

UNIVERZITA KARLOVA V PRAZE

Přírodovědecká fakulta

Katedra parazitologie

Studijní program: Biologie

Studijní obor: Parazitologie



Bc. Ondřej Brzoň

**Analýza regulačních oblastí genů v genomu oxymonády
*Monocercomonoides***

Analysis of gene regulatory regions in the genome of oxymonad
Monocercomonoides

DIPLOMOVÁ PRÁCE

Školitel: Doc. Mgr. Vladimír Hampl, Ph.D.

Praha 2016

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 13. 8. 2016

Ondřej Brzoň

Poděkování

V první řadě bych rád poděkoval svému školiteli, Doc. Vladimíru Hamplovi, za všechny rady a za jeho trpělivost a čas, který věnoval přípravě této práce během mého magisterského studia. Poděkovat bych také chtěl Dr. Anně Karnkowské za pomoc s bioinformatickou částí práce a Mgr. Petru Soukalovi za pomoc při psaní skriptů. Dále pak všem členům laboratoře evoluční protistologie i dalších skupin na katedře parazitologie za vytvoření příjemného a motivujícího prostředí.

Abstrakt

Regulace genové exprese je klíčovou schopností každé buňky v jejím vývoji, diferenciaci a v udržování homeostázy. Prozatím je ale dostupných poměrně málo informací o regulaci genové exprese u prvoků a naše poznatky o tomto procesu u eukaryot pocházejí pouze z několika modelových organismů. Náš výzkum probíhá na oxymonádách, skupině málo prozkoumaných anaerobních protist, vyskytujících se ve střevech některých živočichů. V této práci se zabýváme rodem *Monocercomonoides*.

Genová exprese je ovlivňována mnoha mechanismy na několika úrovních. My jsme se zaměřili na strukturu promotorových oblastí, 5' nepřekládané oblasti a základní transkripční a translační iniciační faktory. Naše výsledky srovnáváme s nejbližšími prozkoumanými příbuznými rodu *Monocercomonoides*, kterými jsou *Trichomonas vaginalis* a *Giardia intestinalis*.

V promotorových oblastech *Monocercomonoides* jsme identifikovali několik konzervovaných motivů, které se potenciálně podílí na regulaci transkripce, včetně TATA boxu a TATA-like motivu. 5' nepřekládané oblasti jsou poměrně krátké (většinou 20 – 30 nukleotidů) a obsah GC párů v těchto oblastech je nižší, než je obvyklé u modelových organismů. U vybraných genů jsme délky těchto oblastí ověřovali pomocí RACE. V genomu *Monocercomonoides* jsme anotovali sady základních transkripčních (23 proteinů) a iniciačních translačních faktorů (30 proteinů). Sady nalezených proteinů se výrazně neliší od sad v genomu *Trichomonas vaginalis*. Pokusili jsme se purifikovat transkripční faktory vázající se na „upstream“ oblast genu pro GAPDH, ale s neprůkaznými výsledky.

Klíčová slova

5' UTR, regulační oblast genu, TATA box, genom, oxymonáda, transkripční faktor, translační faktor

Abstract

Regulation of gene expression is a key ability of every single cell in its development, differentiation and homeostasis. On the other hand, rather sparse amount of information is available for protists and our understanding of regulation of gene expression in eukaryotes is limited to a few model organisms. Our research is aimed at oxymonads, poorly studied group of anaerobic protists, which inhabit digestive tract of some animals. In this study we focus on the genus *Monocercomonoides*.

Gene expression is modulated at multiple levels by many mechanisms. This thesis is focused on structure of promoter regions, 5' untranslated regions and basal transcription and translation initiation factors. Our results are compared to the closest studied relatives of *Monocercomonoides* – *Trichomonas vaginalis* and *Giardia intestinalis*.

We have identified several conserved motifs in promoter regions of *Monocercomonoides*, including TATA box and TATA-like motif. These motifs potentially play a role in the transcription regulation. 5' untranslated regions are relatively short (typically 20 – 30 nucleotides) and GC content in these regions is low compared to model organisms. In selected genes, the quality of the automatic prediction of UTR was verified by RACE. We have annotated sets of basic transcription (23 proteins) and initiation translation factors (30 proteins) in *Monocercomonoides* genome. These sets are similar to sets annotated in the genome of *Trichomonas vaginalis*. We have tried to purify transcription factors binding to the upstream region of a gene for GAPDH, but with inconclusive results.

Key words

5' UTR, gene regulatory region, TATA box, genome, oxymonads, transcription factor, translation factor

Obsah

1. Úvod.....	1
2. Literární přehled.....	2
2.1. Regulace genové exprese	2
2.2. Promotorové oblasti.....	4
2.2.1. TATA box.....	5
2.2.2. Iniciátor.....	6
2.2.3. Další motivy.....	7
2.3. Základní transkripční faktory	9
2.4. Nepřekládané oblasti.....	11
2.5. Translační iniciační faktory	13
2.6. Eukaryotická diverzita.....	17
2.7. Oxymonády.....	18
3. Cíle práce	20
4. Materiál a metody	21
4.1. Kultivace buněk <i>Monocercomonoides</i> sp. PA203.....	21
4.1.1. Příprava TYSGM - 9.....	21
4.2. Filtrace kultury <i>Monocercomonoides</i>	22
4.3. Kultivace buněk <i>Trichomonas vaginalis</i> G3.....	22
4.3.1. Příprava TYM	22
4.4. Práce s nukleovými kyselinami.....	23

4.4.1.	Izolace gDNA	23
4.4.2.	Izolace RNA pomocí ethanol-chloroformové extrakce	23
4.4.3.	Selekce mRNA.....	24
4.4.4.	Syntéza cDNA	24
4.4.5.	Polymerázová řetězová reakce – PCR.....	24
4.4.6.	Agarózová elektroforéza.....	24
4.4.7.	Přečištění a koncentrace DNA	25
4.4.8.	Sekvenace DNA Sangerovou metodou.....	26
4.4.9.	5' Rapid amplification of cDNA ends – 5' RACE.....	26
4.5.	Práce s bakteriemi	27
4.5.1.	Použitá média	27
4.5.2.	Transformace bakterií.....	28
4.5.3.	Izolace plazmidů	28
4.6.	Práce s proteiny	28
4.6.1.	Izolace transkripčních faktorů	28
4.6.2.	Polyakrylamidová denaturační proteinová elektroforéza (SDS PAGE)	30
4.6.3.	Hmotnostní spektroskopie	32
4.7.	Bioinformatické metody	33
4.7.1.	Základní práce se sekvenčními daty	33
4.7.2.	Charakterizace nepřekládaných oblastí	33
4.7.3.	Anotace proteinů.....	33
4.7.4.	Vyhledávání motivů v promotorových oblastech.....	34

5. Výsledky	35
5.1. Promotorové oblasti.....	35
5.2. Bioinformatická analýza UTRs.....	39
5.3. 5' RACE.....	44
5.4. Anotace základních transkripčních faktorů	45
5.5. Anotace iniciačních translačních faktorů.....	47
5.6. Purifikace transkripčních faktorů	48
6. Diskuse.....	52
6.1. Promotorové oblasti.....	52
6.2. 5' nepřekládané oblasti	52
6.3. Transkripční a translační faktory	55
7. Závěry	57
8. Literatura	58

1. Úvod

Největší část eukaryotické diverzity je tvořena protisty. Znalostí o tom, jak prvoci ovlivňují expresi svých genů, je ale poměrně málo. Z tohoto hlediska jsou zkoumány především druhy patogenní pro člověka. Například *Trichomonas vaginalis*, *Entamoeba histolytica*, *Giardia intestinalis*, plasmodia, trypanosomy a leishmanie.

U některých skupin prvoků (např. nálevníci nebo Euglenozoa, kam patří mimo jiné trypanosomy) se vyvinuly naprosto unikátní mechanismy regulace genové exprese, výrazně odlišné od učebnicové představy eukaryotické buňky. Ta je založena zejména na studiu kvasinek, octomilky nebo savčích buněk. Zkoumáním prvoků můžeme poznat odlišné způsoby fungování eukaryotické buňky.

Tato práce se zaměřuje na regulaci genů kódujících proteiny, tedy genů transkribovaných v eukaryotické buňce pomocí RNA polymerázy II. Z velkého množství způsobů, jakými buňka expresi genů může ovlivňovat, se blíže věnuje struktuře promotorových oblastí, 5' nepřekládaným oblastem a základním transkripčním a iniciačním translačním faktorům.

Modelovým organismem pro tuto práci je *Monocercomonoides* sp., kmen PA203. Mikroaerofilní, nepatogenní prvok žijící ve střevě činčily, který patří do skupiny oxymonád. Pro tento organismus se zatím nepodařilo vytvořit funkční transfekční systém. Tato práce by mohla pomoci tento systém vytvořit.

2. Literární přehled

2.1. Regulace genové exprese

Regulace genové exprese je jednou z nejdůležitějších schopností, kterou disponuje každá eukaryotická buňka. Tento proces umožňuje buňkám reagovat na změny vnějšího prostředí, udržet homeostázu a regulovat svůj vlastní vývoj. Přestože se jedná o velmi důležitý děj, prozatím není známo příliš mnoho informací o nemodelových organismech. Z protist jsou z tohoto hlediska velmi kuse prozkoumány zejména druhy patogenní pro člověka (např. *Trichomonas vaginalis*, *Entamoeba histolytica*, *Giardia intestinalis*, plasmodia nebo trypanosomy) (Gomez et al., 2010; Vaňáčová et al., 2003), o ostatních nevíme téměř nic.

Eukaryotické buňky jsou schopny regulovat expresi genů v každém jejím kroku. Ve většině případů je nejdůležitějším krokem iniciace transkripce. Vliv na tuto fázi má zejména primární struktura promotorových oblastí a následné rozeznání sekvenčních motivů v těchto oblastech pomocí transkripčních faktorů. Tyto proteiny umožňují nasednutí RNA polymerázy II, se kterou vytvářejí preiniciační komplex a umožňují efektivní spouštění transkripce (Gomez et al., 2010).

Vzniklá pre-mRNA dále prochází posttranskripčními úpravami. Dochází k vystřížení intronů, k 5' konci je přidána čepička z 7-methylguanositinu a 3' konec je polyadenylován 100 – 250 molekulami adeninu. Zralá mRNA má tři základní části. 5' nepřekládanou oblast (5' UTR), kódující oblast ohraničenou start a stop kodonem a 3' nepřekládanou oblast (3' UTR) (Mignone et al., 2002).

Nepřekládané oblasti hrají hlavní roli v regulaci genové exprese na posttranskripční úrovni. Mezi známé mechanismy, jakými expresi genů ovlivňují, patří zejména vliv na transport mRNA z jádra, stabilitu mRNA a efektivitu translace. Pro spuštění translace jsou nezbytné také eukaryotické iniciační faktory (viz kap. 2.5.). V dalších fázích se na regulaci genové exprese podílí i posttranslační úpravy a strukturální změny v chromatinu (Hinnebusch et al., 2016; Mignone et al., 2002).

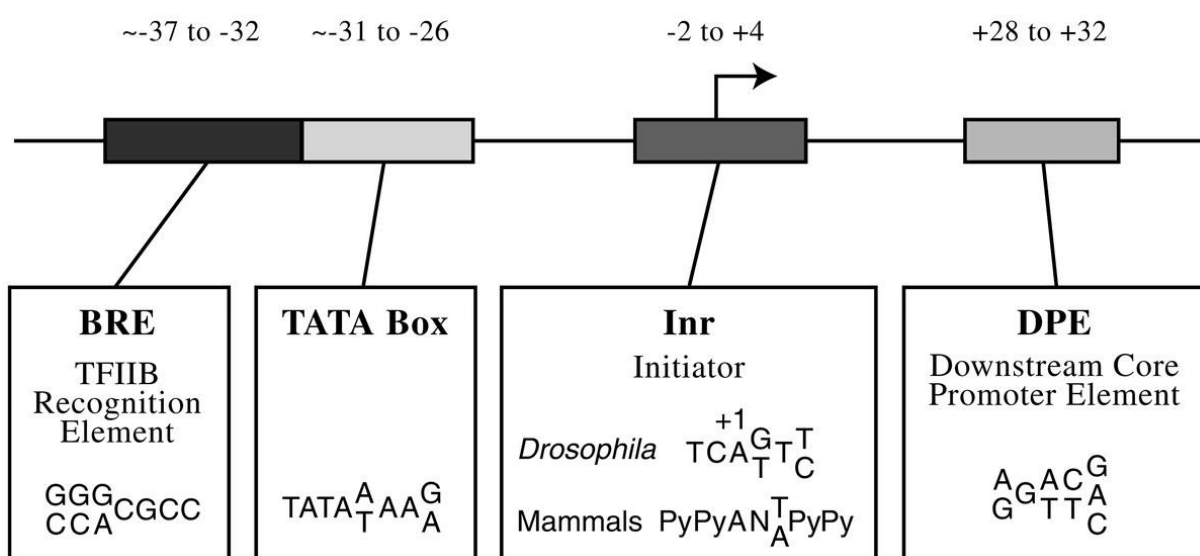
Vzhledem k naprosté odlišnosti transkripce a posttranskripčních úprav u trypanosomatid se jimi tato práce, s výjimkou popisu translačních faktorů, nezabývá. Regulace genové exprese u této skupiny je popsána například v přehledovém článku „*Gene expression regulation in trypanosomatids*“ (De Gaudenzi et al., 2011).

2.2. Promotorové oblasti

Transkripci strukturálních genů (genů kódujících proteiny) v eukaryotických buňkách zajišťuje RNA polymeráza II. Pro zahájení transkripce jsou nezbytné sekvenční motivy, které se nacházejí obvykle v oblasti 35 – 40 nukleotidů před i za místem začátku transkripce. S těmito motivy interagují především základní transkripční faktory (viz kap. 2.3.) a také přímo RNA polymeráza II (Atkinson a Halfon, 2014).

Nejdůležitější motivy jsou obvykle v rámci fylogeneze velmi konzervované. Za nejpůvodnější jsou považovány TATA box a BRE. Tyto motivy, společně s proteiny, které je rozeznávají, byly nalezeny kromě eukaryot i u archeí (Reeve, 2003). Na druhou stranu žádný známý motiv není universální pro všechny promotory a obvykle se vyskytují pouze v poměrně malé části promotorů (Juven-Gershon et al., 2008).

Nejvíce zkoumaným motivem, a pravděpodobně také nejdůležitějším, je TATA box. Dobře prozkoumaný je také motiv iniciátoru (Inr) a poté motivy DPE (downstream promoter element) a BRE (TFIIB recognition element) (obrázek č. 1). Kromě těchto základních motivů obsahují promotorové oblasti často také různé tlumiče a zesilovače transkripce (Atkinson a Halfon, 2014; Juven-Gershon et al., 2008).



Obrázek č. 1: Základní promotorové motivy. Převzato z (Smale a Kadonaga, 2003).

2.2.1. TATA box

TATA box je nejdůležitějším a nejvíce rozšířeným motivem. Je pravděpodobně také evolučně nejstarší. Původně se předpokládalo, že se tento motiv nachází u všech, nebo u většiny genů přepisovaných RNA polymerázou II. S rostoucím množstvím dat se toto nepotvrdilo a různé zdroje uvádějí výskyt TATA boxu mezi 5 – 40 % promotorů (Smale a Kadonaga, 2003). Ve studii provedené na 1031 předpokládaných promotorech lidských genů se TATA box nacházel ve 32 % (Suzuki et al., 2001).

Konsenzuální sekvence TATA boxu se podle různých autorů liší. Jedna z variant je TATA(A/T)AA(G/A) a ve studii provedené *in vitro* na jaderném extraktu z *Acanthamoeba castellanii* byla určena jako optimální pro rozpoznání TATA-binding proteinem sekvence TATATAAG (Huang et al., 1996). TATA box se obvykle nachází 31 až 26 nukleotidů před začátkem transkripce (Smale a Kadonaga, 2003).

U organismů, které nemají klasický TATA box, se ve stejné vzdálenosti od začátku transkripce poměrně často vyskytují AT bohaté motivy, které pravděpodobně plní funkci TATA boxu. Tyto motivy jsou často rozpoznávány TATA-binding proteinem, přestože mohou být sekvenčně poměrně dost odlišné od konsenzuální sekvence TATA boxu (Smale a Kadonaga, 2003).

U *Trichomonas vaginalis* byl popsán motiv M3, který byl nalezen u 12 % genů ve vzdálenosti 20 až 25 nukleotidů před začátkem transkripce. Konsenzuální sekvence tohoto motivu je (A/G/T)(A/G)C(G/C)G(T/C)T(T/A/G). Tato sekvence nejvíce připomíná Myb recognition element (MRE) a je rozpoznávána transkripčním faktorem M3BP. Poloha tohoto motivu a jeho role v začátku transkripce naznačují, že plní roli TATA boxu. Motiv, který by sekvenčně odpovídal TATA boxu, u trichomonády nalezen nebyl (Smith et al., 2011).

U *Giardia intestinalis* byla 30 nukleotidů před začátkem transkripce nalezena konzervovaná oblast obsahující řadu adeninů a thyminů, přímo o TATA box se však nejedná (Adam, 2000; Holberton a Marshall, 1995). Naopak u jiné diplomonády *Spironucleus salmonicida* byl nalezen sekvenčně konsenzuální TATA box u 81 % promotorů, nejčastěji mezi 60 a 20 nukleotidy před start kodonem (Xu et al., 2014).

U *Entamoeba histolytica* byl nalezen v typické vzdálenosti od místa začátku transkripce motiv označovaný jako TATA box, přestože jeho sekvence neodpovídá konsenzuální (GTATTTAAA(G/C)) (Pearson a Singh, 2010; Singh et al., 1997).

Hledání TATA boxu u *Plasmodium falciparum* je komplikováno bohatostí genomu na AT páry, které jsou v nekódujících oblastech zastoupeny v 90 %. Experimentálně bylo dokázáno, že v transkriptu genu *kahrp* se váže TATA-binding protein na sekvenci TATAA 81 nukleotidů před místem začátku transkripce. U genu *gbp-130* interaguje s motivem TGTA 186 nukleotidů před startem transkripce (Ruvalcaba-Salazar et al., 2005).

2.2.2. Iniciátor

Iniciátor (Inr) je motiv, který je vysoce konzervovaný v rámci eukaryot, funkčně je podobný TATA boxu a může fungovat nezávisle na něm. Často ale fungují společně a zajišťují tak velmi efektivní spouštění transkripce. V těchto případech hraje v determinaci místa startu translace důležitější roli TATA box (Smale a Kadonaga, 2003). Iniciátor se nachází kolem místa začátku transkripce (- 2 až + 4) a za jeho konsenzuální sekvenci se u savců považuje PyCANTPyPy (Javahery et al., 1994). Funkčně podobné motivy byly nalezeny i u archeí a některých protist (např. *Trichomonas vaginalis*).

Iniciátor u *Trichomonas vaginalis* má konsenzuální sekvenci TCAPy(T/A), což je poměrně překvapivě velmi blízké konsenzuální sekvenci u savců. Byl nalezen v 75 % promotorů a u *T. vaginalis* plní pravděpodobně nejdůležitější roli v regulaci genové exprese. To je dáno také absencí TATA boxu u tohoto prvoka (Carlton et al., 2007; Schumacher et al., 2003). Dále byl u trichomonády nalezen motiv M5 (CCTTT), který je funkční alternativou Inr a vyskytuje se tak především v genech, u kterých Inr chybí. V 85 % se nachází v promoterech genů pro ribozomální proteiny (Smith et al., 2011).

V promoterech *Giardia intestinalis* se těsně před startem transkripce nachází konzervovaná řada purinů (především adeniny), která je přerušena jedním až čtyřmi

pyrimidiny (hlavně thyminem). Tento motiv také plní funkci iniciátoru (Holberton a Marshall, 1995; Yee et al., 2000). V genomu *Spironucleus salmonicida* se u 33 % genů vyskytuje těsně před iniciačním kodonem sekvence minimálně tři adeninů, která pravděpodobně částečně plní roli iniciátoru (Xu et al., 2014).

U *E. histolytica* byl nalezen konzervovaný motiv s konsenzuální sekvencí AAAAATTCA. Nachází se v typické lokalizaci pro Inr a zřejmě plní jeho funkci, přestože jeho sekvence Inr neodpovídá (Pearson a Singh, 2010; Singh et al., 1997).

2.2.3. Další motivy

Kromě Iniciátoru a motivů M3 a M5 byl u genů, které kódují histony, v genomu *T. vaginalis* popsán motiv M2. Nachází se 20 až 40 nukleotidů před startem translace (Cong et al., 2010).

Ve všech genech kódujících histony byl u *G. intestinalis* nalezen 15 bp sekvenční motiv označovaný jako him (histone motif), u některých byly nalezeny sekvence identifikované jako CAT boxy (Yee et al., 2007).

Třetím identifikovaným motivem v genomu *E. histolytica* je konzervovaná sekvence GAAC (GAACT). Jeho pozice v rámci promotorů různých genů je variabilní a může zajišťovat transkripci nezávisle na TATA boxu i iniciátoru (Singh a Rogers, 1998). Na genu *hgl5* (galaktose-inhibitable lectin heavy subunit gene) bylo popsáno několik regulačních elementů (URE), některé transkripci tlumí, jiné zesilují (Bhattacharya et al., 2000; Gilchrist et al., 1997; Purdy et al., 1996).

U nálevníků jsou promotorové oblasti prozkoumány jen málo. U *Tetrahymena thermophila* nebyl potvrzen TATA box ani jiné typické eukaryotické motivy. Při studiu genů pro metallothioneiny byl nalezen konzervovaný motiv MTCM1 (Metallothionein Conserved Motif 1) se sekvencí GTGTGAATCATGAGT. V promotorové oblasti se vyskytuje dvakrát až třináctkrát a experimentálně bylo dokázáno, že mutace v této sekvenci vede až ke kompletní inhibici transkripce (Díaz et al., 2007; Formigari et al., 2010).

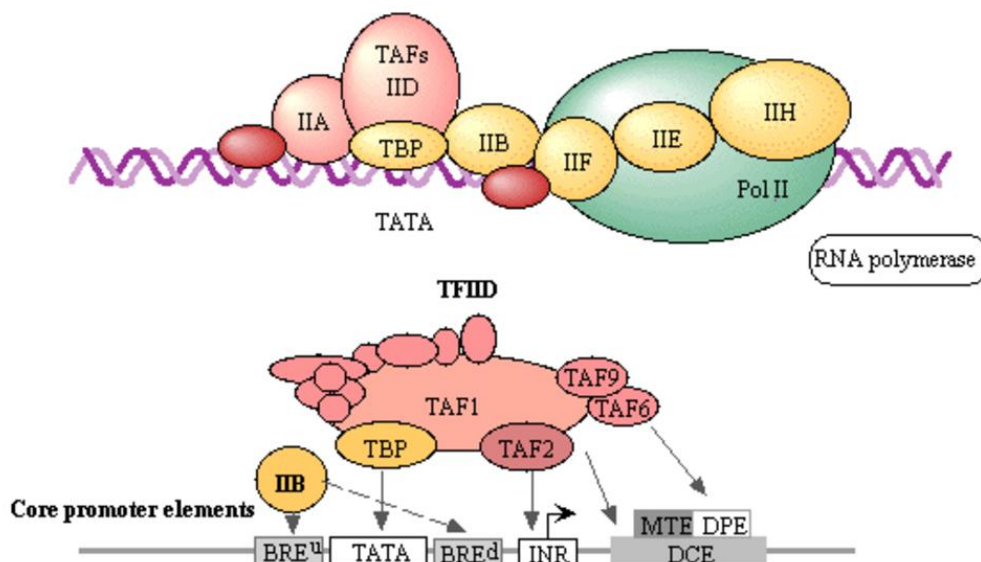
V genech pro heat shock proteiny (*hsp*) u *Plasmodium falciparum* byl nalezen motiv nazvaný G-box se sekvencí (A/G)NGGGG(C/A). Nachází se v 7 z 18 *hsp* genů, často v několika kopiích (Militello et al., 2004). Dále byly v jednotlivých genech identifikovány motivy nezbytné pro správnou transkripci daného genu, které nejsou podobné žádnému konzervovanému motivu u jiných eukaryot (Dechering et al., 1999; Horrocks a Lanzer, 1999; Osta et al., 2002).

2.3. Základní transkripční faktory

Transkripční faktory jsou proteiny, které rozpoznávají a vážou se na specifické sekvence DNA a tím regulují transkripci. Popřípadě mohou transkripci ovlivňovat nepřímo vazbou na jiné transkripční faktory.

Pro zahájení transkripce jsou nezbytné základní transkripční faktory („General transcription factors“, nebo „Basal transcription factors“). Jedná se o šest proteinů (TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, a TFIIH), které se s RNA polymerázou II podílí na tvorbě preiniciačního komplexu a rozpoznávají konzervované motivy v promotorových oblastech (obrázek č. 2). Tyto proteiny jsou minimálně v rámci metazoi velmi konzervované (Orphanides et al., 1996).

V prvním kroku rozpozná promotorové sekvence transkripční faktor IID, proteinový komplex složený z TATA-binding proteinu (TBP) a několika TBP asociovaných faktorů (TAFs). Vazba TFIID způsobí ohnutí DNA, což umožní navázání TFIIA a následně dalších faktorů. TFIIB stabilizuje vazbu DNA a TFIID a pomáhá vazbě mezi RNA polymerázou II a TFIIIF. Následně se navážou TFIIE a TFIIH a zkompletují preiniciační komplex (Albright a Tjian, 2000; Reese, 2003).



Obrázek č. 2: Základní transkripční faktory.

Převzato z www.genome.jp/kegg-bin/show_pathway?ko03022, 18. 6. 2016.

V genomu *Trichomonas vaginalis* se nacházejí všechny základní transkripční faktory s výjimkou TFIIA. V případě transkripčních faktorů IIF a IIH chybí v porovnání s typickou eukaryotickou buňkou některé podjednotky (data z databáze KEGG, Kanehisa a Goto, 2000).

Giardia intestinalis je z hlediska základních transkripčních faktorů unikátní organismus. V genomu byly nalezeny pouze některé podjednotky TFIIH a velmi divergentní TBP. Tento protein je přitom obvykle velmi konzervovaný v rámci všech eukaryot a jeho homology jsou známy i z archeí. To dělá z giardie důležitý model pro studium evoluce transkripčního aparátu (Best et al., 2004; Morrison et al., 2007). Přestože se u *Spironucleus salmonicida* nachází konsenzuální TATA box, TBP nalezený v tomto genomu je překvapivě ještě divergentnější než u giardie (Xu et al., 2014).

TATA-binding protein popsáný u *Plasmodium falciparum* je také poměrně divergentní na úrovni primární struktury (pouze 42 % sekvenční identity s TBP u kvasinky), na druhou stranu většina oblastí vázající DNA je konzervována (McAndrew et al., 1993). Díky své divergenci je TBP u plasmodia považován za možný cíl pro antimalarika. Navíc nezbytnost TBP pro iniciaci transkripce snižuje možnost vzniku rezistence (Buendía-Orozco et al., 2005; Ruvalcaba-Salazar et al., 2006). U plasmodia byly zkoumány také interakce TBP a TFIIIE s promotory (Gopalakrishnan et al., 2009).

TATA-binding protein byl podrobně popsán u *Entamoeba histolytica*. Transkript pro tento protein má dlouhou 5' nepřekládanou oblast (420 nukleotidů), sekvenční identita k lidskému TBP je 55 %, 37 % k homologu u *P. falciparum* a 54 % k TBP popsánému u *Acanthamoeba castellanii* (Luna-Arias et al., 1999; Wong et al., 1992). TBP nalezený u *Tetrahymena thermophila* má k lidskému homologu sekvenční identitu 70 % (Stargell a Gorovsky, 1994).

Zvláštní transkripční iniciační faktor byl nalezen u obrněnky *Cryptosporidium parvum*. Jde o protein se silnou sekvenční homologií k TATA-binding proteinu, který obsahuje substituce čtyř aminokyselin na místech, které vážou DNA. Tyto substituce jsou typické pro skupinu TBP-like faktorů. Tento protein váže silněji než klasický TATA box sekvenci TTTT (Guillebault et al., 2002).

2.4. Nepřekládané oblasti

Nepřekládané oblasti (UTRs) ovlivňují stabilitu mRNA, její lokalizaci v buňce a společně s translačními faktory ovlivňují úroveň translace. Hrají tedy nejdůležitější úlohu v rámci posttranskripční regulace genové exprese a zajišťují její přísnou kontrolu (Bashirullah et al., 2001; Jansen, 2001).

Mechanismů regulace pomocí nepřekládaných oblastí je několik. Sekvence UTR může interagovat s RNA vazebnými proteiny, přičemž na rozdíl od DNA vazebných proteinů v promotorových oblastech je vazba v těchto případech závislá na primární i sekundární struktuře UTR. Stejně tak může UTR interagovat s komplementární nekódující RNA. Pro regulaci jsou také důležité repetitivní oblasti, které se v nich mohou vyskytovat. V neposlední roli se nepřekládané oblasti také mohou podílet na „processingu“ pre-mRNA (Mignone et al., 2002; Sweeney et al., 1996).

Hlavními regulačními prvky u 5' UTRs jsou uORFs (upstream open reading frames) a upstream AUGs. Upstream ORFs jsou sekvence, které mají start a stop kodon umístěné před hlavním čtecím rámcem, uAUGs mají start kodon bez stop kodonu umístěného před kódující oblastí. V lidském genomu tyto elementy obsahuje přibližně polovina genů. Jejich přítomnost obvykle snižuje množství produkovaného proteinu (Araujo et al., 2012).

Průměrná délka 5' UTR se udává mezi 100 až 200 nukleotidy napříč taxonomickými skupinami. 3' UTRs jsou obvykle delší a také rozpětí jejich délek je vyšší, a to i mezi poměrně příbuznými organismy (například v rámci savců). Průměrná délka je mezi 200 nukleotidy u rostlin až po více než 1000 nukleotidů u člověka. Tato porovnání ale nezahrnují prvky (Pesole et al., 2001). I jeden nukleotid může tvořit funkční 5' UTR, jak ukázala *in vitro* studie na savčích buňkách (Hughes a Andrews, 1997).

Nepřekládané oblasti obsahují poměrně často introny, přičemž v 5' UTRs jsou častější (30 % genů u metazoi) než v 3' UTRs (1 – 11 % genů u metazoi). Introny ve

3' nepřekládaných oblastech jsou obvykle výrazně delší. 5' nepřekládané oblasti jsou bohaté na GC páry (60 % u teplokrevných obratlovců), 3' nepřekládané oblasti obsahují v průměru 45 % GC párů u teplokrevných obratlovců. Obsah GC párů také negativně koreluje s délkou nepřekládaných oblastí (Pesole et al., 2001).

Nepřekládané oblasti u *Giardia intestinalis* jsou díky kompaktnosti celého genomu obvykle velmi krátké. 5' UTRs mají délku většinou do 14 nukleotidů, 3' UTRs mezi 10 a 30 nukleotidy a obsahují konsenzuální polyadenylační sekvenci AGUAAA (Adam, 2000; Morrison et al., 2007; Yee et al., 2000).

U *Trichomonas vaginalis* jsou nepřekládané oblasti také velmi krátké, 3' UTR obsahují polyadenylační signál s konsenzuální sekvencí UAAA, který se překrývá s terminačním kodonem UAA (Espinosa et al., 2002; Fuentes et al., 2012).

Polyadenylační signál ve 3' nepřekládaných oblastech *Entamoeba histolytica* je UA(A/U)UU, nepřekládané oblasti jsou opět krátké (Clark et al., 2007; Zamorano et al., 2008).

5' nepřekládané oblasti u *Plasmodium falciparum* jsou naopak neobvykle dlouhé, v průměru 346 nukleotidů, a obsahují velké množství uAUGs a uORFs. Nejvíce se jich vyskytuje u *var* genů, tedy u genů zodpovědných za antigenní variabilitu plasmodia (Kumar et al., 2015; Watanabe et al., 2002). *Var* geny mají upstream sekvence čtyř konzervovaných typů: *upsA*, *B*, *C* a *E*. V *upsC* byl popsán 101 bp dlouhý regulační element MEE (Brancucci et al., 2014). U *P. vivax* je medián délky 5' UTR 295 nukleotidů, u 3' UTR 203 nukleotidů (Zhu et al., 2016).

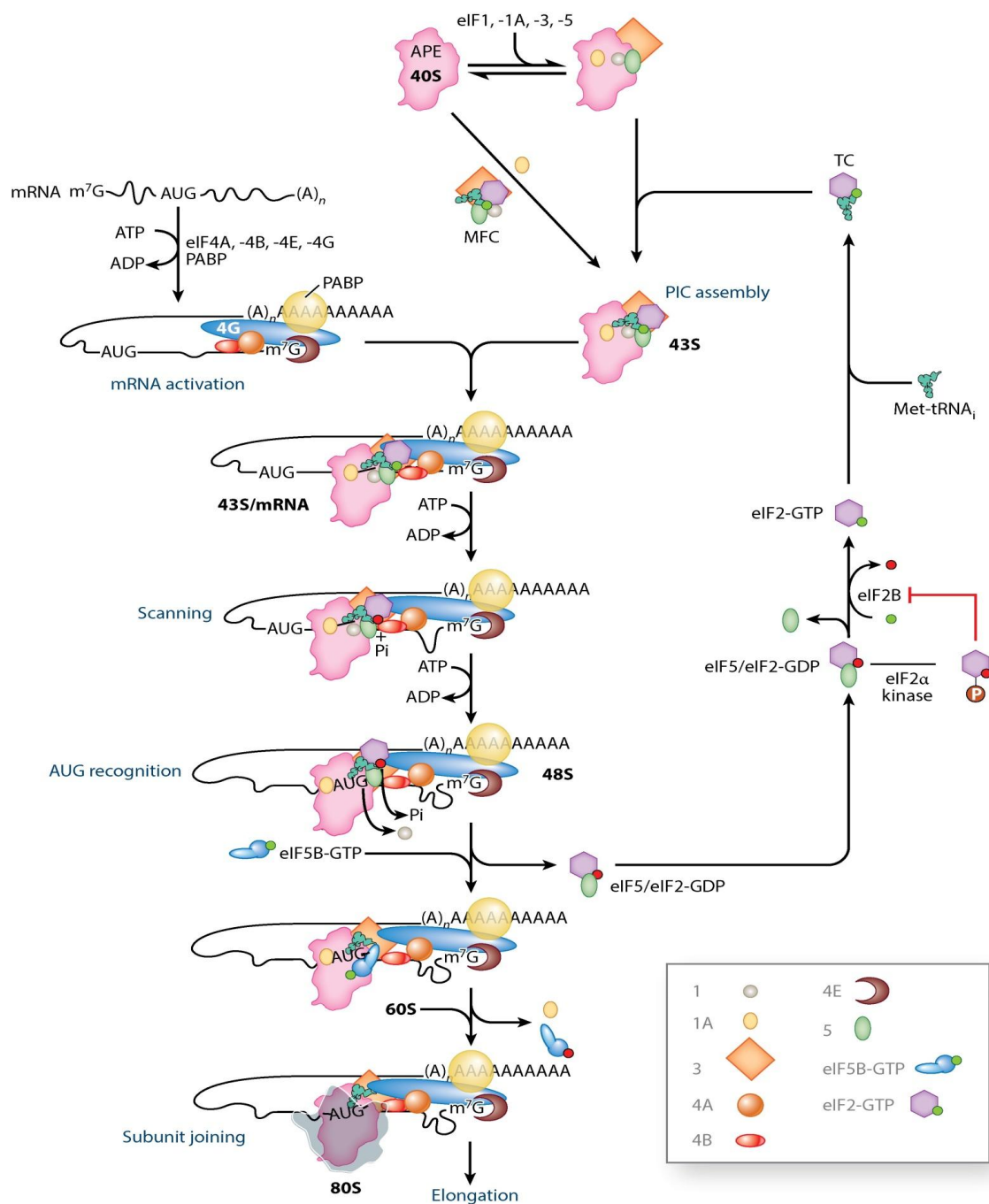
2.5. Translační iniciační faktory

Eukaryotické translační iniciační faktory jsou proteiny, které se podílí na tvorbě funkčního ribozomu, umožňují vazbu ribozomu na mRNA a mají vliv na regulaci translace. Naše znalosti průběhu translace jsou založeny zejména na *Saccharomyces cerevisiae*, octomilce, rostlinách a savčích buňkách.

Start kodon je v případě eukaryot u většiny mRNA nalezen skenovacím mechanismem. V prvním kroku dojde k navázání speciální iniciační tRNA nesoucí methionin (Met-tRNA_i) a aktivní formy eukaryotického iniciačního faktoru 2 (eIF2) nesoucí GTP na malou ribozomální podjednotku (40S). Této reakci pomáhají také eIF1, eIF1A, eIF5 a eIF3. Výsledkem je vytvoření 43S preiniciačního komplexu (PIC). Ten se následně váže na 5' konec mRNA (viz obrázek č. 3) (Hinnebusch, 2014; Hinnebusch et al., 2016).

Navázání PIC předchází aktivace mRNA, při které čepičku z 7-methylguanosu na 5' konci mRNA rozpoznává proteinový komplex eIF4F. Ten se skládá ze tří podjednotek: eIF4E (protein vázající čepičku), eIF4G (centrální protein s vazebnými místy pro mRNA, jiné iniciační faktory a poly(A)-binding protein) a eIF4A, což je helikáza, která rozplétá sekundární struktury v mRNA. Její helikázová aktivita je zesilována pomocí proteinů eIF4G a eIF4B (Hinnebusch, 2014; López-Lastra et al., 2005).

Pomocí štěpení ATP se molekula mRNA posunuje po 40S podjednotce ribozomu, dokud Met-tRNA_i nerozpozná start kodon AUG. Nejčastěji je to první AUG kodon od 5' konce mRNA. Rozpoznání start kodonu je usnadněno konzervovaným motivem v okolí nazývaným jako sekvence Kozakové s konsenzuální sekvencí pro savce (A/G)CCAUGG (Kozak, 1986). Po rozpoznání start kodonu je na eIF2 hydrolyzováno GTP, dojde k disociaci iniciačních faktorů a připojení velké ribozomální podjednotky (60S). To je zajišťováno GTPázou eIF5B (Jackson et al., 2010).



Obrázek č. 3: Iniciační translace. Převzato z (Hinnebusch, 2014). Translace začíná sestavením 43S preiniciačního komplexu (PIC), znázorněno jedním krokem přes „multifactor komplex“ (MFC) a ve dvou krocích, kde se eukaryotické iniciační faktory (eIFs) eIF1, -1A, a -3 vážou na malou ribozomální podjednotku. Následně se naváže ternární komplex (TC) a eIF5. 43S PIC je poté navázán na 5' konec aktivované mRNA s proteiny. Po nalezení start kodonu disociuje eIF1 a vytvoří se uzavřená konformace PIC. eIF5B zajistí připojení velké ribozomální podjednotky za disociace eIF5B-GDP a eIF1A. eIF2-GDP je následně recyklován zpět na eIF2-GTP. GTP je znázorněno zeleně, GDP červeně.

Kromě klasického spouštění translace, pro kterou je nezbytná přítomnost 5' čepičky, je možná regulace translace pomocí IRES (Internal ribosome entry site). Jedná se o sekvenci v upstream oblasti genu, nejčastěji v 5' nepřekládané oblasti, která svou sekundární strukturou umožňuje nasednutí ribozomu a spuštění translace. Tento mechanismus využívají kromě eukaryotických RNA zejména některé viry, u kterých byl také poprvé objeven (Jang et al., 1988; López-Lastra et al., 2005).

Základní set translačních iniciačních faktorů, které zdělila všechna eukaryota od svého archeálního předka, zahrnuje eIF1, eIF1A, eIF2 (všechny tři podjednotky), eIF2B (podjednotky 1, 2 a 5), eIF4A, eIF5B a eIF6 (Aravind a Koonin, 2000; Jagus et al., 2012).

Nejvíce je u protist prozkoumaný protein eIF4E. Ve studii zahrnující prvoky ze všech „superskupin“ bylo popsáno, že často kódují několik typů tohoto faktoru (podobně jako mnohobuněčné organismy). Tyto genové rodiny se však liší od těch, které se nalézají u živočichů nebo rostlin (Jagus et al., 2012). V genomu *Giardia intestinalis* jsou blíže popsány dva homology proteinu eIF4E. Oba dva se vážou na čepičku na 5' konci mRNA, pouze jeden z nich je ale nezbytný pro úspěšné zahájení translace (Li a Wang, 2005).

U *Trichomonas vaginalis* byly zkoumány dva homology proteinu eIF5A, zejména jeho posttranslační modifikace. Ty zahrnují fosforylaci, glykosylaci a také neobvyklou hypusinaci (Carvajal-Gamez et al., 2012; Quintas-Granados et al., 2016). Stejný protein byl popsán i u *Leishmania donovani* (Singh et al., 2014) a *Trypanosoma brucei*, u které byla prokázána pomocí RNA interference nezbytnost tohoto proteinu a jeho modifikace pro přežití trypanosomy (Chung et al., 2013; Nguyen et al., 2015).

Ve srovnávací bioinformatické analýze eIF3 u patogenních exkavát bylo u trichomonády objeveno pět podjednotek tohoto proteinového komplexu, u giardie čtyři a u trypanosomatid (*Leishmania major* a *Trypanosoma brucei*) jedenáct. Kompletní eIF3 se skládá u savců ze 13 podjednotek, u *Saccharomyces cerevisiae* je přítomno šest z nich. Přítomnost většiny podjednotek u trypanosomatid naznačuje časnou evoluci

tohoto komplexu, zjednodušená stavba u ostatních eukaryot je pravděpodobně způsobena sekundární ztrátou některých podjednotek (Rezende et al., 2014).

Na základě sekvencí genů pro podjednotku 3 eIF2 ze zástupců diplomonád, parabasalidů, mikrosporidií a archeálních homologů této podjednotky byla zkoumána fylogenetická pozice tohoto proteinu v rámci GTPáz (Keeling et al., 1998).

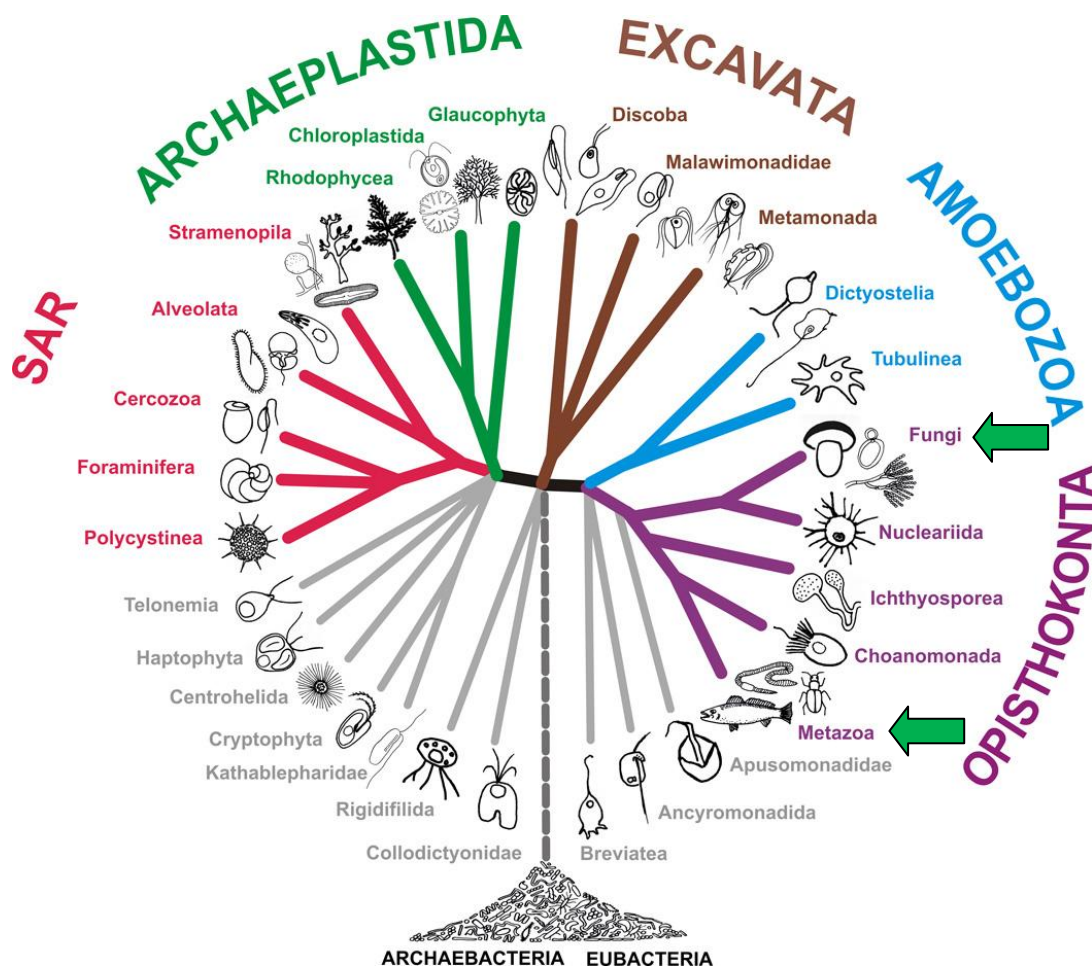
Fosforylace podjednotky 1 eIF2 je konzervovaný způsob, jakým eukaryotické buňky potlačují translaci. Zejména v latentních fázích infekce ho ve velké míře využívají *Plasmodium* a *Toxoplasma*. U *P. falciparum* byl také identifikován eIF5 a dvě podjednotky eIF3 (Bunnik et al., 2016; Tellier et al., 2016; Zhang et al., 2013). Vliv fosforylace eIF2 u podjednotky 1 na expresi genů byl také zkoumán u *Trypanosoma cruzi* (da Silva Augusto et al., 2015; Tonelli et al., 2011).

Vzhledem k omezené možnosti regulace genové exprese na transkripční úrovni a zásadní roli regulace na úrovni translace jsou u trypanosomatid poměrně podrobně popsány jednotlivé podjednotky proteinového komplexu eIF4F. V genomu je obvykle kódováno několik homologů podjednotek, které se mohou funkčně mírně lišit (Dhalia et al., 2006, 2005; Freire et al., 2014a, 2014b, 2011; Moura et al., 2015; Pereira et al., 2013).

2.6. Eukaryotická diverzita

Eukaryota jsou dnes podle převládající koncepce řazena do pěti říší („superskupin“): Excavata, Opisthokonta, Amebozoa, Archaeplastida a SAR (obrázek č. 4). Poměrně velký počet organismů se prozatím do žádné z těchto skupin zařadit nepodařilo, tyto organismy jsou označovány jako „*incertae sedis*“ – nejasného zařazení (na obrázku č. 4 vyznačeny šedě) (Adl et al., 2012; Simpson a Roger, 2004).

Většina informací o genové expresi pochází z modelových organismů, zejména kvasinek nebo některých živočichů. V rámci celkové eukaryotické diverzity jsou však kvasinky a živočichové poměrně blízce příbuzné organismy, součást říše Opisthokonta (na obrázku č. 4 označeno zelenými šipkami). Vyvozovat pouze z těchto dat obecně platná pravidla pro všechna eukaryota tedy není možné. Proto jsou potřeba údaje i z jiných říší, kde je zdaleka největší biodiverzita tvořena protisty.



Obrázek č. 4: Diverzita eukaryot, převzato z (Adl et al., 2012), upraveno.

2.7. Oxymonády

Oxymonády jsou skupinou mikroaerofilních, endosymbioticky žijících prvoků. Společně s volně žijícími rody *Trimastix* a *Paratrimastix* tvoří skupinu Preaxostyla (Excavata: Metamonada) (Adl et al., 2012; Zhang et al., 2015).

Zástupci rodů *Trimastix* a *Paratrimastix* se nejčastěji vyskytují v hypoxických sladkovodních i mořských sedimentech. Jsou to bičíkovci s ventrální rýhou, kterou prochází zpětný bičík, mají tedy typickou exkavátní morfologii. Oxymonády jsou morfologicky značně diverzifikovaná skupina. Nejčastěji se vyskytují v zadním střevě xylofágního hmyzu (především nižší termiti a švábi rodu *Cryptocercus*), ale některé druhy rodu *Monocercomonoides* se vyskytují i u obratlovců a naxylofágního hmyzu. Role oxymonád v celém ekosystému střev xylofágního hmyzu není dobře objasněna. U větších druhů se předpokládá podíl na trávení celulózy (Kulda a Nohýnková, 1978; Radek, 1999).

Oxymonády patří k nejméně prozkoumaným prvokům a až na naprosté výjimky je nelze kultivovat mimo jejich hostitele. Tito prvoci jsou velmi zajímaví zejména z pohledu evoluce mitochondrie. U oxymonád došlo k její naprosté ztrátě, což u žádného jiného organismu popsáno nebylo (Karnkowska et al., 2016). Naopak organela mitochondriálního původu byla popsána u jejich volně žijících příbuzných rodů *Trimastix* a *Paratrimastix*, přestože její fyziologická funkce není příliš objasněna (Zubáčová et al., 2013).

Je popsáno asi 150 druhů oxymonád, které se dělí do pěti čeledí: Polymastigidae, Saccinobaculidae, Pyrsonymphidae, Streblomastigidae a Oxymonadidae, a samostatného rodu *Opisthomitus* (Brugerolle a Lee, 2000). Rod *Monocercomonoides*, kterým se zabývá tato práce, patří do čeledi Polymastigidae.

Tato čeleď zahrnuje malé prvoky se čtyřmi bičíky uspořádanými do dvou párů, jeden je zpětný a zčásti přisedlý na buňku. Buňky mají tenký, nekontraktilní axostyl (pás mikrotubulů, který prochází podélnou osou buňky), jádro je kryto mikrotubuly, vytvářejícími peltu (Radek, 1994). Do této čeledi jsou řazeny rody

Polymastix, *Paranotila*, *Monocercomonoides* a *Tubulimonoides* (Brugerolle a Lee, 2000; Krishnamurthy a Sultana, 1976).

Zástupci rodu *Monocercomonoides* se vyskytují zejména ve střevech larev tiplic a švábů. Poměrně velký počet zástupců však můžeme nalézt i u obratlovců, například hlodavců, dobytka, plazů nebo obojživelníků (Kulda a Nohýnková, 1978). Diverzitou tohoto rodu se blíže zabývala ve své diplomové práci Jitka Vlasáková (Vlasáková, 2014).

Kmen PA203, kterému se tato práce věnuje, byl izolován ze střeva činčily. Jako u jediné oxymonády je pro tento kmen dostupné velké množství transkriptomických a genomických dat (Karnkowska et al., 2016).

3. Cíle práce

Cíle této diplomové práce jsou:

- Najít konzervované motivy v promotorových oblastech oxymonády *Monocercomonoides* sp. PA203.
- Charakterizovat 5' nepřekládané oblasti v genomu *Monocercomonoides* sp. PA203 pomocí bioinformatických metod a ověřit automatické predikce nepřekládaných oblastí pomocí 5' RACE experimentů.
- Anotovat základní transkripční a iniciační translační faktory v genomu *Monocercomonoides* sp. PA203.
- Purifikovat transkripční faktory *Monocercomonoides* sp. PA203 a *Trichomonas vaginalis* G3 pomocí vazby na „upstream“ sekvenci konkrétního genu.

4. Materiál a metody

4.1. Kultivace buněk *Monocercomonoides* sp. PA203

Rod *Monocercomonoides* je kultivovatelný pouze v polyxenické kultuře s bakteriemi. Kmen PA203 jsme pěstovali v médiu TYSGM - 9 (Diamond, 1982), do kterého jsme několik dní před samotnou inokulací prvoka přidali necharakterizovanou směs bakterií separovanou v minulosti z kultury *Monocercomonoides* sp. 203 a pasážovanou odděleně. Přibližně 1 ml kultury jsme inokulovali do 10 ml média se směsí bakterií jednou týdně a kultivovali v 15ml plastových zkumavkách při 37 °C.

Pro získání většího počtu buněk jsme oxymonády nejprve inokulovali do 45 ml bakterizovaného média v 50ml plastových zkumavkách a po třech až čtyřech dnech jsme celý objem přidali do 500 ml bakterizovaného média v kultivační láhvi.

4.1.1. Příprava TYSGM - 9

- Složení TYSGM - 9
 - trypton: 2 g
 - kvasničný autolyzát: 1 g
 - K₂HPO₄: 2,8 g
 - KH₂PO₄: 0,4 g
 - NaCl: 7,5 g
 - dH₂O: 950 ml
 - inaktivované koňské sérum (GIBCO): 15 ml
- Všechny složky média kromě koňského séra jsme rozpustili přibližně v 800 ml destilované vody, upravili pH na 7,2 a doplnili jsme destilovanou vodou do 950 ml. Médium jsme sterilizovali v autoklávu (120 °C, 20 minut) a po vychladnutí jsme sterilně přidali 30 ml inaktivovaného bovinního, nebo koňského séra. Inaktivace séra probíhala 30 minut při 56 °C. Médium jsme uchovávali při 4 °C.

4.2. Filtrace kultury *Monocercomonoides*

Pro minimalizování bakteriální kontaminace jsme kulturu buněk před izolací DNA, RNA nebo buněčného lyzátu přefiltrovali. Nejprve jsme kulturu filtrovali přes filtrační papír, abychom odstranili velké shluky bakterií a následně přes filtry Whatman s 3 μ m póry. *Monocercomonoides* se hromadil nad filtrem. Jeden filtr jsme používali maximálně k filtraci 250 ml kultury, abychom zabránili ucpávání filtru. Buňky oxymonády jsme následně jednou nebo dvakrát promyli 250 ml média TYSGM - 9. Pro minimalizaci ztrát a zachování co možná nejlepšího fyziologického stavu buněk je nutné filtraci provést co nejrychleji, proto jsme po celou dobu filtrování urychlovali protékání média přes filtr vířením kapaliny pomocí plastových Pasteurových pipet.

4.3. Kultivace buněk *Trichomonas vaginalis* G3

Trichomonas vaginalis jsme kultivovali v axenické kultuře v médiu TYM (Diamond, 1957) při 37 °C. Do přibližně 10 ml čerstvého média jsme inokulovali buňky jednou za 48 hodin, kultivovali jsme v 15ml plastových zkumavkách.

4.3.1. Příprava TYM

- Složení TYM
 - K₂HPO₄: 0,8 g
 - KH₂PO₄: 0,8 g
 - kyselina askorbová: 0,2 g
 - L-cystein: 1 g
 - maltóza: 5 g
 - kvasničný autolyzát: 10 g
 - trypton: 20 g
 - citrát železnato-amonný: 1 ml
 - dH₂O: 900 ml
 - inaktivované koňské sérum (GIBCO): 100 ml

- Všechny složky média kromě koňského séra jsme rozpustili přibližně v 800 ml destilované vody, upravili pH na 6,2 a doplnili jsme destilovanou vodou do 900 ml. Médium jsme sterilizovali v autoklávu (120 °C, 20 minut) a po vychladnutí jsme sterilně přidali 100 ml inaktivovaného koňského séra. Inaktivace séra probíhala 30 minut při 56 °C. Médium jsme uchovávali ve 4 °C.

4.4. Práce s nukleovými kyselinami

4.4.1. Izolace gDNA

K izolaci genomické DNA jsme použili DNeasy Blood & Tissue Kit (QIAGEN), postupovali jsme podle přiloženého protokolu. Koncentraci DNA jsme změřili na spektrofotometru Nanodrop a izolovanou DNA uchovávali v -20 °C.

4.4.2. Izolace RNA pomocí ethanol-chloroformové extrakce

- Chemikálie
 - Tri Reagent (Ambion)
 - chloroform
 - isopropanol
 - 75% ethanol (-20 °C)
 - dH₂O
- Přefiltrovanou kulturu *Monocercomonoides* jsme centrifugovali 10 minut při 3000 g ve 4 °C. Odstranili jsme supernatant a pelet jsme resuspendovali v 1 ml Tri Reagent. Směs jsme převedli do 1,5 ml mikrozkušavky a přidali jsme 200 µl chloroformu. Reakci jsme důkladně protřepali na vortexu a poté jsme ji nechali 2 minuty stát při laboratorní teplotě. Následně jsme mikrozkušavku centrifugovali 15 minut při 15 000 g ve 4 °C. Reakce se rozdělila do tří fází. Svrchní fázi, která obsahuje RNA, jsme opatrně odsáli a přenesli do nové mikrozkušavky a přidali jsme k ní 500 µl isopropanolu. Opět jsme reakci protřepali na vortexu a nechali stát 10 minut při laboratorní teplotě. Znovu jsme směs centrifugovali ve 4 °C při 15 000 g, tentokrát 10 minut. Odstranili jsme

supernatant a k peletu jsme přidali 1 ml vychlazeného ethanolu. Zopakovali jsme předchozí centrifugaci. Odstranili jsme supernatant a pelet nechali částečně vysušit. Ten jsme následně rozpustili v 20 μ l destilované vody. RNA jsme uchovávali v -80 °C, koncentraci jsme změřili pomocí spektrofotometru Nanodrop.

4.4.3. Selekcce mRNA

K selekci mRNA z celkové RNA jsme použili Dynabeads® mRNA purification kit (Invitrogen), postupovali jsme podle přiloženého protokolu. Koncentraci mRNA jsme změřili pomocí spektrofotometru Nanodrop a mRNA jsme uchovávali v -80 °C.

4.4.4. Syntéza cDNA

K syntéze cDNA jsme použili SMARTer PCR cDNA Synthesis Kit (Clontech) a postupovali jsme podle přiloženého protokolu. cDNA jsme uchovávali v -20 °C.

4.4.5. Polymerázová řetězová reakce – PCR

Pro amplifikaci konkrétních úseků DNA jsme využívali předpřipravené směsi pro PCR: EmeraldAmp MAX PCR Master Mix (TaKaRa), PrimeSTAR Max Premix (TaKaRa) a DNA polymerázu Ex Taq Hot Start version (TaKaRa). Program thermocycleru jsme nastavovali podle doporučených protokolů a teplotu annealingu podle konkrétních primerů. Používali jsme thermocyclery MyCycler thermal cycler (Bio-Rad) a T100 thermal cycler (Bio-Rad).

Pro ověření přítomnosti požadovaných plazmidů v transformovaných bakteriích jsme používali Colony PCR. Reakce probíhala podle doporučených protokolů, jako templát byl použit vzorek bakteriální kolonie.

4.4.6. Agarózová elektroforéza

Pomocí horizontální agarózové elektroforézy jsme zjišťovali přítomnost a délku úseků DNA z PCR reakcí a také jsme ji využívali k izolaci úseků DNA o konkrétní délce.

- Chemikálie
 - prášková agaróza
 - TAE pufr (Bio-Rad)
 - SYBR Safe (Thermo Scientific)
 - GeneRuler DNA Ladder Mix (Thermo Scientific)
 - 6x DNA loading dye (Thermo Scientific)
- V baňce jsme k 40 ml 1 × TAE pufru přidali 0,4 g práškové agarózy a nechali směs zahřívát v mikrovlnné troubě, dokud nezačala vřít a agaróza se dokonale nerozpustila. Poté jsme roztok ochladili a přidali 40 µl SYBR Safe a nechali 20 minut zatuhnout. Na gely jsme aplikovali elektrické pole 10 V/cm přibližně 30 minut. Gely jsme fotografovali pod UV zářením a případně jsme z nich pod modrým světlem vyřezávali fragmenty DNA. V tomto případě jsme k čištění DNA používali Zymoclean Gel DNA recovery kit (Zymo research), nebo QIAquick Gel Extraction Kit (QIAGEN). Postupovali jsme podle příložených protokolů.

4.4.7. Přечиštění a koncentrace DNA

K přечиštění PCR produktů jsme používali High Pure PCR Product Purification Kit (Roche), postupovali jsme podle příloženého protokolu. V případě, že toto přечиštění nebylo dostatečné nebo jsme potřebovali DNA zakoncentrovat použili jsme metodu přesrážení DNA pomocí ethanolu a octanu sodného.

- Chemikálie
 - Octan sodný (3M)
 - Ethanol
- K 20 µl PCR produktu jsme přidali 2 µl sodného acetátu a 50 µl 96% ethanolu. Směs jsme promíchali a nechali stát při laboratorní teplotě 15 minut. Poté jsme reakci centrifugovali 30 minut při 13 000g ve 4 °C. Opatrně jsme odsáli supernatant a přidali jsme 250 µl 70% ethanolu. Zopakovali jsme předchozí centrifugaci, ale tentokrát jsme centrifugovali 15 minut. Následně jsme odsáli ethanol, pelet vysušili a nakonec rozpustili v destilované vodě.

4.4.8. Sekvence DNA Sangerovou metodou

PCR produkty a plazmidová DNA byly sekvenovány v sekvenační laboratoři PŘF UK v Praze. Objem sekvenační reakce byl 8 μ l, obsahovala 0,4 μ l primeru (zhruba 3,2 pmol) a 5 – 10 ng PCR produktu na 100 bp jeho délky, nebo 3 – 5 ng plazmidové DNA na 100 bp délky plazmidu. Reakce byla doplněna vodou PCR kvality.

4.4.9. 5' Rapid amplification of cDNA ends – 5' RACE

Pro 5' RACE experimenty jsme použili kity FirstChoice RLM-RACE Kit (Ambion) a 5'/3' RACE Kit, 2nd Generation (Roche). Postupovali jsme podle příložených protokolů. Pro PCR amplifikaci jsme používali DNA polymerázu Ex Taq Hot Start version (TaKaRa). Použité primery jsou uvedeny v tabulce 1. „Forward“ primery byly použity pro kontrolu funkčnosti reverzních primerů pomocí klasické PCR s cDNA jako templátem.

Tabulka č. 1: Přehled použitých primerů pro 5' RACE

gen	primer	sekvence	poznámka
GAPDH (MONOS_5021)	GAP5030-F	CCATGGGGACCAATGGGAAT	forward
	GAP5030-Rout	AACACGGCATGAGTATCCCC	reverse
	GAP5030-Rin	TGGAAGTGAACACGGAGTG	reverse
enoláza (MONOS_3916)	ENO3919-F	AATGTTTCGTGCGGCTGTTC	forward
	ENO3919-Rin	TTGCTGTTCTCTCGCCTCTG	reverse
	ENO3919-Rout	TTCAGGGCACGTAACAAGCA	reverse
	ENO_seq	AAGCGGGGGCAATAACAGTG	sekvenačn
SufB (MONOS_1242)	SufB_f	ACTCGATCCTCTATCGCTTGC	forward
	SufB_out	GGAACGACCGAGGCAAGATA	reverse
	SufB_in	TTGCTTGTGTGGCTGGAGA	reverse
SufC (MONOS_1226)	SufC_f	ATCCTGGAGAAGTTCACGCC	forward
	SufC_out	TCACAACACGCAATGCATCA	reverse
	SufC_in	CAGGTATGCTTGGAGGCGAT	reverse
SufSU (MONOS_9203)	SufSU_f	GCATCGCTTCATCCACATCG	forward
	SufSU_out	GGGTGAGATGTGGTGCATGT	reverse
	SufSU_in	CCGTCTGAGAATCACTGGCT	reverse
Cpn60 (MONOS_13132)	Cpn60_f	TAAAGTCATCCTTCGGCCCAG	forward
	Cpn60_out	GCCTTGAGCAGCTCGTGTAT	reverse
	Cpn60_in	AGAGAGAAACGCAAAGCACG	reverse
H3 (MONOS_8194)	5J3_RACE_GS	ATGGTTCGCACAAAACAAGTTGCTCAC	forward
	5H3_RACE_O	GAGTTGCATATCCTTTGGCATAATTGT	reverse
	5H3_RACE_IN	GATTGTGTCTCGAATAATCCC	reverse
PFO (MONOS_4520)	PFO_121_f	TGTTGATGGATGTACCGCCG	forward
	PFO_121_out	CGAAGAAATGCAGGAACGGC	reverse
	PFO_121_in	AACCGTGAAGTGCACCAGAT	reverse

PFO (MONOS_12154)	PFO_653_f	ACTCCTGCATCCACAATGGG	forward
	PFO_653_out	ATTTGTGCACGAACGGTTGC	reverse
	PFO_653_in	TCCAACGGAACGAGATGCAA	reverse
Acetyl-CoA synthetáza (MONOS_10363)	AcCoASyn_f	GAATCAGTTGGTGTGGGCG	forward
	AcCoASyn_out	AACCTCATCAGCACCTGCAA	reverse
	AcCoASyn_in	CCCATTCAGCGAAACCTGC	reverse

4.5. Práce s bakteriemi

4.5.1. Použitá média

100% LB médium

- Složení
 - LB Broth (Sigma): 20 g
 - dH₂O: 1000 ml
- LB Broth jsme rozpustili ve vodě a sterilizovali autoklávováním (120 °C, 20 minut). Hotové médium jsme uchovávali ve 4 °C.
- Výsledné chemické složení média
 - trypton: 10 g/l
 - kvasničný autolyzát: 5 g/l
 - NaCl: 5 g/l

LB médium s agarem

- Do 1000 ml 100% LB média jsme přidali 15 g agaru a sterilizovali autoklávováním (120 °C, 20 minut). Po zchlazení média jsme ho sterilně rozlili do plastových Petriho misek (přibližně 20 ml na jednu misku) a nechali zatuhnout.

SOC médium

- Složení
 - trypton: 20 g
 - kvasničný autolyzát: 5 g
 - NaCl: 0,5 g
 - KCl: 0,186 g
 - MgCl₂: 0,952 g

- MgSO₄: 2,408 g
- glukóza: 3,603 g
- dH₂O: 1000 ml

4.5.2. Transformace bakterií

Přečištěný úsek DNA jsme zaligovali do plazmidu pJET 1.2 pomocí CloneJET PCR Cloning kitu (Thermo scientific), postupovali jsme podle přiloženého návodu. Tímto plazmidem jsme transformovali kompetentní buňky *Escherichia coli* kmene TOP10. Transformaci jsme provedli metodou teplotního šoku (Froger a Hall, 2007). Bakterie jsme následně kultivovali v 1 ml SOC média jednu až dvě hodiny na třepačce při 220 rpm a 37 °C. Narostlé buňky jsme poté převedli na LB agarové plotny s rozetřeným ampicilinem (100 µg/ml) a přes noc inkubovali při 37 °C. Poté jsme vybrali několik kolonií, na kterých jsme provedli Colony PCR. Po kontrole pomocí agarózové elektroforézy jsme pozitivní kolonie naočkovali do 3 ml 100% LB média s přidaným ampicilinem (100 µg/ml) a inkubovali přes noc v 15ml zkumavkách na třepačce při 220 rpm a 37 °C. Následně jsme z narostlých bakterií izolovali plazmidy.

4.5.3. Izolace plazmidů

Plazmidy jsme izolovali pomocí kitu High Pure Plasmid Isolation Kit (Roche), postupovali jsme podle protokolu dodaného výrobcem. Koncentraci DNA jsme změřili na spektrofotometru Nanodrop a izolované plazmidy jsme uchovávali v -20 °C.

4.6. Práce s proteiny

4.6.1. Izolace transkripčních faktorů

K izolaci transkripčních faktorů jsme používali µMACS FactorFinder Kit (Miltenyi Biotec). Centrifugovali jsme 5 minut 10⁷ buněk *Monocercomonoides* (přečištěných filtrací) nebo *Trichomonas* ve 4 °C při 3000 g. Odstranili jsme supernatant a poté jsme buňky dvakrát promyli v 1 ml PBS s 5mM EDTA, centrifugovali jsme stejným

způsobem. Do ledově vychlazeného Cell Lysis Bufferu (součást kitu) jsme přidali inhibitory proteáz (cOmplete, Mini, Roche) a inhibitory fosfatáz (PhosSTOP, Roche). Buňky jsme resuspendovali ve 100 μ l Cell Lysis Bufferu a přenesli do mikrozkuřavky. Následně jsme buňky umístili na dvě minuty do 37 °C a poté na 5 minut do -80 °C, nebo na několik sekund do kapalného dusíku. Toto střídání teplot jsme dvakrát opakovali. Lyzát buněk jsme centrifugovali 5 minut ve 4 °C při 15 000 g. Supernatant jsme převedli do nové mikrozkuřavky a uchovávali ho v -80 °C nebo přímo použili k izolaci transkripčních faktorů.

Pomocí PCR s 5' biotinylovanými primery (tabulka č. 2) jsme amplifikovali úsek DNA ležící v „upstream“ oblasti cílového genu o délce přibližně 130 – 180 bp. Předpokládáme, že tento úsek obsahuje místa pro vazbu transkripčních faktorů.

K lyzátu buněk jsme přidali 200 μ l Binding Bufferu (součást kitu), do kterého jsme přidali 2 μ l Binding Enhanceru (součást kitu) a inhibitory proteáz (cOmplete, Mini, Roche) a fosfatáz (PhosSTOP, Roche). Ke směsi jsme přidali 5 μ g přečištěného biotinylovaného úseku DNA (přibližně 50 pmol) a inkubovali jsme reakci 20 minut při laboratorní teplotě. Poté jsme přidali 100 μ l magnetických kuliček se streptavidinem, směs krátce promíchali a nechali dalších 10 minut stát. Mezitím jsme kolonu (μ Column, Miltenyi Biotec) umístili do magnetického stojánku a nanесли na kolonu 100 μ l Binding Bufferu. Celou reakci jsme napipetovali na kolonu a nechali protéct. Kolonu jsme poté čtyřikrát promyli Wash Bufferem LS a čtyřikrát Wash Bufferem HS (oba součást kitu). Poté jsme na kolonu nanесли 30 μ l Native Elution Bufferu (součást kitu). Pokud jsme vzorek analyzovali pomocí SDS PAGE, byl místo tohoto pufru použit vzorkový I pufr (složení viz kap. 4.6.2) zahřátý na 95 °C. Po dvou minutách jsme na kolonu opět nanесли 50 μ l Native Elution Bufferu, nebo 50 μ l I pufru a teprve tato druhá eluce byla použita pro další analýzu.

Alternativně jsme pro získání většího množství transkripčních faktorů k experimentu použili $2,5 \times 10^7$ buněk. V tomto případě jsme adekvátně zvětšili objem PBS pro promývání buněk, Cell Lysis Bufferu a Binding Bufferu. Množství DNA a magnetických kuliček zůstalo stejné.

Tabulka č. 2: Přehled použitých primerů pro amplifikaci „upstream“ oblastí genů použitých při izolaci transkripčních faktorů

<i>Monocercomonoides PA203</i>		
Gen	Primer	Sekvence
GAPDH (MONOS_5021)	gap5021F-biot	(biot) TGTCTAAGCCTTATGATCTGATCTACA
	gap5021R-biot	(biot) AGTTGCGTTGTAACCTCAAACCTTTATTT
enoláza (MONOS_3916)	enolase3916F-biot	(biot) TGCAAAACAATAAGATCATAACCCA
	enolase3916R-biot	(biot) AGGCTGAATCGAAAGAAAAATTCA
H3 (MONOS_8194)	MONO_H3_biot_F	(biot) AATTGTGAACATTTTCAAAGCGAA
	MONO_H3_biot_R	(biot) GTGAGCAACTTGTTTTGTGCG
	MONO_inH3_biot_F	(biot) GCCTCGAAAGCAATTGGCA
	MONO_inH3_biot_R	(biot) TGATTTCTCGCAACGCAACA
<i>Trichomonas vaginalis G3</i>		
Gen	Primer	Sekvence
GAPDH (TVAG_382880)	Tv_GAP_biot_F	(biot) TCACCGGATTACAATTCAGAACA
	Tv_GAP_biot_R	(biot) GGAGCCGATCATGCATACCT

Pomocí primerů MONO_inH3_biot_F a MONO_inH3_biot_R byl amplifikován úsek DNA uvnitř genu pro histon H3, který byl použit jako negativní kontrola pro oba organismy.

4.6.2. Polyakrylamidová denaturační proteinová elektroforéza (SDS PAGE)

- Chemikálie

Roztok A

- 30% Acrylamide/Bis solution 29:1 (Bio-Rad)

Roztok B

- SDS (sodium dodecyl sulfát): 1 g
- doplněno do 10 ml dH₂O

Roztok C

- Tris (Trizma base): 9,1 g
- SDS, pH=8,8: 0,2 g
- doplněno do 100 ml dH₂O

Roztok D

- Tris, pH=6,8: 3 g
- SDS: 0,2 g
- doplněno do 100ml dH₂O

Roztok E

- 1x TGS (Bio-Rad)

Roztok F:

- 10 mg bromfenolové modři na 1ml roztoku D bez SDS

Roztok G

- persíran amonný: 1 g
- dH₂O: 10 ml

Barvicí roztok CBB (Coomasie brilliant blue)

- CBB: 250 mg
- ethanol: 225 ml
- dH₂O: 225 ml
- kyselina octová: 50 ml

Odbarvovací roztok na CBB

- ethanol: 250 ml
- dH₂O: 650 ml
- kyselina octová: 100 ml

Vzorkový pufr I

5x I

- roztok D (bez SDS): 2,4 ml
- roztok C: 2 ml
- glycerol: 2,5 ml
- dH₂O: 1,6 ml
- roztok F: 1 ml
- 2-merkapt ethanol: 0,5 ml

1x I

- roztok D (bez SDS): 2,4 ml
- roztok C: 2 ml
- glycerol: 1 ml
- dH₂O: 4 ml

- roztok F: 0,1 ml
- 2-merkaptoethanol: 0,5 ml

TEMED (Sigma)

Velikostní standardy

- PageRuler Plus prestained protein ladder (Thermo scientific)
- PageRuler Prestained Protein Ladder (Thermo scientific)
- *Tabulka č. 3: Složení používaných gelů*

	10%	12%	Stack 5%
A (ml)	1,36	1,6	0,24
C (ml)	2	2	D (ml): 0,75
H ₂ O (ml)	0,64	0,32	0,47
TEMED (μl)	5	5	5
G (μl)	40	40	20

- Pro SDS PAGE jsme používali elektroforézu Mini-protean 3 (Bio-Rad) s gely o rozměrech 7 x 8 cm, spacer 0,75 mm. Gely jsme nalévali podle předchozí tabulky, nebo jsme používali připravené gradientové gely (8–16% Mini-PROTEAN® TGX™ Precast Protein Gels, Bio-Rad). Na gely jsme zpočátku aplikovali napětí 70 V, po průchodu vzorků „stacking“ gelem jsme napětí zvýšili na 120 V. Pro zviditelnění proteinů jsme barvili running gel přes noc v roztoku CBB. Následně jsme gel promývali v odbarvovacím roztoku až do zvýraznění jednotlivých proteinů.

4.6.3. Hmotnostní spektroskopie

Proteinové vzorky (eluáty z kolony a vyřezané pruhy z SDS PAGE) získané podle postupu z kap. 4.6.1. byly analyzovány tandemovou hmotnostní spektrometrií (MS/MS) v laboratoři hmotnostní spektrometrie PŘF UK v Praze.

4.7. Bioinformatické metody

4.7.1. Základní práce se sekvenčními daty

K zobrazení a úpravě dat ze Sangerova sekvenování jsme používali programy Geneious v9.0.5 (Kearse et al., 2012) a BioEdit v7.2.5 (Hall, 1999). Pro navrhování primerů jsme využívali zejména nástroj Primer-BLAST (Ye et al., 2012). K odhalování sekvenční podobnosti jsme používali algoritmus BLAST v několika variantách (Altschul et al., 1990), alignment více sekvencí jsme vytvářeli pomocí algoritmu ClustalW (Thompson et al., 1994), který je součástí programů Geneious i BioEdit.

4.7.2. Charakterizace nepřekládaných oblastí

Pro práci s genomovými a transkriptomickými daty byl použit program IGV v2.3 (Robinson et al., 2011). Predikce 5' UTR byla provedena pomocí programového balíku PASA (Haas et al., 2008), tento balík jsme dále použili pro automatickou extrakci konkrétních částí genů z celého genomu. Sekvence 5' UTR z ostatních druhů organismů, které jsme použili pro porovnání s našimi daty, jsou extrahované z databáze UTRdb (Grillo et al., 2010). K charakterizaci nepřekládaných oblastí jsem vytvořil skripty napsané v programovacím jazyce Python v2.7 a v3.5 (www.python.org).

4.7.3. Anotace proteinů

Automatická anotace základních transkripčních a translačních iniciačních faktorů byla provedena pomocí databáze KEGG (Kanehisa a Goto, 2000) ze sady předpovězerných proteinů z genomu *Monocercomonoides*. Pro tyto proteiny jsme následně s využitím transkriptomických dat a databáze GenBank (Benson et al., 2012) ověřovali správnost genových modelů a případně je upravovali. V případě nízké sekvenční podobnosti jsme k anotaci proteinů využívali i databázi Pfam (Finn et al., 2008).

4.7.4. Vyhledávání motivů v promotorových oblastech

Pro analýzu konzervovaných motivů v promotorových oblastech jsme použili nástroje v balíku MEME Suite. Vyhledávání nových motivů bylo provedeno pomocí algoritmů MEME (Bailey et al., 2009) a DREME (Bailey, 2011). Pozici konzervovaných motivů jsme analyzovali pomocí programu Centrimo (Bailey a Machanick, 2012), podobnost motivů pomocí Tomtom (Gupta et al., 2007).

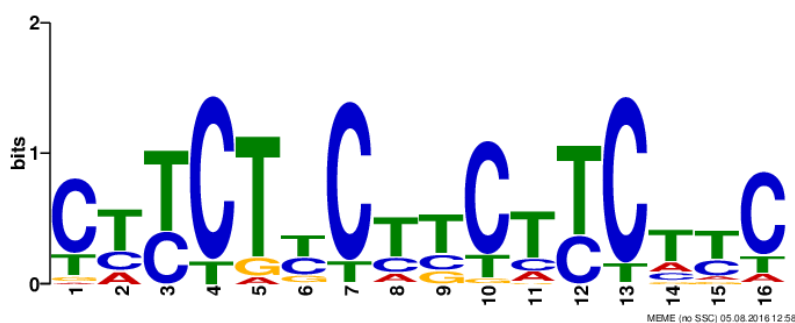
5. Výsledky

5.1. Promotorové oblasti

Konzervované motivy jsme hledali v oblastech 200 nukleotidů před predikovaným začátkem transkripce a 50 nukleotidů za ním. Nalezené motivy byly stejné pro dataset všech genů s predikovanými 5' UTR (6983 genů) i pro dataset genů s anotovanou funkcí (3675 genů). Lišila se pouze statistická signifikance, která byla vyšší u datasetu všech genů a proto se budu dále zabývat jen výsledky z tohoto datasetu.

- Algoritmus MEME

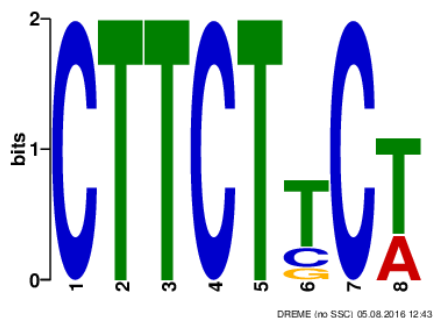
Nalezli jsme pouze jeden statisticky signifikantní motiv (motiv č. 1) s e-value $3,3e-138$.



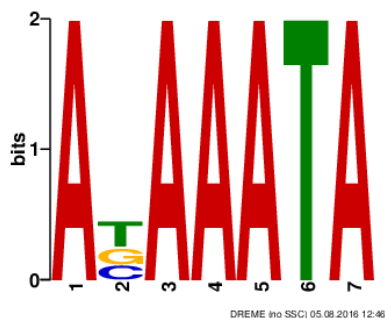
Motiv č. 1, e-value $3,3e-138$

- Algoritmus DREME

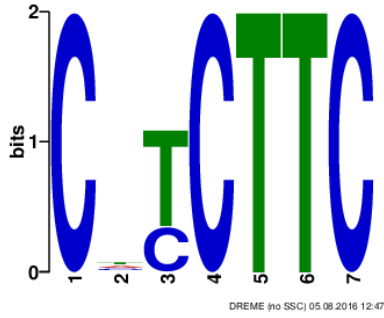
Pomocí tohoto algoritmu jsme našli poměrně velké množství motivů, uvedeno je 10 motivů s e-value do $10e-20$ (motivы 2 – 11).



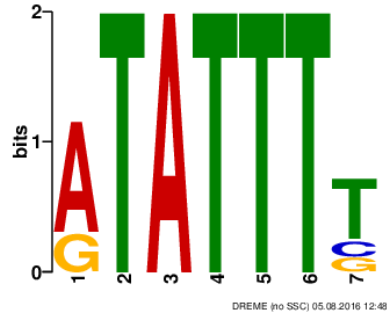
Motiv č. 2: CTTCT(T/C/G)C(T/A),
e-value $4,9e-157$



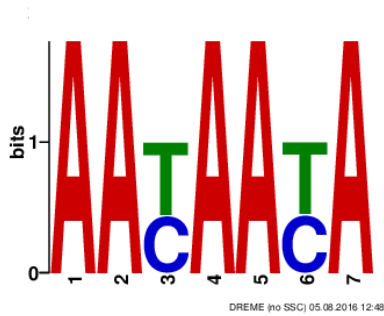
Motiv č. 3: A(T/C/G)AAATA,
e-value $5,6e-92$



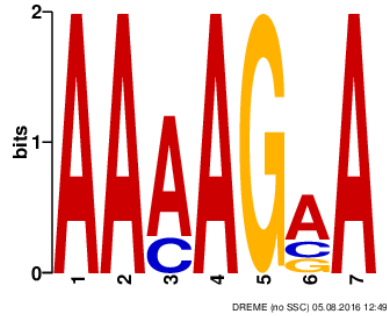
Motiv č. 4: CN(T/C)CTTC,
e-value 4,2e-67



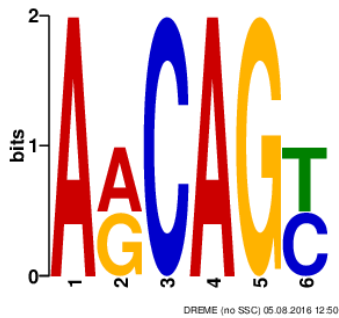
Motiv č. 5: (A/G)TATTT(T/C/G),
e-value 1,8e-64



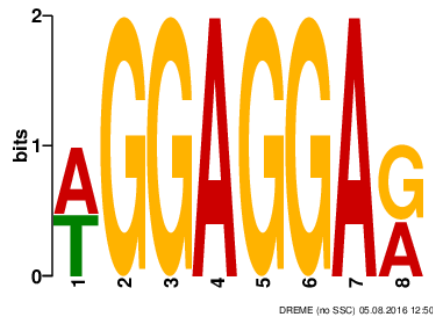
Motiv č. 6: AA(T/C)AA(T/C)A;
e-value 2,1e-54



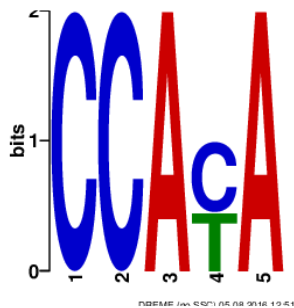
Motiv č. 7: AA(A/C)AG(A/C/G)A;
e-value 1,2e-041



Motiv č. 8: A(A/G)CAG(T/C);
e-value 1,4e-31



Motiv č. 9: (A/T)GGAGGA(A/T);
e-value 1,4e-28



Motiv č. 10: CCA(C/T)A;
e-value 1,5e-26

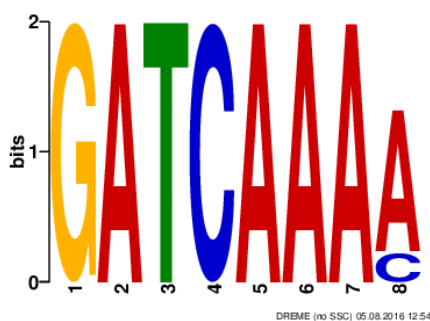


Motiv č. 11: (A/G)AA(A/G)ATA;
e-value 3,2e-021

- Centrimo

Pomocí tohoto algoritmu jsme zjišťovali, zda se jednotlivé motivy nacházejí v promotorových oblastech v určité konzervované pozici. Použili jsme naše výsledky z MEME a DREME a motivy z databází JASPAR a TFBSshape.

Pro motivy č. 1 – 4, č. 6 – 9 a motiv č. 11 nebyla zjištěna konzervovaná pozice. Motiv č. 10 CCA(C/T)A nalezený algoritmem DREME se nejčastěji nachází 15 nukleotidů před začátkem transkripce (graf č. 1, vyznačen modře). V grafech 1 až 3 jsou vyznačeny pozice motivů v 250 nukleotidů dlouhých sekvencích, ve kterých jsme hledali konzervované motivy. Predikovaný začátek transkripce se v těchto grafech nachází na hodnotě 75 na ose x (označeno zelenou přerušovanou čarou). Motiv GATCAAA(A/C) byl identifikován také pomocí DREME, ale s e-value pouze $4,7e-4$ (motiv č. 12). Byl nalezen jen v 649 případech. Z důvodu nízké e-value z DREME jsme se tímto motivem začali zabývat až poté, co jsme pro něj pomocí Centrimo našli konzervované pozice v sekvencích s velkou statistickou signifikancí. Tento motiv je v sekvencích lokalizován okolo predikovaného místa začátku transkripce, nebo 25 nukleotidů před ním (graf č. 1, vyznačen červeně).



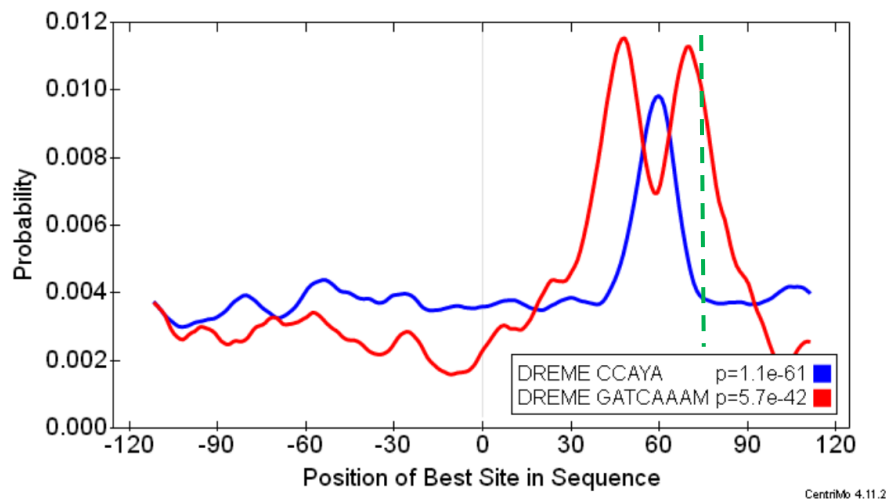
Motiv č. 12: GATCAAA(A/C),
e-value $4,7e-4$

Nepříliš průkazné je umístění motivu č. 5, který připomíná TATA box. Nejčastěji se vyskytuje přibližně 45 nukleotidů před začátkem transkripce (graf č. 2), nalezen byl přibližně u 52 % sekvencí.

Při hledání motivů zahrnutých v databázích JASPAR DNA a TFBSshape DNA jsme se zaměřili na TATA box. V tomto případě jsme nehledali motivy *de novo*, ale zkoumali jsme obohacení známých motivů v sekvencích *Monocercomonoides*. Nejčastěji byl konsenzuální TATA box nalezen 15 nukleotidů před startem transkripce, celkově se nachází u 44 % sekvencí, ale s e-value pouze $9,9e-10$ (graf č. 3).

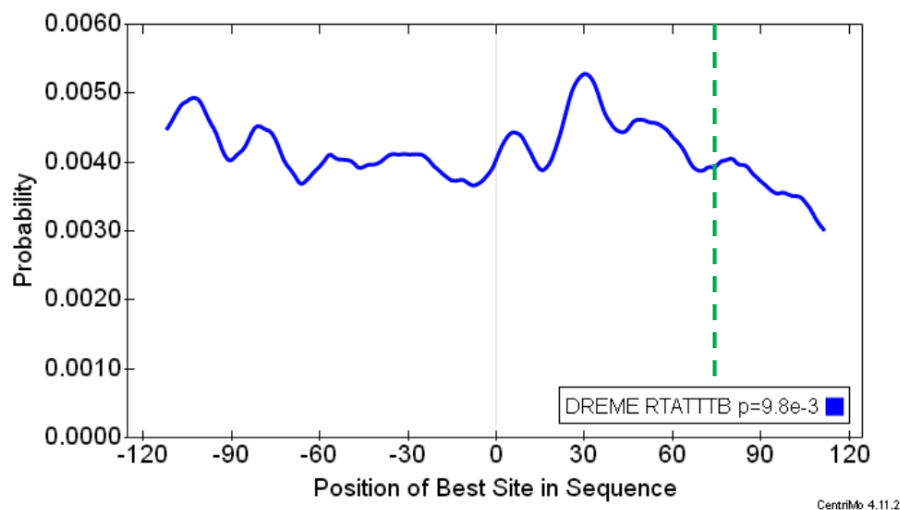
Graf č. 1: Pozice motivů č. 10 a 12 nalezených DREME. Zelenou přerušovanou čarou je vyznačen předpokládaný začátek transkripce.

Motif Probability Graph (score ≥ 5 bits) [?](#)

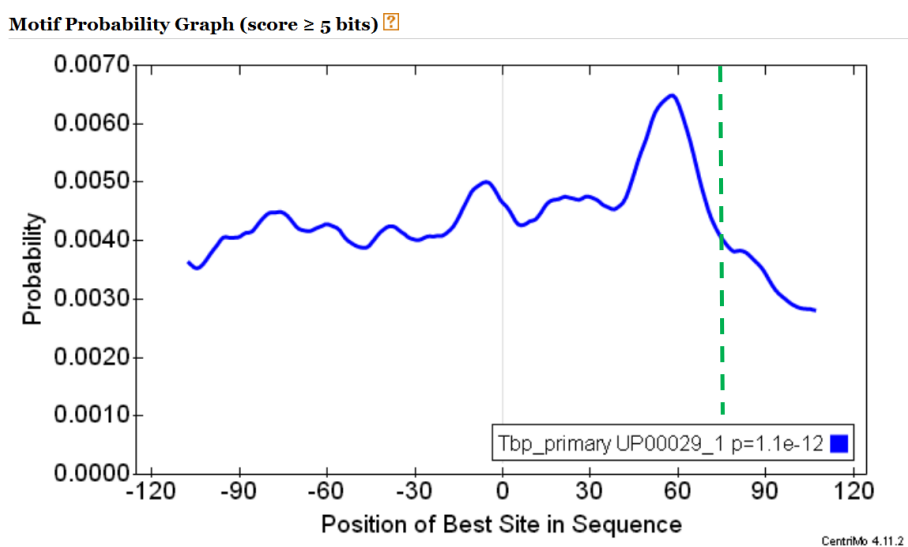


Graf č. 2: Pozice TATA-like motivu č. 5 nalezeného DREME. Zelenou přerušovanou čarou je vyznačen předpokládaný začátek transkripce.

Motif Probability Graph (score ≥ 5 bits) [?](#)



Graf č. 3: Pozice konsenzuální sekvence TATA boxu. Zelenou přerušovanou čarou je vyznačen předpokládaný začátek transkripce.



5.2. Bioinformatická analýza UTRs

Pro porovnání průměrné délky a obsahu GC párů 5' nepřekládaných oblastí s ostatními oblastmi genomu *Monocercomonoides* jsme použili dataset všech předpovězených genů s predikovanou 5' UTR (6983 genů) (tabulka č. 4).

Tabulka č. 4: Porovnání 5' nepřekládaných oblastí s jinými oblastmi genomu *Monocercomonoides*

	průměrná délka	GC	počet sekvencí
3' UTR	167	0,33	6907
5' UTR	109	0,36	6983
introny	125	0,25	31478
cDNA	2480	0,43	16769
kódující oblasti	2366	0,43	16769
geny	2727	0,41	16608
mezigenové oblasti	869	0,29	8765

Spočítali jsme průměrnou délku a průměrný obsah GC párů v 5' nepřekládaných oblastech v genomu *Monocercomonoides* a dalších vybraných prvoků (tabulka č. 5). V případě *Monocercomonoides* jsme v tomto případě vycházeli ze dvou sad genů. Průměrná délka 65 nukleotidů vychází z datasetu všech genů, u kterých byla

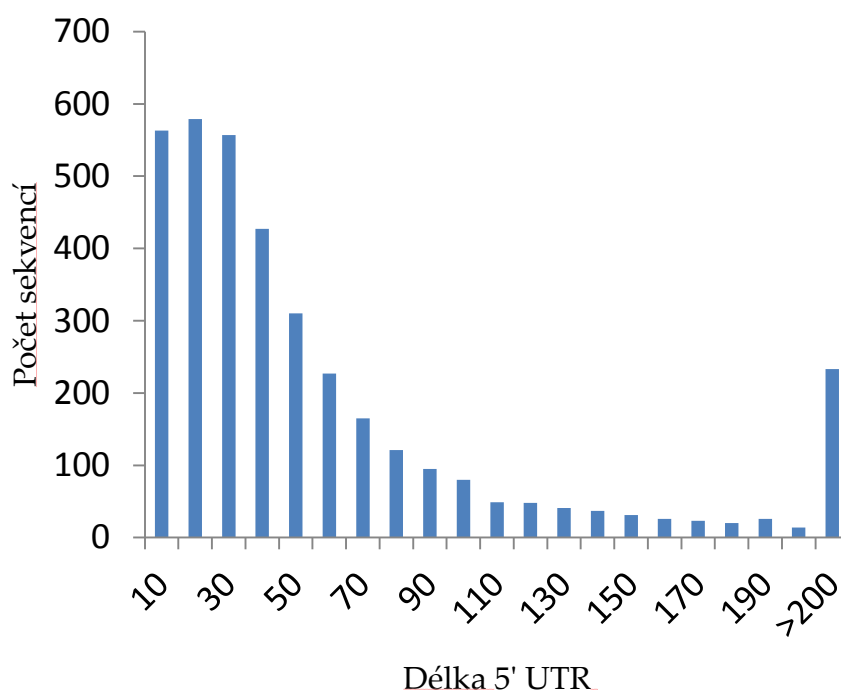
automaticky anotovaná funkce a které mají zároveň predikovanou i 5' UTR (3672 genů). Délka 46 nukleotidů je pro dataset, který obsahuje pouze geny s manuálně anotovanou funkcí a s predikovanou 5' UTR (718 genů). Předpokládáme, že použitím funkčně anotovaných genů, a zejména pouze manuálně anotovaných genů, jsme odstranili špatné genové predikce nebo pseudogeny a výsledky jsou přesnější.

Tabulka č. 5: Charakteristika 5' nepřekládaných oblastí vybraných prvoků

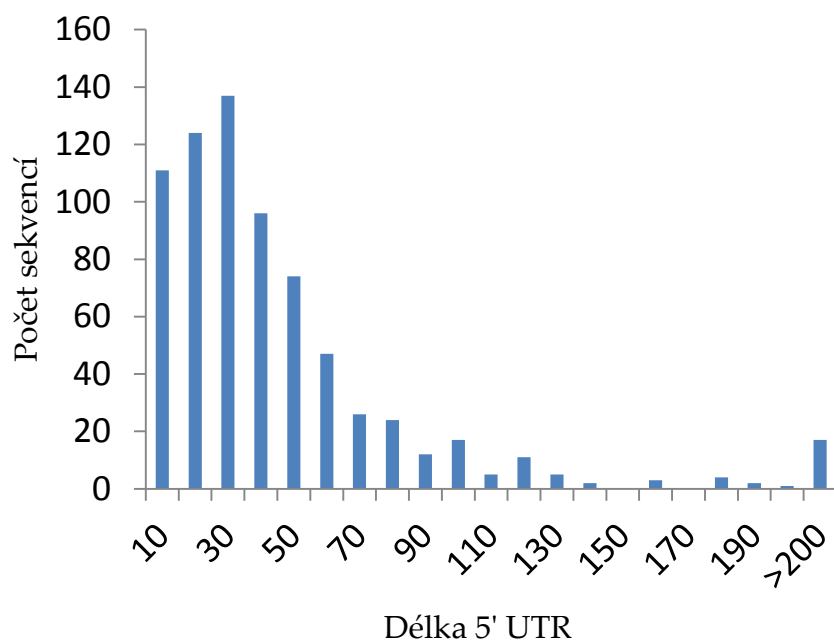
	průměrná délka	GC	počet genů
<i>Monocercomonoides</i>	65 (46)	0,33 (0,33)	3675 (718)
<i>Babesia bovis</i>	123	0,38	861
<i>Cryptosporidium muris</i>	135	0,28	56
<i>Entamoeba histolytica</i>	53	0,24	1115
<i>Trichomonas vaginalis</i>	82	0,31	2546
<i>Tetrahymena thermophila</i>	137	0,23	764
Obratlovci	141	0,50	61390

Vytvořili jsme histogramy znázorňující distribuci délek 5' nepřekládaných oblastí u vybraných prvoků (graf č. 4 – 10).

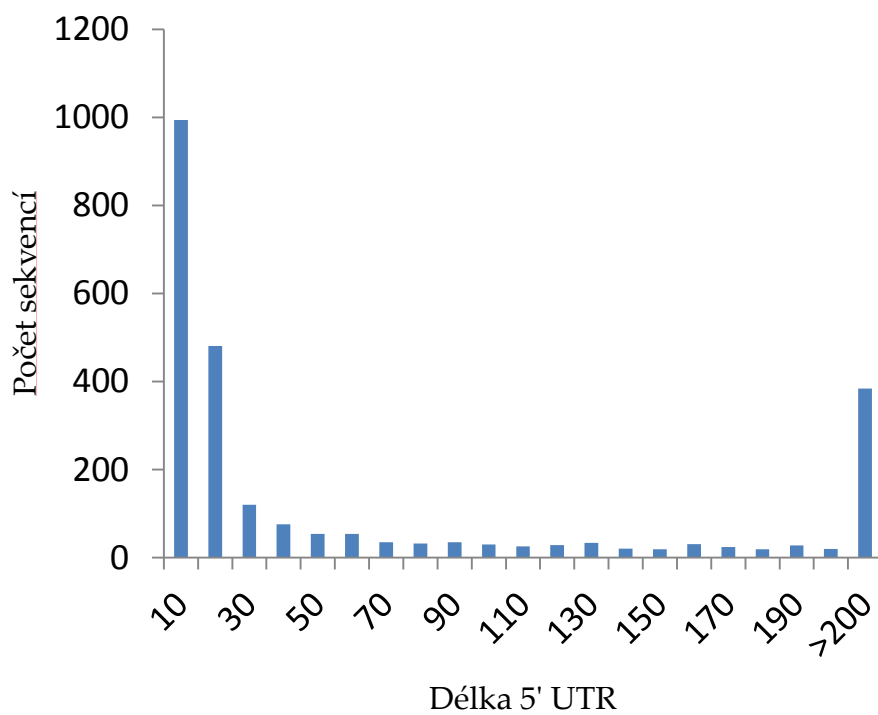
Graf č. 4: Délková distribuce 5'UTR: *Monocercomonoides*, automaticky anotované geny



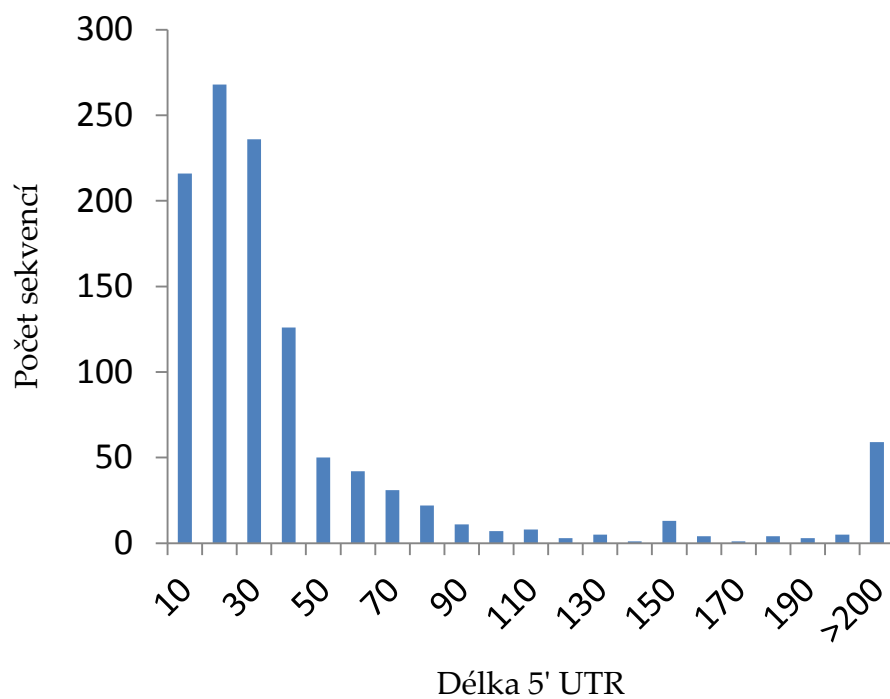
Graf č. 5: Délková distribuce 5'UTR: *Monocercomonoides*, manuálně anotované geny



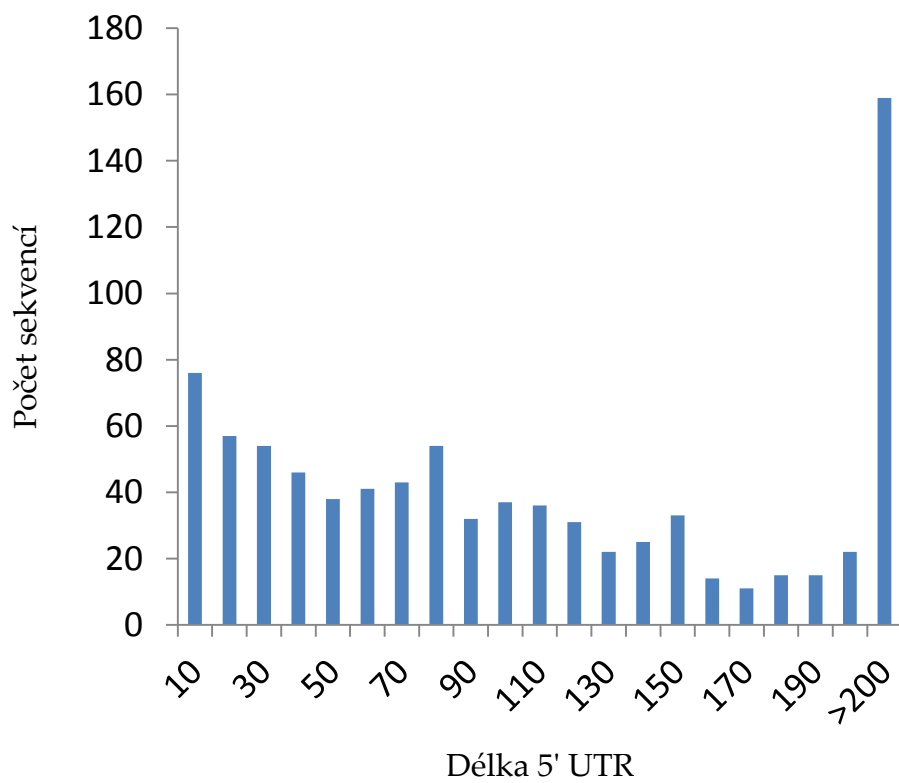
Graf č. 6: Délková distribuce 5'UTR: *Trichomonas vaginalis*



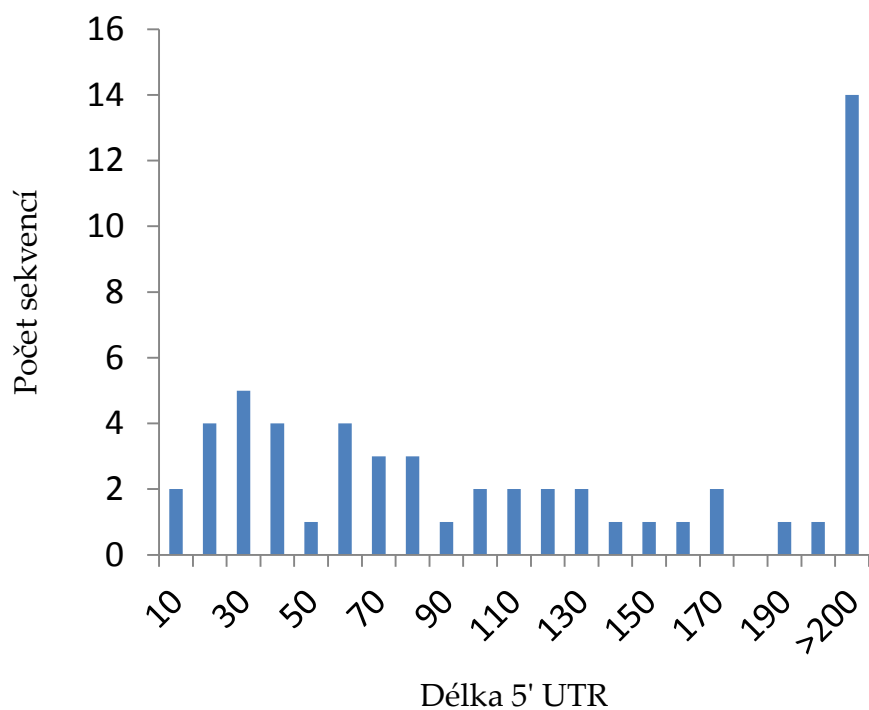
Graf č. 7: Délková distribuce 5'UTR: Entamoeba histolytica



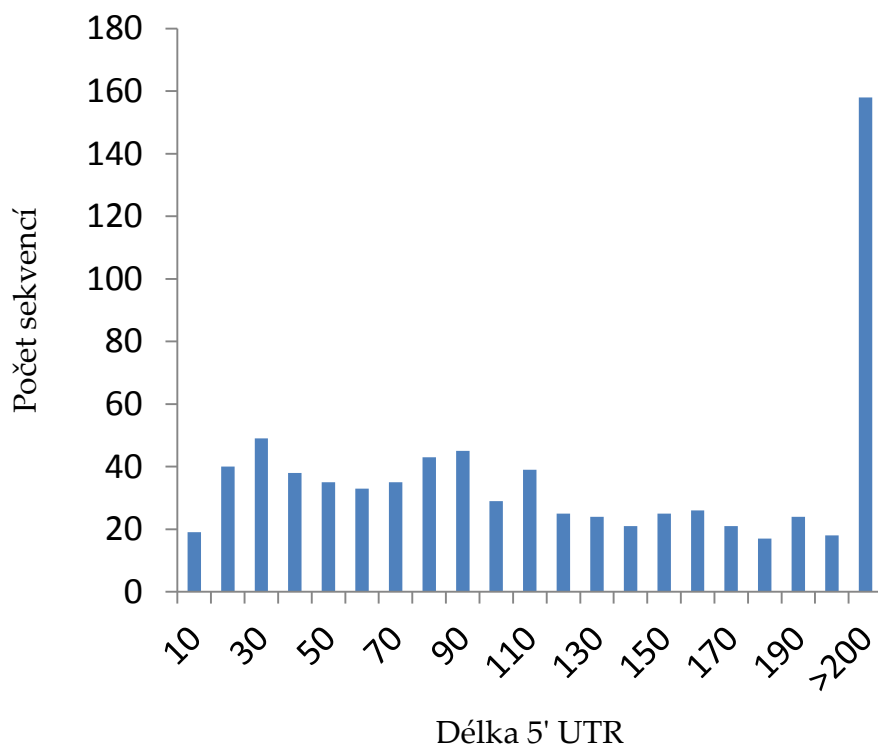
Graf č. 8: Délková distribuce 5'UTR: Babesia bovis



Graf č. 9: Délková distribuce 5'UTR: *Cryptosporidium muris*



Graf č. 10: Délková distribuce 5'UTR: *Tetrahymena thermophila*



5.3. 5' RACE

Pomocí kitu FirstChoice RLM-RACE Kit (Ambion) se nám nepodařilo amplifikovat žádnou 5' nepřekládanou oblast. Kitem 5'/3' RACE Kit, 2nd Generation (Roche) jsme amplifikovali 5' UTR transkriptů pro GAPDH, enolázu, Cpn 60, histon H3 a Acetyl-CoA syntetázu. Nepodařilo se amplifikovat sekvence pro dvě kopie PFO, SufB, SufC a SufSU. Získané sekvence jsme porovnali s automatickou predikcí (Alignment 1 – 5).

Alignment č. 1: GAPDH (MONOS_5021)

	10	20	30	40	50	60
kódující sekvence			ATGTC	AATTAAAGTTGCGATTAA	CGGATTTGGC	CAGAAT
predikovaná 5' UTR		GATCATAAGGCTTAGACATG	TCAATTAAAGTTGCGATTAA	CGGATTTGGC	CAGAAT	
5' RACE	TAGATCAGATCATAAGGCTTAGACATG	TCAATTAAAGTTGCGATTAA	CGGATTTGGC	CAGAAT		
genomová sekvence	TTGTAGATCAGATCATAAGGCTTAGACATG	TCAATTAAAGTTGCGATTAA	CGGATTTGGC	CAGAAT		

Alignment č. 2: enoláza (MONOS_3916)

	10	20	30	40	50	60
kódující sekvence			TTTGT	TCCAACTTGCAAAAAT	TGAATTTTTC	TTTCGATTCAGCCTTTAT
predikovaná 5' UTR			TTTGT	TCCAACTTGCAAAAAT	TGAATTTTTC	TTTCGATTCAGCCTTTAT
5' RACE	TAAAAGCTTTCTGAATTTTGT	TCCAACTTGCAAAAAT	TGAATTTTTC	TTTCGATTCAGCCTTTAT		
genomická sekvence	TAAAAGCTTTCTGAATTTTGT	TCCAACTTGCAAAAAT	TGAATTTTTC	TTTCGATTCAGCCTTTAT		

	70	80	90	100	110	120	13
kódující sekvence	AAAAATTT	CAAATAGCAAGAA	TTTTTTT	TGCGAATGTTG	AAAAGAAGTTG	TTTACAATTTATATA	
predikovaná 5' UTR	AAAAATTT	CAAATAGCAAGAA	TTTTTTT	TGCGAATGTTG	AAAAGAAGTTG	TTTACAATTTATATA	TATA
5' RACE	AAAAATTT	CAAATAGCAAGAA	TTTTTTT	TGCGAATGTTG	AAAAGAAGTTG	TTTACAATTTATATA	
genomická sekvence	AAAAATTT	CAAATAGCAAGAA	TTTTTTT	TGCGAATGTTG	AAAAGAAGTTG	TTTACAATTTATATA	

	140	150	160	170	180	190
kódující sekvence	TTACTGAATAAC					
predikovaná 5' UTR	TTACTGAATAAC					
5' RACE	TTACTGAATAAC					
genomická sekvence	TTACTGAATAACG	TATGCAGTTT	ATTGGTTTCTTTA	ATTTTAAAATTT	TGTGGCGTTATTC	CAAAAC

	200	210	220	230	240	250	26
kódující sekvence			ATGTC	AAACAATCAGAAAGAT	TTTGCACGCGAGATA	CTCGATTCACGTGC	
predikovaná 5' UTR		ACAACGCAATG	TCAACAATCAGAAAGAT	TTTGCACGCGAGATA	CTCGATTCACGTGC		
5' RACE		ACAACGCAATG	TCAACAATCAGAAAGAT	TTTGCACGCGAGATA	CTCGATTCACGTGC		
genomická sekvence	AA	TCCAGACAACGCAATG	TCAACAATCAGAAAGAT	TTTGCACGCGAGATA	CTCGATTCACGTGC		

Alignment č. 3: Cpn60 (MONOS_13132)

	10	20	30	40	50	60
kódující sekvence			ATG	TTC	CAT	CCAAAAAGTCGTTTGAAATACGTC
predikovaná 5' UTR		GAA	TAAAAATG	TTC	CAT	CCAAAAAGTCGTTTGAAATACGTC
5' RACE		ATAAGGAA	TAAAAATG	TTC	CAT	CCAAAAAGTCGTTTGAAATACGTC
genomová sekvence	AAA	TAAGGAA	TAAAAATG	TTC	CAT	CCAAAAAGTCGTTTGAAATACGTC

Alignment č. 4: H3 (MONOS_8194)

	10	20	30	40	50	60
kódující sekvence					ATGG	TCGCACAAAACAAGT
predikovaná 5' UTR		ATT	CCTCAATAAA	TTG	CTTCATTTCTT	GATGGTCGCACAAAACAAGT
5' RACE		ATCAAT	ATTTCC	CAATAAA	TTGCTTCATTTCTT	GATGGTCGCACAAAACAAGT
genomová sekvence	TATAT	ATTGATCAAT	ATTTCC	CAATAAA	TTGCTTCATTTCTT	GATGGTCGCACAAAACAAGT

Alignment č. 5: Acetyl-CoA synthetáza (MONOS_10363)

	10	20	30	40	50	60
kódující sekvence						
predikovaná 5' UTR		CGG	TATTTTGG	TTTCG	TTGGAAACAAAAAC	
5' RACE			GTT	TCGTTGGAAACAAAAAC		
genomová sekvence	ATT	GCGGTATTTTGG	TTTCGTTGGAAACAAAAAC	CGTTTGAAAA	TTCTTTGTTAAAA	TTGGAAG

	70	80	90	100	110	120	13
kódující sekvence							
predikovaná 5' UTR							
5' RACE							
genomová sekvence	TATTTTCTTTT	GATTTATA	TACTTTATTT	ATTTATTC	ATTAATTG	CTTAAATG	CGAAAAATATI

	140	150	160	170	180	190	
kódující sekvence							
predikovaná 5' UTR							
5' RACE							
genomová sekvence				ATG	TCTGATCTTC	ATATTC	TTTTCAATCCAAAGACTGI

5.4. Anotace základních transkripčních faktorů

Zkontrolovali a upravili jsme 23 genových modelů v genomu *Monocercomonoides* automaticky anotovaných s pomocí databáze KEGG. Přítomnost transkripčních faktorů jsme vyznačili zeleně v tabulce č. 5, kde uvádíme rovněž srovnání s *Giardia intestinalis* a *Trichomonas vaginalis*. Žlutě je označen protein, který byl automaticky anotován, ale pomocí jiných databází (NCBI, Pfam, Interproscan) jsme příslušnému genovému modelu tuto anotaci nedokázali potvrdit ani vyvrátit. Pokud se v genomu *Monocercomonoides* nachází více kopií jednoho genu, je to v tabulce uvedeno.

Tabulka č. 5: Anotované základní transkripční faktory

	<i>Monocercomonoides</i>	<i>G. intestinalis</i>	<i>T. vaginalis</i>
<i>TFIIA</i>			
TFIIA1			
TFIIA2			
<i>TFIIB</i>			
TFIIB			
<i>TFIID</i>			
TBP	(2 kopie)		
TAF1			
TAF2			
TAF3			
TAF4			
TAF5			
TAF6			
TAF7			
TAF8			
TAF9			
TAF10			
TAF11			
TAF12			
TAF13			
TAF14			
TAF15			
<i>TFIIE</i>			
TFIIE1			
TFIIE2			
<i>TFII-I</i>			
<i>TFIIF</i>			
TFIIF1			
TFIIF2			
TAF14			
<i>TFIIH</i>			
TFIIH1			
TFIIH2			
TFIIH3			
TFIIH4			
XPB			
XPD			
TTDA			
CDK7			
MAT1			
CCNH			

5.5. Anotace iniciačních translačních faktorů

Zkontrolovali a upravili jsme 30 genových modelů v genomu *Monocercomonoides* automaticky anotovaných s pomocí databáze KEGG. Přítomnost translačních faktorů jsme vyznačili zeleně v tabulce č. 6, kde uvádíme rovněž srovnání s *Giardia intestinalis* a *Trichomonas vaginalis*. Žlutě jsou označeny proteiny, které byly automaticky anotovány, ale pomocí jiných databází (NCBI, Pfam, Interproscan) jsme příslušným genovým modelům tuto anotaci nedokázali potvrdit ani vyvrátit. Pokud se v genomu *Monocercomonoides* nachází více kopií jednoho genu, je to v tabulce uvedeno.

Tabulka č. 6: Anotované eukaryotické iniciační translační faktory

	<i>Monocercomonoides</i>	<i>G. intestinalis</i>	<i>T. vaginalis</i>
<i>eIF1</i>			
eIF1A			
eIF1			
<i>eIF2</i>			
eIF2S1			
eIF2S2			
eIF2S3			
<i>eIF2A</i>			
eIF2A			
<i>eIF2B</i>			
eIF2B1			
eIF2B2			
eIF2B3	(2 kopie)		
eIF2B4			
eIF2B5	(2 kopie)		
<i>eIF2D</i>			
eIF2D			
<i>eIF5</i>			
eIF5			
eIF5A			
eIF5B			
<i>eIF6</i>			
eIF6			
<i>eIF3</i>			
eIF3A			
eIF3B			
eIF3C			
eIF3D			
eIF3E			

eIF3F			
eIF3G			
eIF3H			
eIF3I			
eIF3J			
eIF3K			
eIF3L			
eIF3M			
<i>eIF4</i>			
eIF4A			
eIF4A3			
eIF4B			
eIF4E	(6 kopií)		
eIF4G	(3 kopie)		
CAF20			
<i>MTIFs</i>			
infB			
infC			
<i>Asociované faktory</i>			
ABCF1			
ABCF3			
TIF31			
EAP1			
DHX29			
IGHMBP2			

5.6. Purifikace transkripčních faktorů

Proteinové vzorky pro izolaci transkripčních faktorů jsme v prvních pokusech separovali pomocí SDS PAGE a následně bandy podrobovali proteomické analýze. Především kvůli malému množství proteinů ve vzorcích jsme touto metodou nezískali žádná použitelná data. V dalších pokusech jsme analyzovali proteinové složení kompletních eluátů.

V pilotních experimentech s použitím „upstream“ oblastí genů pro GAPDH, enolázu a histon H3 se jeví nejnadějněji data získaná z pokusu na GAPDH. Kvůli přítomnosti velkého množství ribozomálních proteinů v eluátu jsme snížili vstupní množství buněk. Konečný experiment s „upstream“ oblastí genu pro GAPDH (schéma č. 1) jsme provedli v triplicátu s jednou negativní kontrolou. Identifikované proteiny, které byly ve vzorcích obohaceny vůči kontrole minimálně 2krát, nebo se vyskytovaly s dostatečnou intenzitou pouze ve vzorku, jsou uvedeny v tabulce č. 7.

Hodnoty představují dvojkový logaritmus intenzit vzorku, kontroly a jejich rozdíl. Intenzity ve vzorku jsou vypočteny jako medián ze tří měření. Pokud byl protein nalezen pouze ve dvou měřeních, jedná se o medián dvou hodnot. V případě, že byl protein identifikován pouze v jednom z měření, není v tabulce uveden. Stejný experiment jsme provedli i u *Trichomonas vaginalis*, také s pomocí „upstream“ oblasti genu pro GAPDH. Výsledky jsou uvedeny v tabulce č. 8.

Schéma č. 1: Amplifikovaná „upstream“ oblast genu pro GAPDH. Polohy primerů jsou vyznačeny podtržením, červeně je označen start kodon, zeleně automaticky predikovaný začátek transkripce a modře začátek transkripce určený pomocí RACE.

GAAGTTGCGTTGTAACTCAAACTTTATTTTATTTTGGAAAATTTTCTTGAAAC
AAATCAAATATTGATCAATAATATGTATTTTTCATTCTTCATCACATGAAAAT
GGTTGTAGATCAGATCATAAGGCTTAGACATGTCAATTAA

Tabulka č. 7: Výsledky hmotnostní spektrometrie eluátu *Monocercomonoides* pro „upstream“ oblast GAPDH. Hodnoty naměřených intenzit ve vzorku a negativní kontrole jsou uvedeny ve dvojkových logaritmech. Seřazeno podle velikosti rozdílu těchto hodnot.

gen	rozdíl	vzorek	kontrola	anotace
MONOS_6461	4,8538	23,9461	19,0923	Leucine Rich repeats (2 kopie)
MONOS_3475	4,7953	23,2168	18,4215	SAR DNA binding protein
MONOS_2785	3,8407	22,4755	18,6348	
MONOS_11085	3,7214	23,6535	19,9321	Nrap protein
MONOS_8660; 8659	3,3646	27,0208	23,6562	
MONOS_11014	3,1129	24,6539	21,5410	Dynein regulatory complex protein 1
MONOS_6486	3,1074	27,1368	24,0294	
MONOS_11571	2,9763	24,7255	21,7492	SAP domain
MONOS_2583	2,8786	22,8270	19,9484	RNA recognition motif
MONOS_1405	2,8425	22,7384	19,8959	
MONOS_11508	2,8108	23,3296	20,5188	Growth arrest-specific protein 8
MONOS_7248	2,7865	22,5063	19,7198	
MONOS_10118	2,7219	23,6459	20,9240	Coatomer subunit alpha
MONOS_339	2,6356	22,9119	20,2763	Bms11 protein
MONOS_2493	2,4488	23,2960	20,8472	
MONOS_9528	2,3994	25,0036	22,6042	
MONOS_14713; 3264	2,3553	22,8685	20,5132	
MONOS_16793	2,3491	23,108	20,7589	Ribosomal protein
MONOS_11531	2,2782	22,4264	20,1482	

MONOS_7186	2,2592	24,7191	22,4599	RNA-binding region RNP-1 domain-
MONOS_4051	2,1157	23,7143	21,5986	60S ribosomal protein L35-3
MONOS_2848	2,0899	23,6433	21,5534	RNA-binding region RNP-1 domain-
MONOS_164	2,0486	25,352	23,3034	RNA binding rRNA processing protein
MONOS_5185	2,0332	23,3246	21,2914	Nucleolar family member protein
MONOS_577; 231	1,9307	22,0479	20,1172	Phosphoserine aminotransferase
MONOS_2829	1,916	28,3615	26,4455	Linker histone H1 and H5 family protein
MONOS_1153	1,9006	26,2342	24,3336	Linker histone H1 and H5 family protein
MONOS_2912	1,8059	22,6743	20,8684	Stromal cell-derived factor 2
MONOS_10771	1,7468	23,0963	21,3495	
MONOS_14814	1,7429	21,6346	19,8917	V-type proton ATPase subunit D
MONOS_248	1,7162	21,1215	19,4053	Dihydropyrimidine dehydrogenase
MONOS_6851	1,7096	26,3882	24,6786	Ubiquitin-like protein
1	1,6975	27,7288	26,0313	Histone H2B
MONOS_1252	1,5633	23,6988	22,1355	Growth-arrest specific micro-tubule
MONOS_15415; 11988	1,5626	25,2242	23,6616	40S ribosomal protein S14-1
MONOS_4284	1,5516	27,254	25,7024	Histone H2A
MONOS_5711; 12526	1,5191	26,1017	24,5826	Large subunit ribosomal protein L4e,
MONOS_6203; 16573; 10323	1,5004	22,7428	21,2424	40S ribosomal protein S17-1
MONOS_8947; 3033	1,4966	23,3221	21,8255	Vacuolar protein sorting 26A (Vps26A)
MONOS_11734	1,4332	24,123	22,6898	40S ribosomal protein S20-1
MONOS_13329; 13328	1,4302	20,9962	19,5660	
2	1,4279	25,8536	24,4257	Histone H2A
3	1,4041	29,0188	27,6147	Histone H4
MONOS_14917	1,4029	23,7689	22,366	RNA-binding domain-containing
MONOS_862	1,3879	21,8839	20,4960	
MONOS_13412	1,3841	22,8355	21,4514	60S acidic ribosomal protein P2-1
MONOS_3824	1,3811	22,5078	21,1267	Brix domain-containing protein 2-like
MONOS_4567; 3180; 13168	1,3362	24,4914	23,1552	40S ribosomal protein S8-1
MONOS_180; 13171	1,3346	22,6356	21,3010	40S ribosomal protein S10-1
MONOS_11527	1,3292	23,4237	22,0945	
MONOS_15499	1,3221	23,6066	22,2845	Acetyl-CoA synthetase (ADP-forming)
MONOS_1967	1,3116	25,5675	24,2559	60s ribosomal protein L34e
MONOS_9987; 333; 2201	1,3019	23,9049	22,603	60S ribosomal protein L5-2
MONOS_1005	1,2994	21,9106	20,6112	Threonyl-tRNA synthetase
MONOS_9644	1,2804	25,7362	24,4558	RNA recognition motif
MONOS_9530	1,2638	20,8368	19,573	Pyruvate-ferrodoxin oxidoreductase
MONOS_4805; 4487	1,2196	24,9992	23,7796	Ribosomal protein L36e
MONOS_10479	1,1883	23,4158	22,2275	ATP-dependent RNA helicase Uap56
MONOS_5073	1,1798	25,4751	24,2953	Nucleosome assembly protein
MONOS_6964; 3400; 11718	1,1790	23,2939	22,1149	Glutamate dehydrogenase 3
MONOS_7390	1,1535	26,4306	25,2771	Nucleosome assembly protein
MONOS_10112; 9470	1,1499	24,5069	23,3570	40S ribosomal protein S24
MONOS_3728; 3403	1,1373	23,7363	22,5990	Small subunit ribosomal protein 7
MONOS_10289; 14516	1,1268	23,5258	22,3990	Bifunctional acetaldehyde-CoA/alcohol
MONOS_6798; 1943	1,0978	23,7757	22,6779	60S ribosomal protein L13a-1
MONOS_8272	1,0892	20,2840	19,1948	Poly [ADP-ribose] polymerase 1
MONOS_1697; 13174; 861	1,0603	22,8110	21,7507	Arf1f
MONOS_7522; 5302	1,0139	26,2202	25,2063	Putative phosphoglycerate

MONOS_9916	1,0099	24,2891	23,2792	Profilin I
MONOS_16368	1,0068	22,4491	21,4423	Calmodulin
MONOS_3432	1,0048	21,6685	20,6637	Autophagy-related protein 3 B
MONOS_11114; 11452	-	24,2745	-	Lysozyme
MONOS_1776	-	23,7097	-	Gar1/Naf1 RNA binding region
MONOS_10153; 15511	-	23,5428	-	Ribosomal protein L37e
MONOS_1399	-	23,4239	-	
MONOS_12194	-	23,3901	-	
MONOS_3598; 7848	-	23,1546	-	60S ribosomal protein L19-2
MONOS_16491	-	22,9425	-	
MONOS_8697; 12488	-	22,9244	-	Flagellar associated protein
MONOS_8919	-	22,8979	-	
MONOS_6982; 3552	-	22,7823	-	Ribosomal protein S28e
MONOS_9726	-	22,7198	-	
MONOS_10099	-	22,5317	-	
MONOS_12518	-	22,3225	-	Endothelin-converting enzyme 2 B
MONOS_11139	-	22,0998	-	
MONOS_15746; 8916	-	22,0795	-	60S ribosomal protein L26e

1: MONOS_8879; 5027; 321; 2901; 2133; 9813

2: MONOS_8882; 8757; 2903; 2131; 14653; 2128

3: MONOS_7697; 561; 2791; 2067; 13751; 1327; 13157; 11535

Tabulka č. 8: Výsledky hmotnostní spektrometrie eluátu Trichomonas pro „upstream“ oblast GAPDH. Hodnoty naměřených intenzit ve vzorku a negativní kontrole jsou uvedeny ve dvojkových logaritmech. Seřazeno podle velikosti rozdílu těchto hodnot.

gen	rozdíl	vzorek	kontrola	anotace
TVAG_251250	3,04327	21,4732	18,4300	Conserved hypothetical protein
TVAG_450700; 44270	2,47755	22,4940	20,0164	Conserved hypothetical protein
TVAG_296920	2,26959	20,3174	18,0478	Conserved hypothetical protein
4	1,90452	23,2395	21,335	Histone H4, putative
TVAG_298320	1,71849	22,3552	20,6367	Trichohyalin, putative
TVAG_210010	1,42912	19,7781	18,349	Conserved hypothetical protein
TVAG_198110; 254890	1,29121	19,7265	18,4353	Pyruvate-flavodoxin oxidoreductase
5	1,14260	18,0328	16,8902	Chaperonin, putative
TVAG_253630	1,08404	19,9919	18,9079	Heat shock protein
TVAG_539120	-	22,3182	-	Conserved hypothetical protein
TVAG_040700	-	21,3853	-	Conserved hypothetical protein
TVAG_444360	-	21,2359	-	Conserved hypothetical protein
TVAG_499020	-	21,1514	-	Histone H3, putative
TVAG_021100	-	21,1102	-	Replication factor C large subunit

4: TVAG_499010; 470680; 452720; 445260; 414470; 399380; 392270; 351560; 341350; 315310; 270900; 270070; 239320; 224660; 211270; 191680; 137930; 129720; 085660; 014920

5: TVAG_088050; 203620; 167250

6. Diskuse

6.1. Promotorové oblasti

Pro zahájení transkripce pomocí RNA polymerázy II je v eukaryotických buňkách nezbytná přítomnost konzervovaných sekvenčních motivů, které se nejčastěji nacházejí přibližně 40 nukleotidů před a 40 nukleotidů za místem začátku transkripce (Atkinson a Halfon, 2014).

Pomocí algoritmu MEME, který se obvykle používá pro identifikaci konzervovaných motivů v promotorových oblastech, jsme identifikovali v případě *Monocercomonoides* pouze jeden motiv (motiv 1). Jedná se o sekvenci rozpoznávanou zinc finger motivem, který se vyskytuje u velkého počtu transkripčních faktorů (Jolma et al., 2013; Mathelier et al., 2016). Pozice tohoto motivu nebyla konzervována.

Pomocí algoritmu DREME, který je vhodnější pro analýzu velkých datasetů krátkých sekvencí, jsme identifikovali poměrně velký počet konzervovaných motivů. Motiv č. 10 CCA(C/T)A se nejčastěji nachází 15 nukleotidů před začátkem transkripce. Tato sekvence je rozpoznávána transkripčními faktory z rodiny forkhead proteinů (Jolma et al., 2013). Motiv č. 12 GATCAAA(A/C) se nachází v genech kolem místa startu transkripce nebo 25 nukleotidů před ním. Tento motiv u člověka váže transkripční faktor 7 (TCF7) (Jolma et al., 2013; Mathelier et al., 2016). Tato sekvence však byla nalezena pouze u necelých 650 sekvencí. 45 nukleotidů před začátkem transkripce se nachází motiv č. 5 (A/G)TATTT(T/C/G), který může být TATA-like motivem, vyskytuje se v 52 % genů. Pouze pomocí programu Centrimo byl s malou statistickou signifikancí nalezen konsenzuální TATA box a to 15 nukleotidů před začátkem transkripce, nalezen byl u 44 % genů. Pro definitivní potvrzení funkce všech nalezených motivů v promotorových oblastech by byl potřeba další výzkum.

6.2. 5' nepřekládané oblasti

V rámci posttranskripční regulace genové exprese hrají nepřekládané oblasti nejdůležitější úlohu. Ovlivňují stabilitu mRNA, regulují míru její translace a také její

lokalizaci v buňce (Bashirullah et al., 2001; Jansen, 2001). Průměrná délka 5' UTR se uvádí mezi 100 a 200 nukleotidy (Pesole et al., 2001). Tyto hodnoty jsou ale založeny na údajích z rostlin, hub a živočichů. Mezi těmito skupinami se délka 5' UTR příliš neliší. Mezi prvky je rozpětí délek vyšší.

Z celkového počtu přibližně 16 500 genových modelů v genomu *Monocercomonoides* byla v programu PASA predikována 5' UTR pouze necelým 7000 z nich. Vzhledem k tomu, že je tato predikce založena především na transkriptomických datech, je tento počet překvapivě nízký. Možným vysvětlením nízkého počtu predikcí mohou být nesprávně predikované geny, pro které neexistují odpovídající transkripty. To hrozí zejména u genů, které se nepodařilo funkčně anotovat. Dataset tvořený všemi geny jsme použili pro porovnání GC obsahu 5' UTR s ostatními oblastmi genomu, které byly také automaticky predikovány v PASA (tabulka č. 4). Obsah GC párů v 5' UTR je v tomto datasetu 36 %, což je méně než u kódujících oblastí (43 %) a více než u intronů (25 %) nebo mezigenových oblastí (29 %).

5' nepřekládané oblasti jsou uváděny jako oblasti s vysokým počtem GC párů (60 % u teplokrevných obratlovců), vyšším než u 3' UTR i kódujících oblastí (45 % u stejné skupiny) (Pesole et al., 2001; Zhang et al., 2004). Podle našich výsledků toto v případě *Monocercomonoides* neplatí. 36 % GC párů v 5' UTR zhruba odpovídá obsahu GC párů v celém genomu a je jen o málo vyšší než ve 3' nepřekládaných oblastech. Naopak je tato hodnota nižší než u kódujících oblastí (43 %) (tabulka č. 4). Zastoupení GC párů v genomu několika dalších prvků se pohybuje mezi 23 % u *Tetrahymena thermophila* po 38 % u *Babesia bovis* (tabulka č. 5). U *Trichomonas vaginalis* je zastoupení GC párů v 5' UTR (31 %) také podobné jako v celém genomu (33 %) a je také o něco nižší než v kódujících oblastech (35 %) (Woehle et al., 2014).

Průměrnou délku UTR jsme spočítali pro několik datasetů. Pro všechny geny s predikovanou 5' UTR (109 nukleotidů), pro anotované geny (65 nukleotidů) a pro geny, které prošly manuální anotací (46 nukleotidů). Domníváme se, že v datasetu automaticky anotovaných genů se nepřesné genové modely, popřípadě špatně

přiřazené anotace, vyskytují v menší míře. Nejpřesnější hodnota pravděpodobně pochází z datasetu manuálně anotovaných genů, kde jsou zkontrolovány a popřípadě upraveny hranice mezi kódující a nekódující oblastí.

Pokles průměrné délky mezi těmito sadami dat je způsoben zejména odstraněním velmi dlouhých UTR (více než 200 nukleotidů, nebo dokonce více než 1000 nukleotidů), jak je patrné na grafech č. 4 a č. 5. Nejvíce 5' UTR *Monocercomonoides* má délku mezi 20 a 30 nukleotidy.

Z grafů č. 4 až č. 6 vyplývá, že přítomností extrémně dlouhých UTR je zatížen odhad délky 5' nepřekládaných oblastí i u ostatních prvoků. Pro toto porovnání je tedy pravděpodobně vhodnější použít dataset všech anotovaných genů. Délka 65 nukleotidů je srovnatelná s 53 nukleotidy u *Entamoeba histolytica* nebo 82 u *Trichomonas vaginalis* (tabulka č. 5). Hodnoty průměrných délek 5' UTR u prvoků ale kolísají od několika málo nukleotidů u *Giardia intestinalis* až po 346 u *Plasmodium falciparum* (Morrison et al., 2007; Watanabe et al., 2002).

Hodnoty délek 5' UTR u *Monocercomonoides* pocházely pouze z automatické predikce provedené v programu PASA. Tyto předpovědi nemusejí být přesné, anotační programy mají často s predikcemi nepřekládaných oblastí nebo intronů v genech velké potíže (Yandell a Ence, 2012). Rozhodli jsme se tedy provést 5' RACE experimenty, abychom přesnost alespoň několika předpovědí ověřili. Podařilo se nám amplifikovat nepřekládané oblasti pro GAPDH (24 nukleotidů, predikce byla prodloužena o 7 nukleotidů), enolázu (24 nukleotidů, predikce zkrácena o 111 nukleotidů), Cpn 60 (13 nukleotidů, predikce prodloužena o 5 nukleotidů), histon H3 (35, nukleotidů, predikce prodloužena o 6 nukleotidů) a acetyl-CoA syntetázu (38 nukleotidů, predikce zkrácena o 10 nukleotidů). Z výsledků je patrné, že přes poměrně kvalitní transkriptomická data se automatická predikce UTR s našimi experimenty příliš neshoduje. U enolázy a acetyl-CoA syntetázy jsme také potvrdili introny v 5' UTR.

Určení délek pomocí 5' UTR pomocí RACE ovšem také nemusí být přesné. Jediný kit pro RACE experimenty, který fungoval v případě *Monocercomonoides*,

bohužel neamplifikuje selektivně pouze mRNA s připojenou 7-methylguanositovou čepičkou. Z toho vyplývá, že může teoreticky dojít k připojení adaptéru i na nekompletní mRNA a následné amplifikaci této mRNA. Druhým problémem komplikujícím interpretaci výsledků RACE experimentů je to, že máme k dispozici údaje pouze pro pět genů. Z takto malého vzorku lze samozřejmě pouze obtížně vyvozovat závěry s platností pro celý genom. Přesto, výsledky z 5' RACE experimentů naznačují, že na automatické predikce 5' UTR nelze spoléhat s jistotou (u žádného z pěti genů nebyla předpověď 100% přesná). V případě potřeby zjištění přesného místa začátku transkripce pro jeden konkrétní gen na tyto predikce nelze spoléhat a je třeba využít RACE.

6.3. Transkripční a translační faktory

Pomocí databáze KEGG jsme automaticky anotovali, následně zkontrolovali a eventuálně opravili 23 genových modelů pro základní transkripční faktory a 30 pro eukaryotické iniciační faktory.

Monocercomonoides ve svém genomu obsahuje v podstatě totožnou sadu transkripčních iniciačních faktorů jako *Trichomonas vaginalis*. Nejvýraznějším rozdílem je přítomnost obou podjednotek transkripčního faktoru IIA u oxymonády. Tento protein napomáhá vazbě TATA-binding proteinu (TBP) na TATA box a odpovídá tak pravděpodobné přítomnosti TATA boxu nebo TATA-like motivu u *Monocercomonoides*, zatímco u trichomonády tento motiv nalezen nebyl. Z hlediska transkripčních faktorů je *Giardia intestinalis* unikátní organismus, protože u ní bylo nalezeno pouze několik podjednotek TFIIF a TBP (Best et al., 2004; Morrison et al., 2007).

Sada eukaryotických iniciačních translačních faktorů v genomu *Monocercomonoides* je opět velmi podobná té u trichomonády, a tentokrát i u giardie. U oxymonády jsme našli většinu nejdůležitějších iniciačních translačních faktorů, s výjimkou faktorů 5A a 2A.

Po anotaci základních transkripčních faktorů jsme se pokusili některé z transkripčních faktorů purifikovat. Tyto experimenty jsme prováděli s pomocí „upstream“ sekvencí genu pro GAPDH u *Monocercomonoides* (MONOS_5021) i u *Trichomonas* (TVAG_382880). Cílem bylo izolovat proteiny, které regulují transkripci tohoto genu. Ve směsi vypurifikovaných proteinů se bohužel nevyskytoval žádný známý transkripční faktor ani u jednoho organismu. To získané výsledky výrazně zpochybňuje. U *Monocercomonoides* (tabulka č. 7) jsou eluáty obohaceny o ribozomální a různé jaderné proteiny, které však pravděpodobně tvoří pouze nespecifické pozadí. Dále je ve výsledcích množství proteinů, které obsahují RNA vazebný motiv. Tento motiv je typický pro některé jaderné proteiny, ale vyskytuje se i u několika proteinů, které vážou jednovláknovou DNA a regulují transkripci. Tyto proteiny jsou společně s proteiny neznámé funkce možnými transkripčními faktory pro GAPDH. To by však bylo potřeba prozkoumat dalšími metodami. Dnes asi nejrozšířenější metodou je ChIP (Chromatin immunoprecipitation assay). Pro tuto metodu bychom ale museli připravit specifické protilátky proti všem potenciálním transkripčním faktorům. V prvním kroku dojde k formaldehydovému „cross-linku“ mezi DNA a proteiny v živých buňkách. Následně jsou buňky sonikovány. V průběhu sonikace dojde k rozbití DNA na přibližně 500 bp dlouhé úseky. Transkripční faktory s navázanou DNA jsou následně imunoprecipitovány pomocí specifických protilátek. DNA je v další fázi disociována od proteinů a je dále analyzována.

Většinu proteinů identifikovaných v experimentu s *Trichomonas vaginalis* (tabulka č. 8) jsou proteiny neznámé funkce. Vzhledem k množství nalezených proteinů i jejich naměřených intenzit je ale bohužel nutné i tyto výsledky považovat za neprůkazné.

Domníváme se, že na neúspěch naší purifikace mohla mít vliv nedostatečně optimalizovaná doba inkubace označené DNA s buněčným lyzátem. Tato doba mohla být příliš dlouhá, takže v lyzátu došlo již k disociaci transkripčních faktorů od DNA a rozběhnutí transkripce.

7. Závěry

- Identifikovali jsme 12 konzervovaných motivů v promotorových oblastech *Monocercomonoides*, včetně TATA boxu a TATA-like motivu.
- 5' nepřekládané oblasti oxymonády *Monocercomonoides* mají průměrnou délku 46 nukleotidů a obsahují 33 % GC párů.
- Provedené 5' RACE experimenty na pěti transkriptech ukázaly poměrně velkou nepřesnost automatické predikce 5' UTR.
- Anotovali jsme 23 genových modelů základních transkripčních a 30 eukaryotických iniciačních translačních faktorů v genomu *Monocercomonoides*. Sady těchto proteinů se výrazně neliší od sad přítomných u *Trichomonas vaginalis*.
- Pokusili jsme se purifikovat transkripční faktory pomocí „upstream“ oblasti genu pro GAPHD u *Monocercomonoides* sp. a *Trichomonas vaginalis*, ale s neprůkaznými výsledky.

8. Literatura

- Adam, R.D., 2000. The *Giardia lamblia* genome. *Int. J. Parasitol.* 30, 475–84.
- Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., le Gall, L., Lynn, D.H., McManus, H., Mitchell, E.A.D., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C.L., Smirnov, A. a Spiegel, F.W., 2012. The Revised Classification of Eukaryotes. *J. Eukaryot. Microbiol.* 59, 429–514.
- Albright, S.R. a Tjian, R., 2000. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* 242, 1–13.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. a Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–10.
- Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C. a Penalva, L.O.F., 2012. Before It Gets Started: Regulating Translation at the 5' UTR. *Comp. Funct. Genomics* 2012, 475731.
- Aravind, L. a Koonin, E. V, 2000. Eukaryote-specific domains in translation initiation factors: Implications for translation regulation and evolution of the translation system. *Genome Res.* 10, 1172–1184.
- Atkinson, T. a Halfon, M., 2014. Regulation of gene expression in the genomic context. *Comput. Struct. Biotechnol. J.* 9, e201401001.
- Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27, 1653–9.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. a Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
- Bailey, T.L. a Machanick, P., 2012. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* 40, e128.
- Bashirullah, A., Cooperstock, R.L. a Lipshitz, H.D., 2001. Spatial and temporal control of RNA stability. *Proc. Natl. Acad. Sci.* 98, 7025–7028.
- Benson, D.A., Karsch-Mizrachi, I., Clark, K., Lipman, D.J., Ostell, J. a Sayers, E.W., 2012. GenBank. *Nucleic Acids Res.* 40, D48–D53.
- Best, A.A., Morrison, H.G., McArthur, A.G., Sogin, M.L. a Olsen, G.J., 2004. Evolution of Eukaryotic Transcription : Insights From the Genome of *Giardia lamblia* 1537–1547.
- Bhattacharya, A., Satish, S., Bagchi, A. a Bhattacharya, S., 2000. The genome of *Entamoeba histolytica*. *Int. J. Parasitol.* 30, 401–10.

- Brancucci, N.M.B., Witmer, K., Schmid, C. a Voss, T.S., 2014. A var gene upstream element controls protein synthesis at the level of translation initiation in *Plasmodium falciparum*. PLoS One 9, e100183.
- Brugerolle, G. a Lee, J.J., 2000. Order Oxymonadida, in: Lee, J., Leedale, G., a Bradbury, P. (Ed.), The Illustrated Guide to the Protozoa. Allen Press Inc, s. 1186–1195.
- Buendía-Orozco, J., Guerrero, A. a Pastor, N., 2005. Model of the TBP–TFIIB Complex from *Plasmodium falciparum*: Interface Analysis and Perspectives as a New Target for Antimalarial Design. Arch. Med. Res. 36, 317–330.
- Bunnik, E.M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L. a Le Roch, K.G., 2016. The mRNA-bound proteome of the human malaria parasite *Plasmodium falciparum*. Genome Biol. 17, 147.
- Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., Wortman, J.R., Bidwell, S.L., Alsmark, U.C.M., Besteiro, S., Sicheritz-Ponten, T., Noel, C.J., Dacks, J.B., Foster, P.G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G.J., Westrop, G.D., Müller, S., Dessi, D., Fiori, P.L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F.D., Simoes-Barbosa, A., Brown, M.T., Hayes, R.D., Mukherjee, M., Okumura, C.Y., Schneider, R., Smith, A.J., Vaňáčová, Š., Villalvazo, M., Haas, B.J., Perteua, M., Feldblyum, T. V, Utterback, T.R., Shu, C.-L., Osoegawa, K., de Jong, P.J., Hrdý, I., Horváthová, L., Zubáčová, Z., Doležal, P., Malik, S.-B., Logsdon, J.M., Henze, K., Gupta, A., Wang, C.C., Dunne, R.L., Upcroft, J.A., Upcroft, P., White, O., Salzberg, S.L., Tang, P., Chiu, C.-H., Lee, Y.-S., Embley, T.M., Coombs, G.H., Mottram, J.C., Tachezy, J., Fraser-Liggett, C.M. a Johnson, P.J., 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science 315, 207–12.
- Carvajal-Gamez, B.I., Quintas-Granados, L.I., Arroyo, R., Mendoza-Hernández, G. a Alvarez-Sánchez, M.E., 2012. Translation initiation factor eIF-5A, the hypusine-containing protein, is phosphorylated on serine and tyrosine and O-glycosylated in *Trichomonas vaginalis*. Microb. Pathog. 52, 177–183.
- Clark, C.G., Alsmark, U.C.M., Tazreiter, M., Saito-Nakano, Y., Ali, V., Marion, S., Weber, C., Mukherjee, C., Bruchhaus, I., Tannich, E., Leippe, M., Sicheritz-Ponten, T., Foster, P.G., Samuelson, J., Noël, C.J., Hirt, R.P., Embley, T.M., Gilchrist, C. a Mann, B.J., Singh, U., Ackers, J.P., Bhattacharya, S., Bhattacharya, A., Lohia, A., Guillén, N., Duchêne, M., Nozaki, T. a Hall, N., 2007. Structure and content of the *Entamoeba histolytica* genome. Adv. Parasitol. 65, 51–190.
- Cong, P., Luo, Y., Bao, W. a Hu, S., 2010. Genomic organization and promoter analysis of the *Trichomonas vaginalis* core histone gene families. Parasitol. Int. 59, 29–34.

- da Silva Augusto, L., Moretti, N.S., Ramos, T.C.P., de Jesus, T.C.L., Zhang, M., Castilho, B.A. a Schenkman, S., 2015. A Membrane-bound eIF2 Alpha Kinase Located in Endosomes Is Regulated by Heme and Controls Differentiation and ROS Levels in *Trypanosoma cruzi*. *PLOS Pathog.* 11, e1004618.
- De Gaudenzi, J.G., Noé, G., Campo, V.A., Frasch, A.C. a Cassola, A., 2011. Gene expression regulation in trypanosomatids. *Essays Biochem.* 51, 31–46.
- Dechering, K.J., Kaan, A.M., Mbacham, W., Wirth, D.F., Eling, W., Konings, R.N. a Stunnenberg, H.G., 1999. Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell. Biol.* 19, 967–78.
- Dhalia, R., Marinsek, N., Reis, C.R.S., Katz, R., Muniz, J.R.C., Standart, N., Carrington, M. a de Melo Neto, O.P., 2006. The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct. *Nucleic Acids Res.* 34, 2495–507.
- Dhalia, R., Reis, C.R.S., Freire, E.R., Rocha, P.O., Katz, R., Muniz, J.R.C., Standart, N. a de Melo Neto, O.P., 2005. Translation initiation in *Leishmania major*: characterisation of multiple eIF4F subunit homologues. *Mol. Biochem. Parasitol.* 140, 23–41.
- Diamond, L.S., 1957. The establishment of various trichomonads of animals and man in axenic cultures. *J. Parasitol.* 43, 88–490.
- Diamond, L.S., 1982. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and other lumen dwelling protozoa. *J. Parasitol.* 68, 958–959.
- Díaz, S., Amaro, F., Rico, D., Campos, V., Benítez, L., Martín-González, A., Hamilton, E.P., Orias, E. a Gutiérrez, J.C., 2007. *Tetrahymena* Metallothioneins Fall into Two Discrete Subfamilies. *PLoS One* 2, e291.
- Espinosa, N., Hernández, R., López-Griego, L. a López-Villaseñor, I., 2002. Separable putative polyadenylation and cleavage motifs in *Trichomonas vaginalis* mRNAs. *Gene* 289, 81–6.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.L. a Bateman, A., 2008. The Pfam protein families database. *Nucleic Acids Res.* 36, D281–8.
- Formigari, A., Boldrin, F., Santovito, G., Cassidy-Hanley, D., Clark, T.G. a Piccinni, E., 2010. Functional Characterization of the 5'-upstream Region of MTT5 Metallothionein Gene from *Tetrahymena thermophila*. *Protist* 161, 71–77.
- Freire, E.R., Dhalia, R., Moura, D.M.N., da Costa Lima, T.D., Lima, R.P., Reis, C.R.S., Hughes, K., Figueiredo, R.C.B.Q., Standart, N., Carrington, M. a de Melo Neto, O.P., 2011. The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. *Mol. Biochem. Parasitol.* 176, 25–36.

- Freire, E.R., Malvezzi, A.M., Vashisht, A.A., Zuberek, J., Saada, E.A., Langousis, G., Nascimento, J.D.F., Moura, D., Darzynkiewicz, E., Hill, K., de Melo Neto, O.P., Wohlschlegel, J.A., Sturm, N.R. a Campbell, D.A., 2014a. *Trypanosoma brucei* translation initiation factor homolog EIF4E6 forms a tripartite cytosolic complex with EIF4G5 and a capping enzyme homolog. *Eukaryot. Cell* 13, 896–908.
- Freire, E.R., Vashisht, A.A., Malvezzi, A.M., Zuberek, J., Langousis, G., Saada, E.A., Nascimento, J.D.F., Stepinski, J., Darzynkiewicz, E., Hill, K., De Melo Neto, O.P., Wohlschlegel, J.A., Sturm, N.R. a Campbell, D.A., 2014b. eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in *Trypanosoma brucei*. *RNA* 20, 1272–86.
- Froger, A. a Hall, J.E., 2007. Transformation of plasmid DNA into *E. coli* using the heat shock method. *J. Vis. Exp.* 253.
- Fuentes, V., Barrera, G., Sánchez, J., Hernández, R. a López-Villaseñor, I., 2012. Functional analysis of sequence motifs involved in the polyadenylation of *Trichomonas vaginalis* mRNAs. *Eukaryot. Cell* 11, 725–34.
- Gilchrist, C.A., Purdy, J., Mann, B.J. a Petri, W.A., 1997. Control of gene expression in *Entamoeba histolytica* by a cis-acting upstream regulatory element. *Arch. Med. Res.* 28 Spec No, 39–40.
- Gomez, C., Esther Ramirez, M., Calixto-Galvez, M., Medel, O. a Rodríguez, M.A., 2010. Regulation of gene expression in protozoa parasites. *J. Biomed. Biotechnol.* 2010, 726045.
- Gopalakrishnan, A.M., Nyindodo, L.A., Ross Fergus, M. a López-Estraño, C., 2009. *Plasmodium falciparum*: Preinitiation complex occupancy of active and inactive promoters during erythrocytic stage. *Exp. Parasitol.* 121, 46–54.
- Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E. a Pesole, G., 2010. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* 38, D75–80.
- Guillebault, D., Sasorith, S., Derelle, E., Wurtz, J.-M., Lozano, J.-C., Bingham, S., Tora, L. a Moreau, H., 2002. A New Class of Transcription Initiation Factors, Intermediate between TATA Box-binding Proteins (TBPs) and TBP-like Factors (TLFs), Is Present in the Marine Unicellular Organism, the Dinoflagellate *Cryptothecodinium cohnii*. *J. Biol. Chem.* 277, 40881–40886.
- Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. a Noble, W., 2007. Quantifying similarity between motifs. *Genome Biol.* 8, R24.
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R. a Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7.

- Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*
- Hinnebusch, A.G., 2014. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* 83, 779–812.
- Hinnebusch, A.G., Ivanov, I.P. a Sonenberg, N., 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–6.
- Holberton, D. V a Marshall, J., 1995. Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. *Nucleic Acids Res.* 23, 2945–53.
- Horrocks, P. a Lanzer, M., 1999. Mutational analysis identifies a five base pair cis-acting sequence essential for GBP130 promoter activity in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 99, 77–87.
- Huang, W., Wong, J.M. a Bateman, E., 1996. TATA elements direct bi-directional transcription by RNA polymerases II and III. *Nucleic Acids Res.* 24, 1158–63.
- Hughes, M.J.. a Andrews, D.W., 1997. A single nucleotide is a sufficient 5' untranslated region for translation in an eukaryotic in vitro system. *FEBS Lett.* 414, 19–22.
- Chung, J., Rocha, A.A., Tonelli, R.R., Castilho, B.A. a Schenkman, S., 2013. Eukaryotic initiation factor 5A dephosphorylation is required for translational arrest in stationary phase cells. *Biochem. J.* 451, 257–67.
- Jackson, R.J., Hellen, C.U.T. a Pestova, T. V, 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–27.
- Jagus, R., Bachvaroff, T.R., Joshi, B. a Place, A.R., 2012. Diversity of eukaryotic translational initiation factor eIF4E in protists. *Comp. Funct. Genomics.*
- Jang, S.K., Kräusslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C. a Wimmer, E., 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* 62, 2636–43.
- Jansen, R.-P., 2001. mRNA localization: message on the move. *Nat. Rev. Mol. Cell Biol.* 2, 247–256.
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. a Smale, S.T., 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell Biol.* 14, 116–127.
- Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T. a Taipale, J., 2013. DNA-binding specificities of human transcription factors. *Cell* 152, 327–39.

- Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W. a Kadonaga, J.T., 2008. The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell Biol.* 20, 253–9.
- Kanehisa, M. a Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K., Soukal, P., Hroudová, M., Doležal, P., Stairs, C.W., Roger, A.J., Eliáš, M., Dacks, J.B., Vlček, Č. a Hampl, V., 2016. A Eukaryote without a Mitochondrial Organelle. *Curr. Biol.* 26, 1274–1284.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. a Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–9.
- Keeling, P.J., Fast, N.M. a McFadden, G.I., 1998. Evolutionary relationship between translation initiation factor eIF-2gamma and selenocysteine-specific elongation factor SELB: change of function in translation factors. *J. Mol. Evol.* 47, 649–55.
- Kozak, M., 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292.
- Krishnamurthy, R. a Sultana, T., 1976. *Tubulimonoides gryllotalpae* n.g., n.sp. (Mastigophora: Oxymonadida) from the cricket in India. *Proc. Indian Acad. Sci.* 84, 137–140.
- Kulda, J. a Nohýnková, E., 1978. Flagellates of the human intestine and of intestines of other species, in: Kreier, J.P. (Ed.), *Parasitic protozoa*. New York: Academic Press, s. 1–138.
- Kumar, M., Srinivas, V. a Patankar, S., 2015. Upstream AUGs and upstream ORFs can regulate the downstream ORF in *Plasmodium falciparum*. *Malar. J.* 14, 512.
- Li, L. a Wang, C.C., 2005. Identification in the ancient protist *Giardia lamblia* of two eukaryotic translation initiation factor 4E homologues with distinctive functions. *Eukaryot. Cell* 4, 948–59.
- López-Lastra, M., Rivas, A. a Barría, M.I., 2005. Protein synthesis in eukaryotes: The growing biological relevance of cap-independent translation initiation. *Biol. Res.*
- Luna-Arias, J.P., Hernandez-Rivas, R., de Dios-Bravo, G., Garcia, J., Mendoza, L. a Orozco, E., 1999. The TATA-box binding protein of *Entamoeba histolytica*: cloning of the gene and location of the protein by immunofluorescence and confocal microscopy. *Microbiology* 145, 33–40.

- Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A. a Wasserman, W.W., 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–5.
- McAndrew, M.B., Read, M., Sims, P.F. a Hyde, J.E., 1993. Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite *Plasmodium falciparum*. *Gene* 124, 165–71.
- Mignone, F., Gissi, C., Liuni, S. a Pesole, G., 2002. Untranslated regions of mRNAs. *Genome Biol* 1–10.
- Militello, K.T., Dodge, M., Bethke, L. a Wirth, D.F., 2004. Identification of regulatory elements in the *Plasmodium falciparum* genome. *Mol. Biochem. Parasitol.* 134, 75–88.
- Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., Best, A.A., Cande, W.Z., Chen, F., Cipriano, M.J., Davids, B.J., Dawson, S.C., Elmendorf, H.G., Hehl, A.B., Holder, M.E., Huse, S.M., Kim, U.U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J.E.J., Palm, D., Passamaneck, N.E., Prabhu, A., Reich, C.I., Reiner, D.S., Samuelson, J., Svard, S.G. a Sogin, M.L., 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317, 1921–6.
- Moura, D.M., Reis, C.R., Xavier, C.C., da Costa Lima, T.D., Lima, R.P., Carrington, M. a de Melo Neto, O.P., 2015. Two related trypanosomatid eIF4G homologues have functional differences compatible with distinct roles during translation initiation. *RNA Biol.* 12, 305–319.
- Nguyen, S., Leija, C., Kinch, L., Regmi, S., Li, Q., Grishin, N. V a Phillips, M.A., 2015. Deoxyhypusine Modification of Eukaryotic Translation Initiation Factor 5A (eIF5A) Is Essential for *Trypanosoma brucei* Growth and for Expression of Polyprolyl-containing Proteins. *J. Biol. Chem.* 290, 19987–98.
- Orphanides, G., Lagrange, T. a Reinberg, D., 1996. The general transcription factors of RNA polymerase II. *Genes Dev.* 10, 2657–83.
- Osta, M., Gannoun-Zaki, L., Bonnefoy, S., Roy, C. a Vial, H.J., 2002. A 24 bp cis-acting element essential for the transcriptional activity of *Plasmodium falciparum* CDP-diacylglycerol synthase gene promoter. *Mol. Biochem. Parasitol.* 121, 87–98.
- Pearson, R.J. a Singh, U., 2010. Approaches to characterizing *Entamoeba histolytica* transcriptional regulation. *Cell. Microbiol.* 12, 1681–90.

- Pereira, M.M.C., Malvezzi, A.M., Nascimento, L.M., da Costa Lima, T.D.C., Alves, V.S., Palma, M.L., Freire, E.R., Moura, D.M.N., Reis, C.R.S. a de Melo Neto, O.P., 2013. The eIF4E subunits of two distinct trypanosomatid eIF4F complexes are subjected to differential post-translational modifications associated to distinct growth phases in culture. *Mol. Biochem. Parasitol.* 190, 82–86.
- Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F. a Liuni, S., 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* 276, 73–81.
- Purdy, J.E., Pho, L.T., Mann, B.J. a Petri, W.A., 1996. Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Mol. Biochem. Parasitol.* 78, 91–103.
- Quintas-Granados, L.I., Carvajal Gamez, B.I., Villalpando, J.L., Ortega-Lopez, J., Arroyo, R., Azuara-Liceaga, E. a Álvarez-Sánchez, M.E., 2016. Bifunctional activity of deoxyhypusine synthase/hydroxylase from *Trichomonas vaginalis*. *Biochimie* 123, 37–51.
- Radek, R., 1994. *Monocercomonoides termitis* n. sp., an Oxymonad from the Lower Termite *Kaloterms sinaicus*. *Arch. für Protistenkd.* 144, 373–382.
- Radek, R., 1999. Flagellates, bacteria, and fungi associated with termites: diversity and function in nutrition—a review. *Ecotropica*.
- Reese, J.C., 2003. Basal transcription factors. *Curr. Opin. Genet. Dev.* 13, 114–118.
- Reeve, J.N., 2003. Archaeal chromatin and transcription. *Mol. Microbiol.* 48, 587–598.
- Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R.S. a de Melo Neto, O.P., 2014. The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates—identification of conserved and divergent features based on orthologue analysis. *BMC Genomics* 15, 1175.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. a Mesirov, J.P., 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Ruvalcaba-Salazar, O.K., Ramírez-Estudillo, M. del C., Montiel-Condado, D., Recillas-Targa, F., Vargas, M. a Hernández-Rivas, R., 2005. Recombinant and native *Plasmodium falciparum* TATA-binding-protein binds to a specific TATA box element in promoter regions. *Mol. Biochem. Parasitol.* 140, 183–196.
- Ruvalcaba-Salazar, O.-K., Romero-Ramírez, H., Santos-Argumedo, L., Vargas, M. a Hernández-Rivas, R., 2006. Preparation and Characterization of a Monoclonal Antibody Specific to *Plasmodium falciparum* TATA Binding Protein. *Hybridoma* 25, 367–371.
- Schumacher, M. a, Lau, A.O.T. a Johnson, P.J., 2003. Structural basis of core promoter recognition in a primitive eukaryote. *Cell* 115, 413–24.

- Simpson, A.G.B. a Roger, A.J., 2004. The real 'kingdoms' of eukaryotes. *Curr. Biol.*
- Singh, S., Raju, K., Jatekar, D., Dinesh, N., Paul, M.S. a Sobhia, M.E., 2014. *Leishmania donovani* eukaryotic initiation factor 5A: Molecular characterization, localization and homology modelling studies. *Microb. Pathog.* 73, 37–46.
- Singh, U. a Rogers, J., 1998. The novel core promoter element GAAC in the *hgl5* gene of *Entamoeba histolytica* is able to direct a transcription start site independent of TATA or initiator regions. *J. Biol. Chem.* 273, 21663–8.
- Singh, U., Rogers, J.B., Mann, B.J. a Petri, W. a, 1997. Transcription initiation is controlled by three core promoter elements in the *hgl5* gene of the protozoan parasite *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 8812–7.
- Smale, S.T. a Kadonaga, J.T., 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449–79.
- Smith, A.J., Chudnovsky, L., Simoes-Barbosa, A., Delgadillo-Correa, M.G., Jonsson, Z.O., Wohlschlegel, J. a a Johnson, P.J., 2011. Novel core promoter elements and a cognate transcription factor in the divergent unicellular eukaryote *Trichomonas vaginalis*. *Mol. Cell. Biol.* 31, 1444–58.
- Stargell, L.A. a Gorovsky, M.A., 1994. TATA-binding protein and nuclear differentiation in *Tetrahymena thermophila*. *Mol. Cell. Biol.* 14, 723–34.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Suyama, A., Sakaki, Y., Morishita, S., Okubo, K. a Sugano, S., 2001. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* 11, 677–84.
- Sweeney, R., Fan, Q. a Yao, M.C., 1996. Antisense ribosomes: rRNA as a vehicle for antisense RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8518–23.
- Tellier, G., Lenne, A., Cailliau-Maggio, K., Cabezas-Cruz, A., Valdés, J.J., Martoriati, A., Aliouat, E.M., Gosset, P., Delaire, B., Fréville, A., Pierrot, C. a Khalife, J., 2016. Identification of *Plasmodium falciparum* Translation Initiation eIF2 β Subunit: Direct Interaction with Protein Phosphatase Type 1. *Front. Microbiol.* 7, 777.
- Thompson, J.D., Higgins, D.G. a Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–80.
- Tonelli, R.R., Augusto, L. da S., Castilho, B.A. a Schenkman, S., 2011. Protein synthesis attenuation by phosphorylation of eIF2 α is required for the differentiation of *Trypanosoma cruzi* into infective forms. *PLoS One* 6, e27904.
- Vaňáčová, Š., Liston, D.R., Tachezy, J. a Johnson, P.J., 2003. Molecular biology of the amitochondriate parasites, *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*. *Int. J. Parasitol.* 33, 235–255.

- Vlasáková, J., 2014. Diverzita rodu *Monocercomonoides*. Univerzita Karlova v Praze.
- Watanabe, J., Sasaki, M., Suzuki, Y. a Sugano, S., 2002. Analysis of transcriptomes of human malaria parasite *Plasmodium falciparum* using full-length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs. *Gene* 291, 105–113.
- Woehle, C., Kusdian, G., Radine, C., Graur, D., Landan, G. a Gould, S.B., 2014. The parasite *Trichomonas vaginalis* expresses thousands of pseudogenes and long non-coding RNAs independently from functional neighbouring genes. *BMC Genomics* 15, 906.
- Wong, J.-M., Liu, F. a Bateman, E., 1992. Cloning and expression of the *Acanthamoeba castellanii* gene encoding transcription factor TFIID. *Gene* 117, 91–97.
- Xu, F., Jerlström-Hultqvist, J., Einarsson, E., Astvaldsson, A., Svärd, S.G. a Andersson, J.O., 2014. The Genome of *Spironucleus salmonicida* Highlights a Fish Pathogen Adapted to Fluctuating Environments. *PLoS Genet.* 10.
- Yandell, M. a Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. a Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134.
- Yee, J., Mowat, M.R., Dennis, P.P. a Nash, T.E., 2000. Transcriptional Analysis of the Glutamate Dehydrogenase Gene in the Primitive Eukaryote, *Giardia lamblia*. IDENTIFICATION OF A PRIMORDIAL GENE PROMOTER. *J. Biol. Chem.* 275, 11432–11439.
- Yee, J., Tang, A., Lau, W.-L., Ritter, H., Delport, D., Page, M., Adam, R.D., Müller, M. a Wu, G., 2007. Core histone genes of *Giardia intestinalis*: genomic organization, promoter structure, and expression. *BMC Mol. Biol.* 8, 26.
- Zamorano, A., López-Camarillo, C., Orozco, E., Weber, C., Guillen, N. a Marchat, L.A., 2008. In silico analysis of EST and genomic sequences allowed the prediction of cis-regulatory elements for *Entamoeba histolytica* mRNA polyadenylation. *Comput. Biol. Chem.* 32, 256–63.
- Zhang, L., Kasif, S., Cantor, C.R. a Broude, N.E., 2004. GC/AT-content spikes as genomic punctuation marks. *Proc. Natl. Acad. Sci.* 101, 16855–16860.
- Zhang, M., Joyce, B.R., Sullivan, W.J. a Nussenzweig, V., 2013. Translational control in *Plasmodium* and *Toxoplasma* parasites. *Eukaryot. Cell* 12, 161–7.
- Zhang, Q., Táborský, P., Silberman, J.D., Pánek, T., Čepička, I. a Simpson, A.G.B., 2015. Marine Isolates of *Trimastix marina* Form a Plesiomorphic Deep-branching Lineage within Preaxostyla, Separate from Other Known Trimastigids (*Paratrimastix* n. gen.). *Protist* 166, 468–491.

- Zhu, L., Mok, S., Imwong, M., Jaidee, A., Russell, B., Nosten, F., Day, N.P., White, N.J., Preiser, P.R. a Bozdech, Z., 2016. New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. *Sci. Rep.* 6, 20498.
- Zubáčová, Z., Novák, L., Bublíková, J., Vacek, V., Fousek, J., Rídl, J., Tachezy, J., Doležal, P., Vlček, Č. a Hampl, V., 2013. The Mitochondrion-Like Organelle of *Trimastix pyriformis* Contains the Complete Glycine Cleavage System. *PLoS One* 8, e55417.