exp v internetovém vyhledávači

Je patrné, že hodnoty Obs/Exp je skoro dvojnásobně větší než u vytvořeného vyhledávače. Z toho důvodu jsou výsledky uprostřed sekvenec rozdílné.

4 ZÁVĚR

Cílem bakalářské práce bylo vytvořit program pro vyhledávání CpG ostrůvků v genomu eukaryotických organismů. Poté pro zjištění funkčnosti program porovnat s volně přístupným vyhledávačem CpG ostrůvků.

Práce je rozdělena do dvou částí. Teoretické a praktické. Teoretická část se věnuje problematice DNA, jejímu složení, struktuře, co jsou to báze, vysvětlení pojmu párování bází, co znamená pojem genetická informace a gen a jaké typy genů existují. Podrobně je rozebrán genom, jakou roli hrají chromosomy a část je zaměřena na lidský a šimpanzí genom. V závěru teoretické práce je zařazena důležitá kapitola o CpG ostrůvcích, kde se v genech nalézají, původ ostrůvků a rozdíl mezi CpG ostrůvkem a CpG poor.

V praktické části bakalářské práce je uveden popis vytvořeného programu `Vyhledavac.m` zpracovaného v programovém prostředí MATLAB. Program umí načíst sekvenci ve formátu FASTA, zpracovat ji a nalézt v ní CpG ostrůvky. Ty vyhledává na základě uživatelem vložených podmínek a to velikost okna, které se po sekvenci pohybuje, minimální velikost ostrůvku, procentuální zastoupení dimerů CG a GC a hodnotu Obs/Exp. Ty byly zvoleny proto, že bylo potřeba otestovat funkčnost programu s jiným podobným, založeným na stejné metodě, výpočtu Obs/Exp. Výsledek programu je zaznamenán v grafu, který ukazuje rozložení CpG ostrůvku v délce celého genu. Další možností grafické interpretace má uživatel pomocí prohlížení jednotlivých bází v sekvenci, kde jsou části patřící do CpG ostrůvku přebarveny na červeně. Uživatel se může pohybovat po sekvenci buď přímým zadáním rozsahu, který chce prohlížet nebo posouváním pomocí tlačítek o jeden, deset nebo sto nukleotidů dopředu nebo dozadu po sekvenci. Nalezené úseky CpG ostrůvků lze uložit do tabulkového souboru.

Poslední část tvoří analýza CpG ostrůvků. Pro patřičné vyzkoušení funkčnosti programu, z databáze GenBank byly vybrány geny s velkým počtem bází a jeden s počtem menším. Nejvýše se počet pohyboval okolo 70000 bází. Geny byly společný pro čtyři organismy: Homo sapiens, Pan troglodytes, Bos taurus a Mus musculus.

Pro analýzu byl použit vytvořený program a internetový vyhledávač. Vstupní parametry byly zvoleny pro oba vyhledávače stejné, aby bylo porovnání objektivní. Jako výstup z internetového vyhledávače bylo použito pouze rozložení CpG ostrůvků v sekvenci, původně obsahuje grafy tří, procentuální zastoupení CG párů, hodnoty Obs/Exp a rozložení CpG na sekvenci. Pouze pro srovnání hodnot výpočtu byly použity grafy Obs/Exp obou programů v rámci zdůvodnění rozdílnosti nalezených CpG ostrůvků.

Z výsledku analyzovaných sekvencí je na první pohled zřejmé, že se vyhledávače od sebe liší. Nejvíce je to patrné např. na sekvenci patřící *Bos taurus* genu PTPN1, kde má internetový vyhledávač nalezených ostrůvků mnohem více, než vytvořený program. Přestože se vyhledávané části lišily, byla jasná podobnost sekvencí člověka a šimpanze. Z evolučního hlediska je patrné, že člověk a šimpanz mají společného mnohem více, jak dokazuje i analýza CpG ostrůvků. Ve vlastním programu je tato podobnost mnohem zřetelnější. Další sekvence se od sebe různily v mnohonásobně větší míře, i když některé úseky společné měly.

5 ZDROJE

- [1] PAČES, Václav. Aktuálně.cz [online]. 15. 04. 2009 [cit. 2011-10-22]. DNA, viry, opice a lidi. Dostupné z WWW: <<http://blog.aktualne.centrum.cz/blogy/vaclav-paces.php?itemid=6354>>.
- [2] DOSTÁL, Jiří, et al. Biochemie pro posluchače bakalářských oborů. 1. vydání. Brno: Masarykova univerzita, 2009. 158 s. ISBN 978-80-210-5020-4.
- [3] NEČAS, Oldřich. Obecná biologie : pro lékařské fakulty. 3. přepracované vydání. Jinočany : H & H, 2000. 554 s. ISBN 80-86022-46-3.
- [4] OTOVÁ, Berta, et al. Lékařská biologie a genetika: 1. díl. 1. vydání. Praha: Karolinum, 2008. 123 s. ISBN 978-80-246-1594-3.
- [5] ROSYPAL, Stanislav. Úvod do molekulární biologie. 4. inovované vyd. Brno :, 2006, 289 s. ISBN 80-902-5625-2.
- [6] ROSYPAL, Prof. RNDr. Stanislav DrSc. Molekulární biologie : Druhý díl. 3. vydání. Brno : Masarykova univerzita, 2002. 305 s. ISBN 80-902562-1-X.
- [7] Human Genome Project Information [online]. 2011 [cit. 2011-11-21]. About HGP. Dostupné z WWW: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml.
- [8] TOMKINS, Jeffrey. Human-Chimp Similarities: Common Ancestry or Flawed Research?. Acts & Facts. červen 2009, 38, s. 12-13.
- [9] Nanotrubičky z DNA. VÍTĚZSLAV KŘÍHA. Aldebaran [online]. [cit. 2011-12-30]. Dostupné z: http://www.aldebaran.cz/bulletin/2009_38_nan.php
- [10] Genetika - biologie. Genetika - biologie [online]. [cit. 2011-12-30]. Dostupné z: <http://www.genetika-biologie.cz/>
- [11] ANTEQUERA, F. Structure, function and evolution of CpG island promoters. Cell. Mol. Life Sci. 2003, vol. 60, p. 1647-1658.
- [12] GARDINER-GARDEN, M., Frommer, M. CpG islands in vertebrate genomes. J. Mol. Biol. 1987, vol. 196., p. 261-282.
- [13] BIRD, P. Adrian. CpG islands – ‚A rough guide‘. FEBS Letters. 2009, vol. 583, p. 1713-1720.
- [14] DNA Methylation and CpG Island. Web - books [online]. [cit. 2011-12-30]. Dostupné z: <http://www.web-books.com/MoBio/Free/Ch7F2.htm>
- [15] PTPN1 protein tyrosine phosphatase. NCBI [online]. [cit. 2012-05-11]. Dostupné z: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=retrieve&dopt=full_report&list_uids=5770

- [16] MEFV Mediterranean fever - NCBI. *NCBI* [online]. [cit. 2012-05-03]. Dostupné z: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=retrieve&dopt=full_report&list_uids=4210
- [17] PML promyelocytic leukemia - NCBI. *NCBI* [online]. [cit. 2012-05-03]. Dostupné z: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=retrieve&dopt=full_report&list_uids=5371

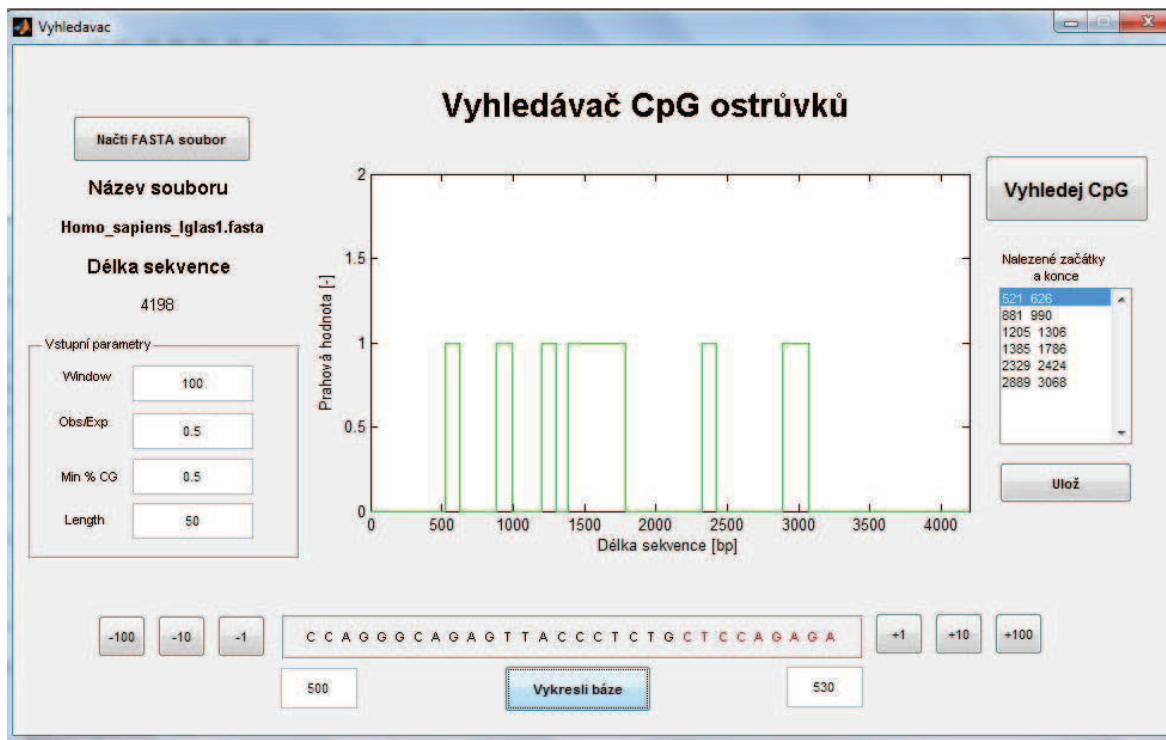
6 PŘÍLOHY

A Zkratky

DNA	deoxyribonukleová kyselina
HUGO	HUman Genome Organization
RNA	ribonukleová kyselina
rRNA	ribozomální RNA
tRNA	transferová RNA
mRNA	mediátorová RNA
bp	páry bází
ELSI	Ethical, Legal, and Social Issues
PIR	Protein Information Resource
kb	kilobáze

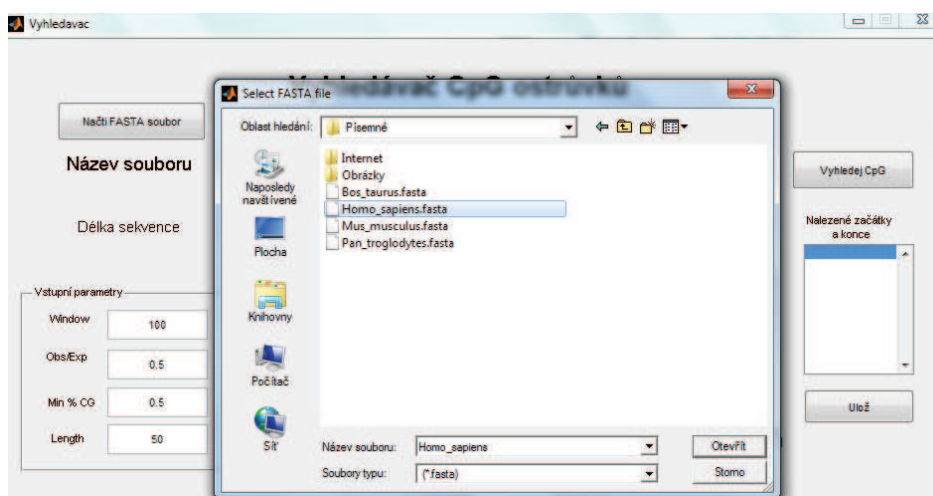
B UŽIVATELSKÝ MANUÁL

Program `Vyhledavac.m` slouží pro vyhledávání CpG ostrůvků v genomech eukaryotických organismů. Používá sekvence ve formátu FASTA a jeho výstupem je grafické vykreslení CpG ostrůvků do grafu a je možnost získaná data uložit nebo prohlížet.



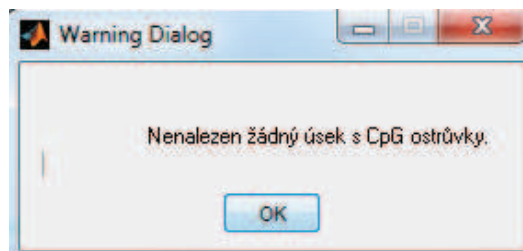
Obr. 37: Uživatelské prostředí programu

- 1) Program `Vyhledavac.m` spustíte v programovém prostředí MATLAB stiskem tlačítka F5 nebo zelenou šipkou v horním příkazovém panelu
- 2) Stiskem tlačítka `Načti FASTA soubor` se vyvolá dialogové okno pro výběr sekvence



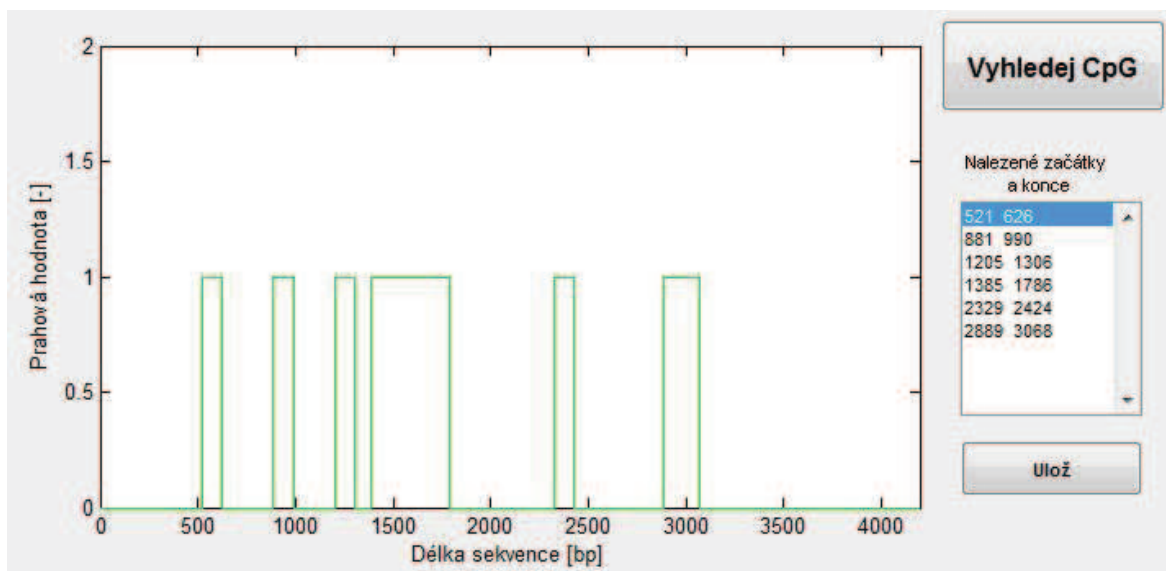
Obr. 38: Správné načtení sekvence

- 3) Zkontrolujte, zda načtení souboru proběhlo správně. Známkou správného nahrání je vypsání názvu sekvence a vypočítání její délky
- 4) Zadejte požadované vstupní parametry pro analýzu. Již dopředu jsou nastaveny výchozí jako doporučené nastavení. Hodnoty zadávané do oken nesmí překročit určitou minimální stanovenou hranici
 - a) Hodnota okna `Window` nesmí být menší než 100
 - b) Hodnota okna `Obs/Exp` nesmí být menší než 0.5
 - c) Hodnota okna `Min % CG` nesmí být menší než 0.5
 - d) Hodnota okna `Length` nesmí být menší než 50
- 5) Pro vyhledání CpG ostrůvků stiskněte tlačítko `Vyhledej CpG`
- 6) Pokud v sekvenci nebyl nalezen žádný ostrůvek, program upozorní vyskakovací hláškou



Obr. 39: Dialogové okno při nenalezení CpG ostrůvku

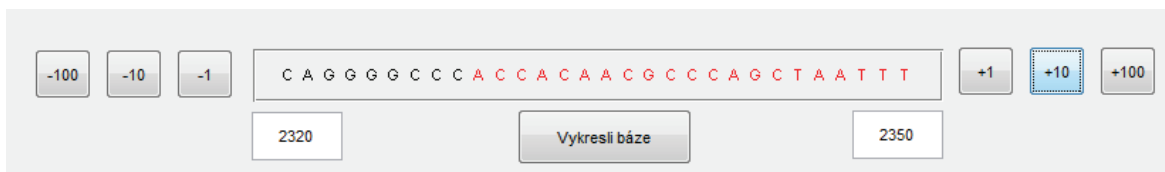
- 7) Pokud jsou ostrůvky nalezeny, zapíší se do listboxu po pravé straně grafu, načtež se souběžně zobrazí jejich rozložení na sekvenci v grafu.



Obr. 40: Vykreslení grafu a zapsání pozic do listboxu

- 8) Jestliže chcete výsledky dále zpracovávat, stiskněte tlačítko `Ulož`. Po stisknutí se objeví vyskakovací okno s přednastaveným názvem souboru vycházející z názvu použité sekvence a možnost uložit soubor ve formátu `*.xls`.

- 9) Po stisknutí tlačítka `Vykresli báze` se zobrazí prvních 30 bází ze zadané sekvence. Pokud se v daném místě nachází CpG ostrůvek budou báze přebarveny na červeně. Pro vybrání jiné části sekvence je možno použít okna na začátku a na konci panelu, kde lze zadat vlastní zobrazený rozsah, ale pouze v rozmezí třiceti nukleotidů. Nelze zadat hodnoty, které přesahují délku sekvence (automatické nastavení hodnot na posledních třicet bází).
- 10) Pro posouvání se po větších rozmezích jsou po stranách panelu umístěny tlačítka pro posun sekvence o jeden, deset nebo sto bází dopředu nebo dozadu.



Obr. 41: Zobrazení nalezeného úseku

- 11) Ukončení programu provedete stiskem Esc nebo kliknutím na křížek v pravém horním rohu

C VÝVOJOVÝ DIAGRAM FUNKCE CPGCOUNT

