

**Univerzita Karlova v Praze**  
Přírodovědecká fakulta  
Katedra zoologie



# RECENTNĚ AKTIVNÍ L1 A B1 RETROTRANSPOSONY V MYŠÍM GENOMU

Diplomová práce

**Bc. Václav Janoušek**

Vedoucí diplomové práce: Mgr. Pavel Munclinger, Ph.D.

Praha 2010

## Poděkování

Chtěl bych poděkovat svému školiteli Dr. Pavlovi Munclingerovi za cenné rady a připomínky, které mi poskytl během psaní diplomové práce.

Práce byla financována z grantu GAČR reg. č. 206/08/0640.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s použitím uvedené literatury.

V Praze dne:

.....

# OBSAH

---

<b>Abstrakt</b> .....	<b>6</b>
<b>1. Úvod</b> .....	<b>7</b>
<b>2. Literární přehled</b> .....	<b>8</b>
2.1 Retrotransposony .....	8
2.1.1 Základní dělení myších retrotransposonů .....	8
2.1.2 Základní charakteristika aktivních retrotransposonů .....	9
2.1.3 Nomenklatura aktivních retrotransposonů.....	10
2.1.4 Počet aktivních retrotransposonů v myším genomu .....	13
2.2 Propagace retrotransposonů.....	14
2.3 Modely šíření retrotransposonů.....	16
2.4 Aktivita retrotransposonů v genomu myši domácí.....	19
2.4.1 SINE.....	19
2.4.2 L1.....	22
2.5 Využití retrotransposonů v populační genetice .....	24
<b>3. Cíle práce</b> .....	<b>26</b>
<b>4. Materiál a metodika</b> .....	<b>27</b>
4.1 Analýza konsenzuálních sekvencí B1 podrodin.....	27
4.2 Sekvenční analýza recentně aktivních L1 a B1 retrotransposonů.....	27
4.2.1 Získání potenciálně polymorfních B1 a L1 retrotransposonů .....	27
4.2.2 Získání polymorfních recentně aktivních L1 a B1 retrotransposonů .....	28
4.2.3 Úprava a analýza sekvencí .....	29
4.2.4 Navržení primerů .....	30

4.3 Genotypizace L1 a B1 u 8 inbredních kmenů .....	31
4.3.1 Výběr vhodných kmenů.....	31
4.3.2 Izolace DNA .....	31
4.3.3 PCR.....	31
4.3.4 Elektroforéza .....	31
4.3.5 Hodnocení výsledků elektroforézy .....	32
<b>5. Výsledky .....</b>	<b>33</b>
5.1 Analýza konsenzuálních sekvencí B1 podrodin.....	33
5.1.1 Analýza konsenzuálních sekvencí B1 podrodin z databáze RepBase .....	33
5.1.2 Evoluční vztahy jednotlivých podrodin B1 .....	35
5.1.3 Porovnání podrodin z RepBase a podrodin definovaných Quentinem (1989) .....	38
5.2 Sekvenční analýza recentně aktivních L1 a B1 retrotransposonů .....	40
5.2.1 Recentně aktivní podrodiny .....	40
5.2.2 MSM/Ms vs. C57BL/6J původ.....	40
5.2.3 Analýza sekvencí.....	41
5.2.4 Srovnání aktivity podrodin B1 retrotransposonů .....	41
5.3 Genotypizace L1 a B1 insercí u inbredních 8 kmenů.....	45
5.3.1 Výsledky genotypizace.....	45
5.3.2 Rozdíly mezi podrodinami na základě zastoupení u testovaných kmenů .....	49
5.3.3 Možné využití insercí jako poddruhově diagnostických markerů .....	51
<b>6. Diskuse .....</b>	<b>52</b>
6.1 Srovnání podrodin z databáze RepBase a podrodin definovaných Quentinem (1989) .....	52
6.2 Rozdíly v počtech nalezených insercí mezi kmeny C57BL/6J a MSM/Ms.....	53
6.3 Zastoupení insercí u poddruhů <i>M. musculus</i> .....	54
6.4 Přítomnost insercí u <i>M. spretus</i> a <i>M. macedonicus</i> .....	55
6.5 Využití L1 a B1 retrotransposonů jako markerů v populační genetice .....	56

6.6 Zhodnocení aktivity jednotlivých podrodin B1 retrotransposonů .....	57
6.7 Zhodnocení aktivity jednotlivých podrodin L1 retrotransposonů .....	58
6.8 Porovnání aktivity mezi rodinami L1, B1 a B2 retrotransposonů .....	59
<b>7. Shrnutí .....</b>	<b>62</b>
<b>Přehled literatury .....</b>	<b>63</b>
<b>Přílohy.....</b>	<b>71</b>
Příloha 1: Originální záznamy B1 podrodin z RepBase .....	71
Příloha 2: Seznam nalezených insercí s koordinátami .....	76
Příloha 3: Sekvence navržených primerů .....	83
Příloha 4: Genotypizace B1 a L1 retrotransposonů .....	87
Příloha 5: Alignment sekvencí podrodiny Mm .....	94
Příloha 6: Alignment sekvencí podrodiny Mus2 .....	96
Příloha 7: Divergence od konsensu podrodin Mm a Mus2 .....	98
Příloha 8: Inserce vhodné jako poddruhově diagnostické markery .....	99

# ABSTRAKT

---

Diplomová práce je zaměřena na studium dvou recentně aktivních rodin retrotransponů v genomu myši domácí. Jsou to L1 a B1 retrotransposony. Cílem práce bylo najít polymorfní inserce vzniklé recentní aktivitou retrotransponů. Byly srovnány genomy dvou inbredních kmenů odvozené od dvou odlišných poddruhů myši domácí. Částečně osekvenovaný genom kmene MSM/Ms odvozený od *M. m. molossinus* byl srovnán s referenčním genomem C57BL/6J odvozeným převážně od *M. m. domesticus*. Nalezené inserce byly zařazeny do několika podrodin B1 a L1 retrotransponů. Následně byla pomocí PCR testována přítomnost/absence těchto insercí u všech tří poddruhů myši domácí a u dvou sesterských druhů (*M. spretus* a *M. macedonicus*). Jednotlivé rodiny se lišily na základě zastoupení u dvou sesterských druhů. Zástupci podrodin F2 a F3 patřící k L1 retrotransponům přetrvávali v polymorfním stavu v populaci myši domácí i přes předpokládaný konec aktivity. Na rozdíl od nich se však podrodiny B1 retrotransponů přes svoji aktivitu již dávno v minulosti zdají být aktivní i v současnosti. Na základě srovnání L1 a B1 se také ukázalo, že rodina L1 je v současnosti více aktivní než rodina B1.

The work focuses on two recently active retrotransposon families in the house mouse genome. They are L1 and B1 retrotransposons. The aim of the work was to find polymorphic retrotransposon insertions caused by their recent activity. Two genomes of mouse inbred strains derived from the different house mouse subspecies were compared. The BACends from MSM/Ms derived from *M. m. molossinus* were compared with the reference genome of C57BL/6J derived mostly from *M. m. domesticus*. The set of output insertions was classified into several subfamilies of B1 a L1 families. The presence/absence of these insertions was tested using PCR in all three house mouse subspecies and also in two sister species (*M. spretus* and *M. macedonicus*). The particular subfamilies differed with regard to presence in latter species. Despite the supposed lack of activity of older L1 families (F2 and F3) they persist in house mouse population as an ancestral polymorphism. Unlike L1 subfamilies, B1 subfamilies appear to be active in house mouse genome for longer period of time. Also the difference between the whole families L1 and B1 was observed. Thus, according to my data L1 family seems to be recently more active than B1 family.

# 1. ÚVOD

---

Je to už téměř 60 let, kdy Barbara McClintock objevila první transposabilní elementy v genomu kukuřice. Významnost tohoto objevu byla uznána až později, kdy se zjistilo, že se nejedná pouze o zajímavost pod čarou v učebnicích genetiky, nýbrž že jsou to právě transposabilní elementy, které tvoří značnou část všech eukaryotních genomů (Burt & Trivers 2006). Tento objev byl naprosto zásadní při objasnění problému nazvaného „C-value paradox“, který trápil genetiky řadu dekad. Totiž před objevením těchto repetitivních sekvencí se předpokládalo, že velikost genomu souvisí s komplexitou daného organismu. Tento předpoklad se ukázal být lichým vzhledem k faktu, že kolikrát nejjednodušší organismy měly ty největší genomy (Gregory 2005). Až v nedávné době si genetici spojili aktivitu transposabilních elementů s touto diskrepancí a následně celogenomové projekty tento předpoklad potvrdily. Přesto však nadále přetrvával názor, že se s největší pravděpodobností jedná o tzv. „junk DNA“, která v genomech nemá žádnou funkci a pouze se zde hromadí. Recentní studie v posledních několika letech však ukazují, že transposabilní elementy nejsou jen „junk DNA“, ale účastní se v jádře řady mechanismů a jsou tak spíše motorem evoluce, který je zodpovědný za dynamiku celého genomu. Tyto elementy se uplatňují například při modulaci exprese genů, nehomologní rekombinaci, inaktivaci X chromosomu, tvorbě telomer aj. (Hurst & Werren 2001). Zároveň jsou zdrojem *de novo* mutací a tedy původcem řady genetických poruch (Ostertag & Kazazian 2001).

Přestože se transposabilním elementům věnuje enormní množství prací, jen zlomek z nich se zaměřuje na detailnější klasifikaci těchto elementů a porovnání aktivity jednotlivých rodin a podrodin, do kterých jsou rozděleny. Tento detailní přístup může velmi pomoci k pochopení šíření těchto elementů v genomech a umožnit tak lépe pochopit jejich funkční význam a třeba i evoluci na organismální úrovni. V mé diplomové práci se proto zaměřuji na zhodnocení aktivity dvou rodin retrotransposonů, které se významně podílely na utváření myšího genomu a sehrály tak v evoluci tohoto organismu významnou roli.



# 2. LITERÁRNÍ PŘEHLED

---

## 2.1 RETROTRANSPOSONY

### 2.1.1 ZÁKLADNÍ DĚLENÍ MYŠÍCH RETROTRANSPOSONŮ

V průběhu evoluce se v genomech eukaryotních organismů nahromadily sekvenční elementy, které ke své propagaci v rámci genomů využívají RNA intermediát a reverzní transkriptázu, a jsou nazývány retroelementy (retrotransposony) (Deininger & Batzer 2002). Retroelementy se dělí do několika skupin majících odlišný evoluční původ. Autoři se víceméně shodují na základním dělení retroelementů na LTR (Long Terminal Repeats) retroelementy a non-LTR retroelementy.

LTR retroelementy pochází z klasických retrovirů a vznikly procesem tzv. endogenizace – tedy ztráty schopnosti propagovat se horizontálním přenosem (myšleno přenosem z jednoho jedince do druhého v rámci jedné generace). K tomu, aby došlo k endogenizaci je nutné, aby se jednak exogenní retrovirus dostal do pohlavních buněk a šířil se tak dál vertikálně, a zároveň je nutné, aby ztratil schopnost se šířit infekčně mezi jedinci v populaci. Nejsnáze se tohoto docílí ztrátou genů kódujících u běžných retrovirů obal. LTR elementy se tedy liší od klasických retrovirů absencí genu *env*, který kóduje obal nutný k šíření viru mimo buňku a následný vstup viru zpět do buňky jiného jedince (Tarlinton et al. 2006).

LTR retroelementy se od non-LTR retroelementů liší jednak přítomností LTR (Long Terminal Repeats), které vznikají na obou koncích elementů v průběhu retrotransposice a také přítomností několika (> 2) ORF (Open Reading Frames), které odpovídají jednotlivým genům u klasických retrovirů (Deininger & Batzer 2002). Problematika jemnějšího dělení a evolučního původu LTR retroelementů je značně komplikovaná a vzhledem k tomu, že se ve své diplomové práci budu zabývat výhradně v myším genomu aktivními non-LTR retroelementy, nebudu dále zabíhat do podrobností.

Non-LTR retroelementy lze rozdělit na dvě skupiny. Na základě toho, zdali si kódují vlastní enzymatický aparát sloužící k propagaci či jen využívají aparát jiných retroelementů, je lze rozdělit na retrotransposony autonomní a neautonomní (např. Deininger & Batzer 2002). Jak poznamenává Trivers & Burton (2003), nelze žádné retrotransposony považovat za zcela

autonomní, vzhledem k faktu, že i během retrotransposice autonomních elementů jsou využívány reparační mechanismy buňky, bez kterých by tento proces nemohl být dokončen.

Mezi autonomní non-LTR retroelementy lze zařadit LINE (Long Interspersed Nuclear Elements), s jednou aktivní skupinou v myším genomu (LINE1) a dvěma reliktními skupinami (LINE2, L3/CR12) (Waterson et al. 2002, Gibbs et al. 2004). LINE1 (L1) mají celkovou délkou 6-7 Kb, obsahují 2 ORF. ORF1 kóduje protein p40, jehož funkce je RNA vazebná. ORF2 pak kóduje samotnou endonukleázu a reverzní transkriptázu. Na 3' konci je pak poly-A konec (detailněji viz níže). Evoluční původ LINE elementů není znám (Deininger & Batzer 2002).

Jako neautonomní non-LTR retroelementy se označují SINE (Short Interspersed Nuclear Elements). V myším genomu bylo identifikováno 5 typů (B1, B2, B4/RSINE, ID a MIR/MIR3) (Waterson et al. 2002), z toho pouze 3 (B1, B2 a ID) jsou aktivní (Gibbs et al. 2004). Na základě faktu, že 3' konec některých SINE je *de facto* identický 3' konci L1 elementů, je široce přijímána hypotéza, že SINE využívají ke své amplifikaci reverzní transkriptázu kódovanou L1 elementy (detailněji viz níže) (Deininger & Batzer 2002).

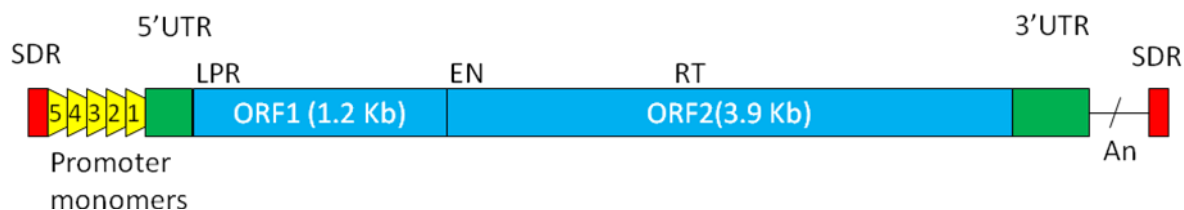
## 2.1.2 ZÁKLADNÍ CHARAKTERISTIKA AKTIVNÍCH RETROTRANSPONŮ

Ve své diplomové práci se zaměřím na 3 skupiny retroelementů (L1, B1, B2), které jsou považovány v myším genomu v současnosti za nejaktivnější non-LTR retrotransposony (Waterson et al. 2002, Gibbs et al. 2004).

### 2.1.2.1 L1

LINE1 (L1) jak už bylo zmíněno, patří mezi autonomní non-LTR retrotransposony. Kompletní retrotransposičně aktivní L1 mají délku 6-7 Kb (Voliva et al. 1983). Většina L1 je však v genomu z 5' konce zkrácena během procesu retrotransposice (Voliva et al. 1983, Pavlíček et al. 2002). L1 jsou tvořeny několika regiony, z toho 2 jsou kódující – ORF1 a ORF2. Délka ORF1 je zhruba 1,2 Kb, délka ORF2 je pak asi 3,9 Kb (Mears & Hutchison 2001). ORF1 kóduje protein p40, u kterého byla zjištěna RNA vazebná aktivita, která byla poprvé prokázána Hohjoh & Singer (1997). ORF2 kóduje 2 hlavní domény. První doména má endonukleázovou aktivitu (EN) (Feng et al. 1996), druhá doména má aktivitu reverzně transkriptázovou (RT) (Mathias et al. 1991). ORF1 a ORF2 jsou obklopeny 3' a 5' UTR (UnTranslated Region). Před 3' UTR se nachází několik promoter monomerů (viz dále). Za 5' UTR pak následuje poly-A konec. Celý L1 retroelement je pak v genomu ohraničen tzv. SDR (Short Direct Repeats), které vznikají

v průběhu retrotransposice (Mears & Hutchison 2001). Celá struktura je znázorněna na obrázku (obrázek 1).



**Obrázek 1:** Schéma L1 retrotransposonu (podle Mears & Hutchison 2001).

### 2.1.2.2 B1 a B2

SINE jsou zastoupeny v genomu myši, na rozdíl od genomu člověka (jen Alu), několika aktivními skupinami (B1, B2, ID) (Waterson et al. 2002). Tyto elementy na rozdíl od L1 nekódují žádné proteiny, jsou velmi krátké – obvykle mají délku 90 – 300 bp (Deininger & Batzer, 2002). Všechny 3 typy jsou závislé na aparátu L1 retrotransposonů, přestože nejsou sekvenčně podobné (Ohshima et al. 1996, Smit 1996). Evolučně jsou SINE elementy odvozeny od RNA genů – konkrétně od dvou typů: tRNA a 7SL RNA genů. Vznikly fúzí 3' konce LINE s jedním z výše zmíněných typů RNA genů. Stejně jako lidské Alu, i myší B1 jako jediné vznikly z 7SL RNA (Rogers 1985). Prvním a nejvýraznějším krokem ke vzniku B1 byla delece 182 nt u 7SL RNA. Následně sérií menších delecí a duplikací v průběhu evoluce hlodavců získaly současné myší B1 délku 135 nt (Quentin 1994, Veniaminova et al. 2007b). Ostatní myší SINE byly odvozeny od tRNA. B2 a ID elementy jsou sekvenčně nejvíce podobné tRNA, přesto však oba vznikly nezávisle na sobě (Daniels & Deininger 1985). Stejně jako v případě L1 elementů obsahují i všechny SINE na 3' konci poly-A konec (Ryskov et al. 1985). Společnou vlastností B1 a B2 elementů je přítomnost RNA Pol III specifického promotoru (Duncan et al. 1981, Elder et al. 1981, Fuhrman et al. 1981). Vzhledem ke stejnému mechanismu propagace jako v případě L1 elementů (viz kapitola Propagace elementů), jsou na obou stranách sekvence také přítomny tzv. „Short Direct Repeats“ (SDR).

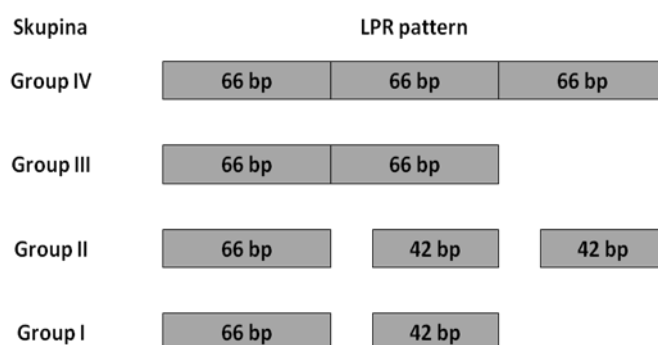
## 2.1.3 NOMENKLATURA AKTIVNÍCH RETROTRANSPONONŮ

### 2.1.3.1 L1

Nomenklatura L1 retroelementů v myším genomu je značně komplikovaná. První snahou o logické třídění elementů byla studie Smit et al. (1995), kde se autoři pokusili smysluplně roztrždit všechny do té doby známé L1 savčí retroelementy. Tento systém spíše než

evoluční vztahy mezi jednotlivými retroelementy bral v potaz liniovou specificitu, kde výchozím bodem byly L1 unikátní u člověka. L1 byly zařazeny do 2 skupin podle exkluzivní přítomnosti u dané taxonomické jednotky. Jednotlivé skupiny byly označeny písmeny P a M podle přítomnosti výhradně u primátů (L1P) a u všech savců (L1M). Dále pak byly elementy klasifikovány na základě struktury 3' UTR, která umožnila zařazení do jednotlivých skupin, které jsou označeny vzestupně písmenem abecedy. V názvu je toto písmeno na 4. místě (např. L1PB<sub>5</sub>). Další pozice je pak již vyhrazena číslu označujícím evoluční příbuznost s tím, že nejnižší čísla zaujímají retroelementy nejrecentněji odvozeny od v současnosti aktivních elementů (tedy např. L1PB<sub>5</sub>).

Tato nomenklatura je založena na strukturálních změnách v 3' UTR a následné divergenci ORF2 sekvence v rámci skupin definovaných 3' UTR. Mears & Hutchison (2001) se zaměřili na L1 výhradně u myši a vytvořili velmi jemný systém třídění těchto elementů založený na odlišnostech mezi promotorovými monomery a tzv. length polymorphic region (LPR) na začátku ORF1 (obrázek 1). U myších L1 jsou 3 typy transkripčních promotorů (A, F a V). Typ A a F je složen z několika tandemově uspořádaných sekvencí zvaných monomery a každý z těchto monomerů obsahuje zhruba 200 nukleotidů. Promotor typu V není v genomu myši velmi prozkoumán. Obecně platí, že L1 elementy mají vždy monomery pouze jednoho typu (Padgett et al. 1988). L1 elementy jsou dále klasifikovány do 4 skupin na základě struktury již zmíněného LPR v ORF1 (obrázek 2).



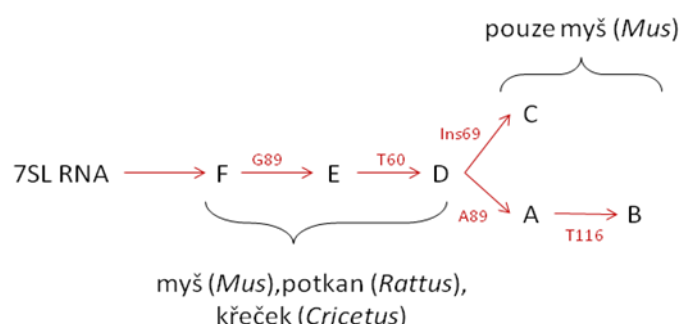
**Obrázek 2:** Struktura LPR v ORF1 L1 transposonu (podle Mears & Hutchison 2001)

Fylogenetické studie ukazují, že tyto 3 skupiny definované na základě promotorových monomerů odrážejí 3 po sobě jdoucí fáze v průběhu evoluce jedné linie L1 elementů v genomu myši. Výsledek analýzy ukázal, že skupina V je nejstarší, následuje skupina F a recentně pak

vznikla skupina A (V->F->A) (Adey et al. 1994). Ze všech 3 skupin je v současnosti aktivní pouze částečně skupina A (Adey et al. 1994). Kromě těchto 3 skupin byla ještě v genomu myši objevena velmi aktivní rekombinantní skupina s F promotorem a zbytkem sekvence skupiny A. Byla tedy pojmenována jako T<sub>F</sub> (transposable F) (Saxton & Martin 1998).

### 2.1.3.2 B1 a B2

Klasifikace B1 elementů je mnohem jednodušší než v případě L1 elementů. Základní dělení, které odráží evoluční vztahy v rámci této rodiny, bylo navrženo v práci Quentin (1989). Autor vytvořil 6 skupin (A, B, C, D, E, F), které jsou fylogeneticky různě vzdáleny od původní 7SL RNA a vzájemně se liší vždy jen jednou záměnou. Fylogenetické vztahy jsou znázorněny na obrázku (obrázek 3). V době, kdy byla tato klasifikace vytvořena, bylo známo pouze několik sekvencí B1. Tento systém tedy zdaleka nepokrývá všechny B1 elementy v myším genomu. Dělení podrodin v databázi RepBase (<http://www.girinst.org/>) je značně odlišné od Quentinova (1989) dělení. RepBase uvádí 10 B1 podrodin (F, F1, F2, Mur1, Mur2, Mur3, Mur4, Mus1, Mus2, Mm), u kterých není příliš zřejmé, kterým rodinám podle Quentina (1989) odpovídají. Mimo jiné i na rozdíly mezi těmito dvěma nomenklaturami se chci zaměřit v mé diplomové práci.



**Obrázek 3:** Fylogenetické vztahy podrodin B1 retrotransposonů (podle Quentina 1989)

Co se týče B2 elementů, je až s podivem, že doposud nebyla publikována studie zabývající se evoluční historií této skupiny. Přestože byly u myši tyto elementy identifikovány již v roce 1979 (Kramerov et al. 1979), žádný z autorů, i po osekvenování myšního genomu, se nezabýval evoluční historií v rámci této rodiny. Databáze RepBase (<http://www.girinst.org/>) B2 elementy v genomu myši dělí na 3 skupiny (Mm2, Mm1a, Mm1t), nicméně chybí jakákoliv publikace, která by zmiňovala evoluční vztahy těchto tří podrodin a přesně definovala mutace, kterými se jednotlivé podrodiny vzájemně liší.

## 2.1.4 POČET AKTIVNÍCH RETROTRANSPONŮ V MYŠÍM GENOMU

### 2.1.4.1 L1

Co se týče počtu L1 elementů v genomu myši, lze říci, že tato skupina je se 660 000 kopiemi v referenčním genomu nejvíc abundantní ze všech skupin retrotransponů. Vzhledem k faktu, že tato skupina patří se svými 6-7 Kb mezi delší retroelementy, pokrývá tudíž i největší frakci genomu (18,78 %) (shrnutí v tabulce 1) (Waterson et al. 2002). Přesto však jen zlomek z celkového počtu L1 dosahuje této délky. Před osekvenováním myšního genomu bylo odhadnuto, že v genomu musí být přítomno kolem 3000 potenciálně aktivních kopií (Goodier et al. 2001). Poté co byl však genom osekvenován, bylo zjištěno, že pouze 400 elementů má plnou délku a jen 12 z tohoto počtu je potenciálně aktivních (tzn. mají ORF1 i ORF2 nepoškozený) (Waterson et al. 2002). Po srovnání všech L1 z myšního genomu s L1 z lidského genomu lze říci, že většina L1 v současnosti přítomná u myši (88 % z celkového počtu L1) je specifická v linii vedoucí k myši, popř. u myši samotné (Waterson et al. 2002).

**Tabulka 1:** Počty retrotransponů v genomu myši. Data z MGSC (2002).

	Tisíce kopií	Počet pokrytých bazí (Mb)	Frakce genomu (%)	Liniově specifických ve srovnání s lidským genomem (%)
<b>LINEs</b>	660	475.3	19.2	16.46
<b>LINE1</b>	<b>599</b>	<b>464.8</b>	<b>18.78</b>	<b>16.46</b>
<b>LINE2</b>	53	9.4	0.38	—
<b>L3/CR1</b>	8	1.2	0.05	—
<b>SINEs</b>	1383	188.8	7.66	7.66
<b>B1 (Alu)</b>	<b>564</b>	<b>67.3</b>	<b>2.66</b>	<b>2.66</b>
<b>B2</b>	<b>348</b>	<b>59.6</b>	<b>2.39</b>	<b>2.39</b>
<b>B4/RSINE</b>	391	57.1	2.36	2.36
<b>ID</b>	79	5.3	0.25	0.25
<b>LTR elements</b>	631	244.3	9.87	8.72
<b>ERV_classI</b>	34	16.8	0.68	0.58
<b>ERV_classII</b>	127	79.1	3.14	3.14
<b>ERV_classIII</b>	37	14	0.58	0.32
<b>MaLRs(III)</b>	388	112.2	4.82	4.02

### 2.1.4.2 B1 a B2

Ze všech SINE v myším genomu, jsou B1 elementy nejpočetnější (564 000 kopií). Přesto však pokrývají pouze 2,66 % genomu. Tento rozpor ve srovnání s L1 elementy je dán faktem, že B1 a vůbec všechny SINE jsou mnohonásobně kratší nežli L1. B2 pak se svými 348 000 kopiemi pokrývají 2,63 % myšního genomu (shrnutí v tabulce 1) (Waterson et al. 2002). Co se

týče liniové specificity, všechny myší SINE se vyskytují výhradně v linii vedoucí k myši. V genomu člověka můžeme nalézt pouze jednu rodinu SINE, a tou jsou Alu elementy, které naopak není možné nalézt v genomu myši (HGSC 2001, Waterson et al. 2002).

## 2.2 PROPAGACE RETROTRANSPONŮ

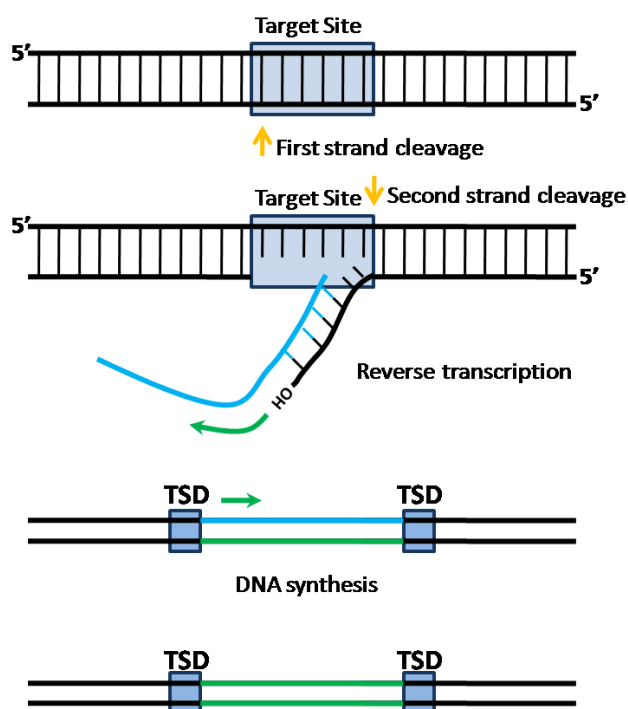
Retrotransposony, jak už jsem zmínil výše, si kódují vlastní enzymatický aparát, který jim slouží k propagaci v rámci genomu, popřípadě využívají enzymatický aparát jiných retroelementů. Velmi důležitým enzymem je reverzní transkriptáza (Mathias et al. 1991), která umožňuje přepsat mRNA retroelementu zpět do DNA. Retrotransposice LTR elementů (detailně popsána v Lewin 2004) se velmi liší od retrotransposice non-LTR elementů. Jak jsem již avizoval výše, ve své práci se zaměřuji na non-LTR elementy, proto popíši tento proces jen u LINE a SINE. Retrotransposice non-LTR elementů je stále velmi diskutované téma a ne všechny detaily jsou známy. Ostertag and Kazazian (2001) shrnuje velmi detailně tuto problematiku u L1 elementů, která však přesahuje rámec mé diplomové práce, proto zde popíši propagaci retrotransponů jen velmi stručně.

LINE i SINE obsahují tzv. vnitřní promotor, který je nezávislý na upstreamové sekvenci (Swergold 1990). Doposud není zřejmé, zdali jsou L1 elementy transkribovány pomocí RNA polymerázy typu III (Pol III) nebo typu II (Pol II). Kurose et al. (1995) prokázali, že L1 retroelementy mohou být skutečně transkribovány Pol III, další důkazy se však přiklání spíše k Pol II. Vzhledem k faktu, že L1 obsahují AATAAA polyadenylační (polyA) signál (např. Moran et al. 1999), který je nutný k terminaci Pol II (Osheim et al. 1999) a zároveň následná polyadenylace transkriptu zase naopak vyžaduje Pol II (Hirose and Manley 1998), je dost pravděpodobné, že L1 jsou transkribovány skutečně Pol II. Toto se podařilo experimentálně prokázat Mizrokhi et al. (1988) u elementů *jockey*, které jsou přítomny u *Drosophila melanogaster* a jsou velmi podobné savčím LINE. Přestože podobnou námitku by bylo možné vznést i v souvislosti se SINE, bylo již v 80. letech v několika studiích nezávisle na sobě prokázáno, že SINE využívají k transkripci Pol III (Duncan et al. 1981, Elder et al. 1981, Fuhrman et al. 1981).

Po dokončení transkripce mRNA putuje z jádra do cytoplazmy, kde dojde k translaci ORF1/ORF2. Následně se předpokládá, že LINE mRNA spolu s translatovanými ORF1 a ORF2 vytvoří ribonukleoproteinový komplex („RNP particle“). Tento komplex již byl experimentálně

prokázán (Martin 1991). Celý tento komplex pravděpodobně vstupuje do jádra, kde dojde k tzv. „Target Primed Reverse Transcription“ – zkráceně (TPRT).

TPRT byla prokázána u R2 elementů. Jedná se o non-LTR elementy, které se nachází u Arthropod (Luan et al. 1993). Podle Ostertag and Kazazian (2001) můžeme v případě L1 retroelementů najít několik náznaků, které podporují TPRT. Jednak jsou všechny L1 v genomu obklopeny na obou koncích velmi dobře zachovanými 7-20 bp dlouhými tzv. TSD (Target Site Duplication), dále původní nukleotidy, ve kterých došlo k inserci, jsou obvykle T bohaté, což umožňuje komplementaritu s polyA koncem retroelementu.



**Obrázek 4:** Schéma target site reverse transcription (TPRT)  
(podle Ostertag & Kazazian 2001)

Samotný proces probíhá následovně. Nejprve RNP vybere vhodné místo, které závisí na tom, zdali je endonukleáza místně specifická či nikoliv. Jak již bylo zmíněno výše, obecně je preferováno místo s vysokým zastoupením thyminů. V případě lidských L1 elementů bylo demonstrováno, že endonukleáza štěpí 5' – TT|AAAA – 3' (Moran et al. 1996). Na základě fylogenetické analýzy L1 endonukleáz u obratlovců bylo prokázáno, že sekvenční specificitou se vyznačují všechny retroelementy skupiny M a mladší. Endonukleázy savčích L1 retroelementů jsou tedy místně specifické, což vysvětluje i velký nárůst počtu elementů v



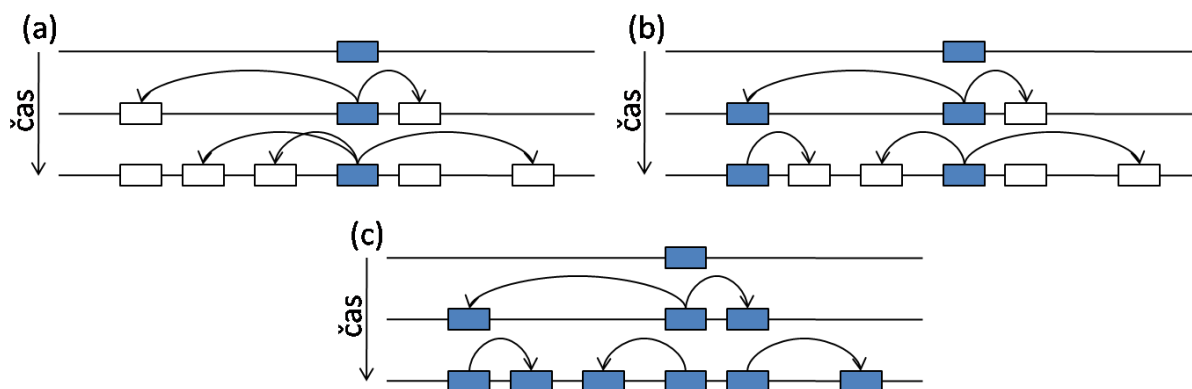
genomech savců, vzhledem k tomu, že motiv 5' – TT|AAAA – 3' se nachází především v oblastech na geny chudých (AT izochory) (Ichiyanagi et al. 2007).

L1 endonukleáza štěpí DNA vlákno tak, aby došlo k vytvoření kohezních konců, na které se může navázat poly-A konec. Poté, co se poly-A konec naváže na +vlákno, začíná reverzní transkriptáza vytvářet cDNA podle mRNA templátu na -vlákně tak, že první nukleotid naváže na 3'OH skupinu kohezního konce -vlákna. Poté, co reverzní transkriptáza dokončí cDNA, dojde k odbourání templátové mRNA a její nahrazení DNA. Celý proces znázorněn na obrázku (obrázek 4).

## 2.3 MODELY ŠÍŘENÍ RETROTRANSPOSONŮ

Vzhledem k charakteru propagace retroelementů vyvstala otázka, jakým způsobem se v genomech organismů tyto elementy šíří. V minulosti byly navrženy dva modely šíření retroelementů (Deininger et al. 1992). Jeden z nich (tzv. transposonový model) předpokládá, že každá nová kopie může být zdrojem dalších kopií. U tohoto modelu se předpokládá exponenciální nárůst počtu kopií v genomu. Zatímco model druhý (tzv. „master genes“ model), předpokládá, že v genomu je pouze pár zdrojových elementů, které produkují kopie, které se již dál nešíří z různých důvodů. Tyto kopie mohou být poškozeny (neobsahují celou sekvenci nebo obsahují mutace) nebo mohou být vloženy do oblastí, které nepodléhají transkripci. V tomto případě se předpokládá lineární nárůst s tím, že sklon je určen schopností dané podrodiny se amplifikovat.

Několik studií z poslední doby provedených na lidských Alu elementech ukázala, že retroelementy v genomech se šíří spíše „master genes“ modelem než transposonovým modelem (Cordaux et al. 2004, Han et al. 2005). Otázkou ale stále zůstalo, zdali jednotlivé skupiny retroelementů mají jen jeden zdroj nebo zdrojů několik (Cordaux et al. 2004). Zde autoři v podstatě rozdělili „master genes“ model na „single master gene“ model a „intermediate“ model, kde v prvním případě, je aktivní vždy jen jediný element, zatímco v případě druhém je několik zdrojových elementů (obrázek 5).

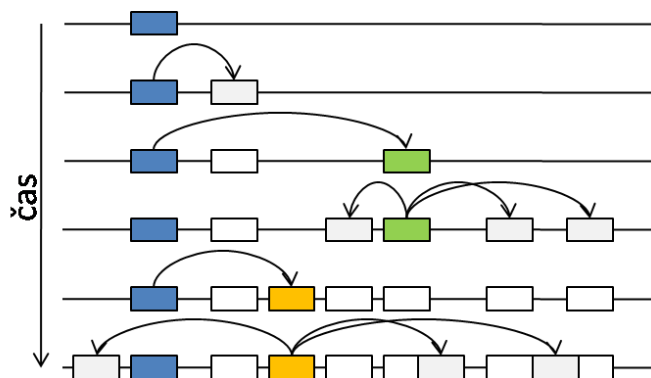


**Obrázek 5:** Modely šíření retrotransponů v genome. a) Single master gene model, b) intermediate model, c) transposonový model (podle Cordaux et al. 2004).

Předpokladem „master gene“ modelu (ať už „single master gene“ modelu nebo „intermediate“ modelu) je, že jednotlivé podrodiny jsou aktivní po určitou dobu a během této doby vytvoří velké množství kopií. Kopie vytváří, dokud zdrojová kopie nezmutuje, popř. nedojde k potlačení hostitelskými mechanismy. Nicméně například v případě podrodiny AluYb tomu tak není. Tato podrodina má několik svých amplifikačních maxim spolu s obdobími relativního útlumu. Aby lépe vystihli strategii této podrodiny, použili Han et al. (2005) název „stealth“ model. Autoři se inspirovali stejnojmennou technologií, která umožňuje letounům či plavidlům uniknout radaru. Tedy v případě retroelementů se jedná o fakt, že zdrojová kopie produkuje jen velmi malé množství dceřiných kopií po velmi dlouhý čas a v podstatě tak uniká jakékoliv pozornosti selekce na hostitelskou buňku. Některé dceřiné kopie však při vhodných podmínkách způsobí tzv. vlnu amplifikace, kdy se počet kopií v genomu zvětší o tisíce, po této vlně následuje zase dlouhé období, kdy dceřiné kopie produkuje pouze jedna zdrojová kopie ve velmi malém množství. Celý proces je znázorněn na obrázku (obrázek 6).

U hlodavců se podařilo identifikovat BC1 RNA jako master gen pro skupinu ID elementů (Kim et al. 1994). Autoři zjistili, že změny v sekvenci BC1 RNA u různých hlodavců v různých dobách korespondují se sekvencemi ID elementů v té době a u dané skupiny aktivních. Zároveň se taky podařilo prokázat, že gen BC1 RNA je exprimován u všech studovaných skupin hlodavců. V následující studii Kass et al. (1995) bylo zjištěno pomocí dot-blot, že počty kopií v genomech jednotlivých druhů hlodavců se liší bez ohledu na evoluční vztahy hlodavců. Tedy amplifikace elementů je liniově specifická. V době kdy bylo toto zjištění publikováno, doposud nebyl znám výše zmíněný „stealth“ model, ale právě tento zjištěný pattern u ID elementů velmi dobře koresponduje s tímto modelem. Kass et al. (1995) vysvětlují tuto odlišnou aktivitu u

různých hlodavčích linií, i přes stejnou míru exprese BC1 RNA, právě mechanismem, který označuje Han et al. (2005) jako „stealth“ – BC1 RNA občas vytvoří kopii, popř. dojde k duplikaci celé oblasti, a tak vzniká nový zdroj, který velmi snadno vysvětlí značně odlišné počty ID elementů i v případě evolučně velmi příbuzných druhů.



**Obrázek 6:** Schématické znázornění stealth modelu.

Tato problematika je důležitá i z důvodu vzniku nových podrodin v rámci jednotlivých skupin (rodin) retroelementů. Deininger et al. (1992) se tímto tématem zabývali v souvislosti s „mater genes“ modelem. Podrodina je v podstatě sekvenční varianta, kterou je možno charakterizovat specifickými mutacemi, a která je také svázána s určitým obdobím, během něhož byla v genomu amplifikována (aktivní). Nejjednodušším modelem evoluce jednotlivých podrodin, který zmiňuje Deininger et al. (1992), je postupný vznik jednotlivých podrodin z podrodin, které jim evolučně předcházely. Takovýto model byl v minulosti běžně akceptován. Je však velmi důležité rozlišovat, na jaké časové škále jsou retroelementy studovány. Doposud se většina analýz týkala kvantifikace počtu kopií *per* genom u jednotlivých hlodavců pomocí dot-blot techniky (Kass et al. 1997, Veniaminova et al. 2007c), popř. srovnání počtu jednotlivých rodin elementů mezi dvěma genomy (Waterson et al. 2002, Gibbs et al. 2004). Tyto studie byly *a priori* zaměřené na porovnání aktivity jednotlivých skupin (rodin) elementů mezi všemi hlodavci spíše než na porovnání aktivity jednotlivých podrodin u několika velmi blízké příbuzných druhů. Až na dvě práce, které se však obě týkají lidských Alu elementů (Cordaux et al. 2004, Han et al. 2005), nebyla doposud detailněji rozpracována problematika aktivity a modelů šíření jednotlivých podrodin retroelementů pomocí celogenomické analýzy sekvencí jednotlivých retrotransposonů.

## 2.4 AKTIVITA RETROTRANSPONŮ V GENOMU MYŠI DOMÁČÍ

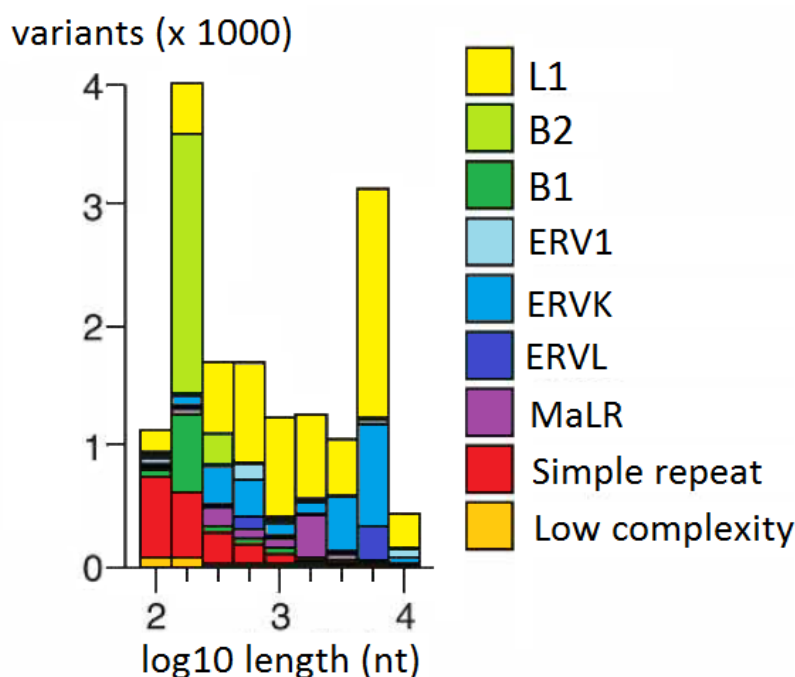
### 2.4.1 SINE

Celogenomové analýzy ukázaly, že v genomu myši domácí jsou přítomny 4 skupiny SINE (viz tabulka 1) (Waterson et al. 2002). Z toho díky výsledkům porovnání genomů myši a potkana víme, že v linii vedoucí k myši byly aktivní pouze 3 z nich (Waterson et al. 2002, Gibbs et al. 2004). Porovnání aktivity u jednotlivých druhů je možné také pomocí techniky dot-blot, která umožní kvantifikovat ne zcela přesně počet kopií v genomu na základě počtu hybridizovaných prób (Kass et al. 1997, Veniaminova et al. 2007c). Vzhledem ke špatné sensitivitě této metody lze porovnávat rozdíly v aktivitě mezi jednotlivými skupinami (rodinami) elementů spíše než odhalovat rozdíly v aktivitě mezi jednotlivými podrodinami. Tedy studií zabývajících se rozdíly mezi aktivitou jednotlivých podrodin je poskrovnu, jako příklad lze uvést práci Kass et al. (1997).

Přesto i porovnání aktivity jednotlivých rodin elementů mezi hlodavci může leccos vypovědět o aktivitě těchto elementů v genomu myši. Z kvantifikačních analýz jasně vyplývá, že v linii vedoucí od společného předka potkana (*Rattus norvegicus*) a myši domácí (*Mus musculus*) k myši byly jednotlivé rodiny různě aktivní. V případě B2 elementů lze vidět výrazné zvýšení jejich aktivity v Muroidní linii (*Cricetidae*, *Muridae*) ve srovnání se zástupci Sciuroidní a Hystricognathní linie (Kass et al. 1997). U vybraných druhů Muroidea tuto autoři pomocí dot-blot analýzy odhadli konstantní počet B2 elementů na 80 000 kopií. V dnešní době již v případě potkana a myši díky sekvenačním projektům známe přesné počty u obou druhů (myš: 348 000, potkan: 328 000) (Waterson et al. 2002, Gibbs et al. 2004). Lze vidět, že výsledky dot-blot analýzy jsou silně podhodnoceny, přesto nám však dávají určitou informaci, díky které je možné porovnat počet elementů mezi jednotlivými druhy. Kass et al. (1997) odhadli u *Sciurus carolinensis* 2500 elementů, u obou zástupců Hystricognathi (*Myocaster coypus*, *Cavia porcellus*) pak méně než 100 kopií na celý genom. Kass et al. (1997) vycházeli ze staré fylogeneze hlodavců (Sarich 1985), kde Hystricognathi tvoří sesterskou skupinu skupině Sciurognathi, která zahrnuje jak Sciuroidní linii, tak i linii Muroidní. Autoři se tedy domnívali, že u hlodavců docházelo k postupnému nárůstu aktivity B2 elementů směrem k myši domácí (*Mus musculus*). U skupiny Muroidea, jak uvádí Kass et al. (1997), pak došlo k velmi výraznému zvýšení jejich aktivity, která je v podstatě typická pro všechny zkoumané druhy. Vyplývá to i z analýzy genomů myši a potkana, kde jsou *de facto* počty velmi podobné (Waterson et al.

2002, Gibbs et al. 2004). Výše zmíněné schéma je však v rozporu s výsledky nejnovějších fylogenetických analýz skupiny Rodentia založených jak na sekvenční divergenci dvou nukleárních genů (Adkins 2003), tak i na přítomnosti/absenci fylogeneticky významných SINE insercí (Veniaminova et al. 2007a, Churakov et al. 2010). Obě analýzy se shodují na faktu, že Sciurognathní linie je bazální skupinou hlodavců a Hystricognathní a Muroidní linie pak tvoří monofylum. To znamená, že v případě aktivity B2 elementů se nejedná o vzestupný trend v jedné linii hlodavců, ale spíše o různou aktivitu u různých skupin hlodavců. Přesto však jejich aktivita u jednotlivých skupin nevaríruje tolik, jako v případě jiných rodin elementů (viz dále).

Přejdeme-li na úroveň podrodin, můžeme zmínit práci Stehlíková (2009). Z této analýzy na základě přítomností polymorfních insercí mezi dvěma poddruhy myši domácí jednoznačně vyplývá, že v současných populacích *Mus musculus* jsou B2 elementy velmi aktivní. Mezi recentně aktivními elementy převažují B2 patřící do podrodiny Mm1a. Zatímco podrodiny Mm1t a Mm2 jsou zastoupeny velmi zřídka. Recentní velmi výrazná aktivita B2 elementů byla potvrzena i díky celogenomovému srovnání délkových polymorfismů mezi 5 klasickými laboratorními kmeny, kde B2 elementy, co se početnosti týče, zaujímaly hned druhé místo za L1 elementy (Akagi et al. 2008) (viz obrázek 7).



**Obrázek 7:** Zastoupení retrotransposonů detekovaných srovnáním několika genomů laboratorních inbredních kmenů odvozených od myši domácí (z Akagi et al. 2008).

V případě B1 elementů, na které se mj. zaměřím v mé diplomové práci, nelze mluvit v rámci skupiny Rodentia o žádném jednoznačném trendu, který by bylo možno spatřit na první pohled. Ze studie Veniaminova et al. (2007c) vyplynulo, že u jednotlivých druhů se počet kopií na genom značně liší i v rámci jednotlivých čeledí. Veniaminova et al. (2007c) použili stejnou techniku jako Kass et al. (1997). Díky znalosti přesných počtů B1 elementů v genomu myši a ID elementů v genomu potkana, však byli schopni velmi přesně kalibrovat počet elementů v genomech. Pokud by skutečně nebyl žádný trend a u jednotlivých druhů by počet kopií varíroval od několika tisíc až po stovky tisíc, ukazovalo by to na velmi recentní paralelní aktivitu těchto elementů u jednotlivých studovaných druhů, jednotlivé druhy by tedy navzájem sdílely jen velmi málo elementů, což je velmi málo pravděpodobné. Toto je případ skupiny Hystricognathi, kde i u velmi příbuzných druhů se počty kopií liší o statisíce s tím, že některé druhy mají jen několik tisíc těchto elementů (*Hystricidae*, *Octodontidae*), zatímco jiné druhy mají až několik stovek tisíc kopií (*Chinchillidae*). Přestože i ve skupině Muroidea jsou znát značné rozdíly v počtu kopií těchto elementů, lze sledovat jejich jasný nárůst. To lze vidět na příkladu potkana (*Rattus norvegicus*), který ze všech studovaných zástupců skupiny Muroidea má kopií nejméně, přesto jejich počet dosahuje 384 000 (RGSPR 2005, Veniaminova et al. 2007c). V genomu myši je však tento počet téměř dvojnásobný (Waterson et al. 2002, Veniaminova et al. 2007c). V případě celé Sciurognathní linie jsou počty nízké a rozdíly mezi jednotlivými druhy minimální. Lze tedy říci, že stejně jako v případě B2 elementů, došlo u skupiny Muroidea ke značnému nárůstu aktivity B1 elementů.

Detailnější analýzu zabývající se aktivitou jednotlivých podrodin B1 elementů provedli dvě studie (Quentin 1989, Kass et al. 2000). Quentin (1989) na základě znalosti 47 sekvencí B1 dostupných v té době v databázi GenBank, definoval 6 podrodin (A, B, C, D, E, F) a určil specifická místa, kterými se jednotlivé podrodiny liší. Schéma evoluce B1 podrodin je znázorněno na obrázku (obrázek 3). Na základě divergence sekvencí v rámci jednotlivých podrodin byl autor schopen datovat stáří těchto podrodin a určit tak, u kterých druhů je možné je nalézt. Podrodiny F, E a D byly přítomny u společného předka *Cricetidae* a *Muridae*. Další 3 podrodiny A, B a C jsou přítomny jen v genomu myši.

Kass et al. (2000) ověřoval aktivitu jednotlivých podrodin definovaných ve studii Quentin (1989) u několika zástupců čeledi *Muridae* (*M. musculus musculus*, *M. m. domesticus*, *M. m. castaneus*, *M. spretus*, *M. hortulanus*, *M. caroli*, *M. pahari*) tak, aby bylo zachyceno co nejvíce variability mezi aktivitou jednotlivých podrodin B1 elementů. Autoři testovali celkem 10

elementů patřících do 4 podrodin – A(2), B(4), C(2) a D(2). Díky znalosti stáří jednotlivých druhů myši (založeno na Boursot et al. 1993), byli schopni víceméně určit, kdy k inserci daného elementu došlo. Analýza ukázala, že rodina B, která je považována Quentinem (1989) za nejmladší, byla aktivní už u společného předka druhů *M. musculus*, *M. hortulanus* a *M. spretus* – tedy mezi 1-3 miliony let. Stáří nejmladších insercí je pak odhadováno na méně než 0,5 milionu let. Stejně schéma *de facto* ukazuje podrodina A. Aktivita podrodiny C je na základě 2 insercí autory odhadována na 1-6 milionu let s přítomností u 3 výše zmiňovaných druhů a navíc druhu *M. caroli*. Podobné schéma lze vidět i v případě poslední podrodiny D, kde aktivita jde více do minulosti a inserci lze nalézt i u druhu *M. pahari* (1-? milionu let). Vzhledem k faktu, že se aktivity v období 1-3 milionu překrývají, autoři naznačují přítomnost více tzv. „master“ genů, které slouží jako zdrojové kopie pro jednotlivé podrodiny nezávisle na sobě.

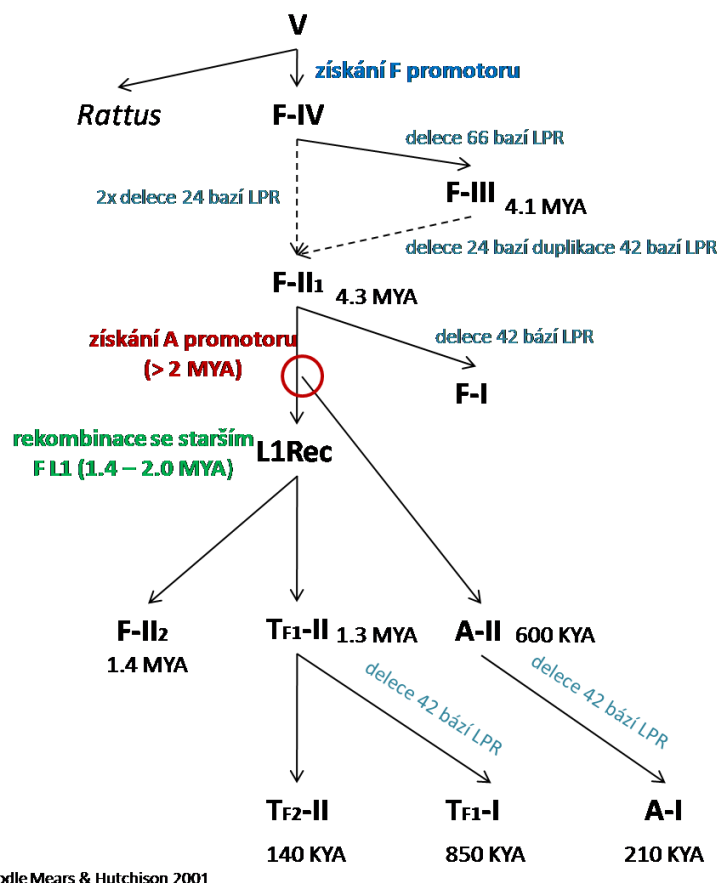
Poté, co byl osekvenován myší genom (Waterson et al. 2002), bylo zveřejněno v databázi RepBase (<http://www.girinst.org/>) 10 podrodin B1 elementů, které, jak už jsem uvedl v úvodní kapitole, nikterak neodpovídají Quentinem (1989) definovaným podrodinám. Lze tedy očekávat, že problematika aktivity jednotlivých podrodin bude ještě komplexnější, než ji uvedli Quentin (1989) a Kass et al. (2000).

#### 2.4.2 L1

Z hlediska porovnávání aktivity jednotlivých podrodin jsou L1 elementy mnohem lépe prozkoumány než výše zmiňované SINE. Důvodem může být fakt, že jednotlivé podrodiny L1, jak už jsem uváděl v první kapitole, jsou velmi dobře definované na základě strukturálních změn 5' UTR a LPR regionu na začátku ORF1 (Mears & Hutchison 2001). Také z biomedicínského hlediska je studium L1 elementů velmi důležité, vzhledem k tomu, že L1 kódují reverzní transkriptázu, a jsou tak zdrojem řady *de novo* mutací (Ostertag & Kazazian 2001). Na rozdíl od SINE, byla také L1 aktivita porovnávána spíše na menší škále u několika příbuzných druhů, nežli u velké skupiny jako jsou Rodentia (Pascale et al. 1990, Casavant & Hardies 1994, Saxton & Martin 1998, Rebuzzini et al. 2009).

Na základě přítomnosti velkého množství recentních kopií L1 v genomu myši lze říci, že celková aktivita těchto elementů je v Muroidní linii velmi vysoká (Pascale et al. 1990, Casavant & Hardies 1994, Saxton & Martin 1998, Waterson et al. 2002, Gibbs et al. 2004, Akagi et al. 2008). Z dlouhodobějšího hlediska probíhala evoluce jednotlivých podrodin postupným vznikem podrodiny novější z podrodiny předcházející (Adey et al. 1994). Autoři zjistili, že

v linii vedoucí od společného předka myši a potkana směrem k myši vznikly postupně 3 podrodiny L1 (A, F, V). Rozdělení do jednotlivých podrodin je dáno typem promotoru. U společného předka potkana (před 15 miliony let (Boursot et al. 1993)) byl přítomen promotor typu V, po odštěpení linie vedoucí k myši od linie vedoucí k potkanovi byl získán promotor typu F.



**Obrázek 8:** Schéma evoluce jednotlivých podrodin L1 transposonů (podle Mears & Hutchison 2000).

Z tohoto typu pak před více jak 2 miliony let vznikl promotor typu A. L1 mající promotor typu A jsou dnes v genomu myši považovány za aktivní. Saxton & Martin (1998) zjistili, že kromě L1 elementů podrodiny A jsou přítomny recentní inserce i elementů, které nelze zařadit do žádné z výše uvedených podrodin. Následný výzkum ukázal, že se jedná o novou podroduinu, která vznikla rekombinací staré, neaktivní F podrodiny a aktivní podrodiny A. Autoři tuto podroduinu pak nazvali T s indexem F ( $T_F$ ) jako „Transposable F“. Na základě kombinace variant promotorových monomerů a již zmiňovaného LPR na 5' konci ORF1, bylo možno rozdělit L1 podrodiny mnohem jemněji (Mears & Hutchison 2001). V současné době jsou tedy aktivní 3 podrodiny (viz obrázek). Dvě podrodiny jsou odvozeny od rekombinační podrodiny  $T_F$ . Jedná



se o  $T_{F2-II}$  a  $T_{F1-I}$ . Předpokládaná doba vzniku těchto dvou podrodin je před 140 000 lety a 850 000 lety. Třetí aktivní podrodina (A-I) obsahuje promotor monomery typu A a její vznik se odhaduje na dobu před 210 000 let (viz obrázek 8).

Vysoká aktivita L1 v genomu myši v současnosti stále přetrvává. Akagi et al. (2008) provedli srovnání CNV (Copy Number Variants) mezi pěti myšími laboratorními kmeny, ze kterého vyplynulo, že největší podíl na inserčních/delečních polymorfismech mají právě L1 elementy (obrázek 7). Studie hodnotící kvantitativní variabilitu L1 u 5 blízce příbuzných druhů myši včetně *M. musculus* zjistila, že směrem k *M. musculus* (zvláště pak poddruhu *M. m. domesticus*) došlo k velkému nárůstu aktivity (Rebuzzini et al. 2009). Ani jedna z dvou výše uvedených studií však nezjišťovala zařazení L1 do konkrétních podrodin. Z tohoto pohledu bude tedy zajímavé zjistit, které z podrodin jsou v populaci myši domácí zastoupeny v polymorfním a tedy aktivním stavu.

## 2.5 VYUŽITÍ RETROTRANSPOSONŮ V POPULAČNÍ GENETICE

Retrotransposony vzhledem k jejich povaze šíření v genomech organismů představují velmi ideální markery pro užití v populační genetice (Shedlock & Okada 2000). Ve srovnání s běžně užívanými typy markerů (mikrosatelity, RFLP, SNP) mají několik výhod. Jednou z nich je, že v případě retrotransposonů je znám počáteční stav (Batzer et al. 1994), kterým je absence inserce. Druhou nespornou výhodou ve srovnání s běžnými typy markerů je, že v případě retrotransposonů je velmi malá pravděpodobnost vzniku homoplázie (Batzer & Deininger 1991). Již tyto dvě vlastnosti předurčují retrotransposony k tomu, aby byly velmi vhodnými markery. Přihlédneme-li ještě k faktu, že genotypy lze velmi jednoduše získat pouze z výsledku elektroforézy bez nutnosti sekvenace (Munclinger et al. 2003), lze je považovat za téměř ideální markery.

O jejich významnosti jako populačně genetických markerů svědčí do dnešní doby velké množství publikovaných studií. Drtivá většina těchto studií je zacílena na lidskou populační genetiku. Významnou roli tyto markery sehrály při potvrzení hypotézy „Out of Africa“ v případě lidské populace (Batzer et al. 1994, Sherry et al. 1997, Stoneking et al. 1997). Následovaly pak další studie zabývající se lidskou populační strukturou (Watkins et al. 2001, Bamshad et al. 2003, Watkins et al. 2003, Whitterspoon et al. 2006).

Přesto však již byly publikovány studie, které přinejmenším snižují význam retrotransposonů jako ideálních markerů. V této souvislosti je nutné zmínit studii Ray et al. (2006), která byla zacílena na diskrepance mezi přítomností insercí a fylogenezí daných blízce příbuzných druhů. Ukázalo se, že tyto diskrepance mohou mít několik příčin. Jedná se o hlavně o přetrvávání ancestrálního polymorfismu vlivem nekompletního lineage sortingu, který zmiňuje již Shedlock et al. (2004). Dále pak jsou to: paralelní inserce, excise inserce a paralogní inserce. Ve zkratce se v případě paralelní inserce jedná o vložení retrotransposonu na stejné místo v genomu u dvou jedinců/druhů nezávisle na sobě. V případě excise inserce dojde k delecii retrotransposonu pomocí TSD (Target Site Duplication). Poslední možností je fakt, že některé oblasti jsou v genomu duplikovány a pokud dojde k vložení retrotransposonu jen do jednoho místa, pak se výsledný genotyp jeví jako heterozygotní. Detailně tuto problematiku shrnula Stehlíková (2009).

### 3. CÍLE PRÁCE

---

1. Najít polymorfní inserce, které vznikly recentní aktivitou retrotransposonů v genomu myši domácí
2. Porovnat aktivitu jednotlivých podrodin v rámci rodin B1 a L1 retrotransposonů
3. Porovnat aktivitu mezi rodinami L1 a B1 a také mezi rodinou B2, která byla testována Stehlíkovou (2009)
4. Určit inserce vhodné jako marker pro využití v populační genetice myši domácí

## 4. MATERIÁL A METODIKA

---

### 4.1 ANALÝZA KONSENSUÁLNÍCH SEKVENCÍ B1 PODRODIN

Jedním z cílů mé diplomové práce bylo porovnat rozdíly mezi dvěma klasifikačními systémy podrodin B1 retrotransposonů. Původní systém definovaný Quentinem (1989) byl porovnáván s novým systémem používaným databází RepBase (<http://www.girinst.org/>). Konsensuální sekvence jednotlivých podrodin byly staženy z databáze RepBase (zdrojový soubor je uveden v příloze 1) a z článku Quentin (1989). Pomocí programu BioEdit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) byly sekvence alignovány a poté manuálně hledány diagnostické mutace u podrodiny pocházejících z databáze RepBase. Tyto sekvence byly následně porovnávány s konsensuálními sekvencemi definovanými Quentinem (1989).

K rekonstrukci fylogeneze byly použity metody Maximum Parsimony v programu MEGA4 (<http://www.megasoftware.net/>) a Median-Joining Networks (Bandelt et al. 1999) v programu Network (<http://www.fluxus-engineering.com/index.htm>).

### 4.2 SEKVENČNÍ ANALÝZA RECENTNĚ AKTIVNÍCH L1 A B1 RETROTRANSPOSONŮ

#### 4.2.1 ZÍSKÁNÍ POTENCIÁLNĚ POLYMORFNÍCH B1 A L1 RETROTRANSPOSONŮ

K získání polymorfních insercí je nejvhodnější srovnání sekvencí alespoň dvou genomů. K porovnání jsme použili referenční genom myši domácí kmene C57BL/6J (Mouse Genome Sequencing Consortium 2002), který je z větší části odvozen od poddruhu *M. m. domesticus* (Yang et al. 2007, Frazer et al. 2008). Jako druhý jsme použili částečně osekvenovaný genom kmene MSM/Ms (Abe et al. 2004). MSM/Ms je odvozen od japonských myší označovaných jako *M. m. molossinus*, které vznikly přirozenou hybridizací poddruhů *M. m. musculus* a *M. m. castaneus* (Boursot 1993). Sekvence MSM/Ms byla zveřejněna ve formě tzv. BACendů (hrubý výstup sekvenační procedury), se kterými jsme pracovali (Abe et al. 2004). BACendy jsme použitím programu BLAT na serveru UCSC (celogenomová databáze) (<http://genome.ucsc.edu/>) alignovali na referenční genom C57BL/6J. Pomocí této strategie se

nám podařilo identifikovat indely (inserce/delece), kterými se mezi sebou liší tyto dva genomy (C57BL/6J vs. MSM/Ms). Koordináty indelů v referenčním genomu (C57BL/6J) byly staženy pomocí webového rozhraní programu BLAT. Z důvodu přesnosti byly sekvence staženy pomocí bioinformatického modulu BioPerl (<http://www.bioperl.org/>) z lokální genomové databáze C57BL/6J (stažena z NCBI ([ftp://ftp.ncbi.nih.gov/genomes/M\\_musculus/](ftp://ftp.ncbi.nih.gov/genomes/M_musculus/))) a MSM/Ms (na vyžádání zaslány autory (Abe et al. 2004)). Charakterizaci indelů jsme provedli pomocí webového rozhraní programu RepeatMasker ([http://www.repeat\\_masker.org/](http://www.repeat_masker.org/)), který nám umožnil verifikovat, zdali se jedná o repetitivní element popř. zařadit jej do rodiny/podrodiny. Jednotlivé inserce byly zařazeny programem RepeatMasker do několika rodin. Následně jsme se rozhodli charakterizovat 3 nejpočetnější a recentně aktivní skupiny retroelementů – L1, B1, B2. B2 elementy testovala a výsledky zveřejnila Jaroslava Stehlíková ve své diplomové práci v r. 2009. Ve své diplomové práci jsem se zaměřil na testování L1 a B1 elementů.

#### 4.2.2 ZÍSKÁNÍ POLYMORFNÍCH RECENTNĚ AKTIVNÍCH L1 A B1 RETROTRANSPONŮ

Postup popisovaný výše poskytl pouze hrubé výsledky. Po vizuální inspekci bylo zřejmé, že řada insercí nebyla způsobena recentní aktivitou retrotransponů výše zmiňovaných rodin. Velmi často se jednalo o případ, kdy retrotranspon samotný tvořil jen část inserce. Zbylá část byla tvořena sekvencí, u níž nebylo možné určit původ, popřípadě byla daná inserce tvořena více retrotransposony. V těchto případech s největší pravděpodobností inserce nevznikly aktivitou retrotransponu samotného (retrotransposicí), ale spíše procesem nehomologní rekombinace. Vzhledem k faktu, že cílem mé diplomové práce nebylo nalézt pouze polymorfismy, ale zjistit, které rodiny retroelementů jsou v současnosti svojí aktivitou zodpovědné za vznik insercí, bylo nutné případy, které vznikly jiným procesem, než je retrotransposice, vyloučit. Toto bylo provedeno několikakrokovým filtračním procesem výstupního souboru programu RepeatMasker, jehož účelem bylo zajistit, aby na konci zůstaly pouze inserce způsobené právě retrotransposicí. V průběhu 3 kroků se původní počet retrotransponů obou skupin (L1 a B1) nalezených insercí snížil asi o 30 %. Počty insercí v jednotlivých krocích jsou shrnuty v tabulce 2.

**Tabulka 2:**

Krok	Počet B1 záznamů	Počet L1 záznamů
1.	66	164
2.	58	127
3.	49	116

#### *Krok 1: Nalezení L1 a B1 retrotransposonů*

Cílem tohoto kroku bylo vyloučit z výstupního souboru RepeatMasker záznamy patřící do jiných rodin, než jsou B1 a L1 retroelementy. Zároveň také záznamy představující „simple repeat“ nebo „low complexity repeat“. Zde je třeba zmínit, že v tomto kroku záznam  $\neq$  inserce. Soubor obsahoval více záznamů pro jednu inserci (viz následující krok).

#### *Krok 2: Vyloučení insercí majících více záznamů*

Soubor z předchozího kroku obsahoval inserce, které měly více záznamů – tzn. inserce byla tvořena 2 a více retrotransposony z rodiny B1 nebo L1. Z tabulky byly tedy následně vyloučeny duplicitní záznamy. Takto byly vyloučeny celkem 4 B1 inserce a 18 L1 insercí.

#### *Krok 3: Nalezení insercí způsobených aktivitou retrotransposonu*

Jak jsem zmínil výše, v případě řady insercí nebyla za jejich vznik zodpovědná retrotransposice, ale spíše nehomologní rekombinace. Cílem tohoto kroku bylo nalézt pouze takové inserce, u kterých lze s jistotou říci, že vznikly recentní retrotransposicí – tzn. délka inserce  $\sim$  délka retrotransposonu. Tohoto bylo využito při nastavení filtru – byly vyřazeny inserce začínající dále než 30 nt v případě B1 a 40 nt v případě L1 od začátku a končící dále než 100 nt od konce sekvence u obou rodin (seznam nalezených insercí je uveden v příloze 2).

### 4.2.3 ÚPRAVA A ANALÝZA SEKVENCÍ

Sekvence byly alignovány pomocí programu ClustalW, který byl součástí programu BioEdit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) a následně pomocí stejného programu ručně upraveny.

#### *Výpočet entropie*

Entropie udává variabilitu daného místa v sekvenci na základě srovnání souboru 2 a více sekvencí. K výpočtu entropie byl použit program BioEdit. Graf byl následně zobrazen pomocí R (<http://www.r-project.org/>).

Vzorec pro výpočet entropie pro dané místo  $H_{(p)}$  je:

$$H_{(p)} = -\sum f_{(b,p)} \log_2 f_{(b,p)},$$

kde  $p$  označuje posici v sekvenci,  $b$  pak dané residuum a  $f_{(b,p)}$  je frekvence daného residua na dané posici.

#### *Mismatch distribution*

K získání frekvencí párových rozdílů mezi sekvencemi v rámci podrodin B1 retrotransposonů byl použit program DNAsp (<http://www.ub.edu/dnasp/>).

#### 4.2.4 NAVRŽENÍ PRIMERŮ

Dalším krokem bylo navržení primerů okolo námi nalezených insercí. Primery byly navrhovány na referenční sekvenci z C57BL/6J bez ohledu, jestli nalezená inserce byla přítomna v tomto kmeni nebo v MSM/Ms. Pomocí bioinformatického modulu BioPerl byla stažena sekvence inserce z referenčního genomu C57BL/6J spolu s 500 bp genomové sekvence na každou stranu od místa inserce. Tedy v případě, že inserce byla v MSM/Ms, byla délka stažené sekvence:  $500 + 500 = 1000$  bp. Pokud se inserce nacházela v C57BL/6J pak:  $500 + (\text{délka inserce}) + 500 = 1000 + (\text{délka inserce})$  bp.

Abych zamezil případným tzv. nespecifickým amplifikacím v průběhu PCR, všechny repetitivní oblasti v 500 bp obklopujících sekvencích byly zamaskovány pomocí N v programu RepeatMasker. Primery byly následně navrženy pomocí programu Primer3 (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>), tak aby se nasedací (annealingová) teplota pohybovala mezi 60 – 62 °C. U řady insercí se však z důvodu vysokého obsahu repetitivních sekvencí v jejich okolí nepodařilo navrhnout primery (viz výsledky). Klíčovým bylo také testovat navržené primery pomocí PCR *in silico* na serveru UCSC (<http://genome.ucsc.edu/>), abych zamezil amplifikaci více oblastí v případě, že tato oblast prošla duplikací. Sekvence jednotlivých primerů jsou uvedeny v příloze 3.

## 4.3 GENOTYPIZACE L1 A B1 U 8 INBREDNÍCH KMENŮ

### 4.3.1 VÝBĚR VHODNÝCH KMENŮ

Pro zjištění variability daných insercí v rámci komplexu *M. musculus* a sesterských taxonů jsem se rozhodl testovat přítomnost/absenci insercí na panelu 8 inbredních kmenů. 6 kmenů bylo odvozeno od tří poddruhů *M. musculus*, kde od každého poddruhu byly testovány 2 kmeny (2 x *M. musculus musculus*, 2 x *M. m. domesticus*, 2 x *M. m. castaneus*), 2 kmeny pak od sesterských taxonů (1 x *M. macedonicus*, 1 x *M. spretus*). Sumarizace jednotlivých kmenů je uvedena v tabulce 3.

**Tabulka 3:**

Název kmene	druh	poddruh	stát	lokalita
BZO	<i>M. musculus</i>	<i>M. m. domesticus</i>	Alžírsko	Oran
Straas 2788	<i>M. musculus</i>	<i>M. m. domesticus</i>	Německo	Straas
MPB	<i>M. musculus</i>	<i>M. m. musculus</i>	Polsko	Bialowieza
CTP	<i>M. musculus</i>	<i>M. m. castaneus</i>	Thajsko	Pathumthani
CIM	<i>M. musculus</i>	<i>M. m. castaneus</i>	Indie	Masinagudi
Stud 2802	<i>M. musculus</i>	<i>M. m. musculus</i>	Česká Republika	Studenec
SEG	<i>M. spretus</i>		Španělsko	Grenade
XBS	<i>M. macedonicus</i>		Bulharsko	Slantchev Briag

### 4.3.2 IZOLACE DNA

Jednotlivé vzorky byly poskytnuty ze sbírek Dr. Pavla Munclinger (Přf UK v Praze, ČR) a Dr. Barbary Dod (Univerzita Montpellier II, Francie). Izolovaná DNA všech kmenů byla normalizována na koncentraci 10 ng/ul.

### 4.3.3 PCR

K amplifikaci chtěných fragmentů jsem použil standardní PCR. Protokol viz tabulka 4. Všechny inserce byly amplifikovány stejným programem na několika typech thermocyclerů (MJ Research Thermal Cycler, BioRad, Eppendorf). Program viz tabulka 5.

### 4.3.4 ELEKTROFORÉZA

PCR produkty byly separovány pomocí horizontální elektroforézy při napětí 100 V po dobu 60 minut. K separaci byl použit 2 % agarózový gel. K odečtení délky amplifikovaných fragmentů byl použit standard (100bp DNA ladder plus Gene Ruler<sup>TM</sup>, Fermentas).



**Tabulka 4:**

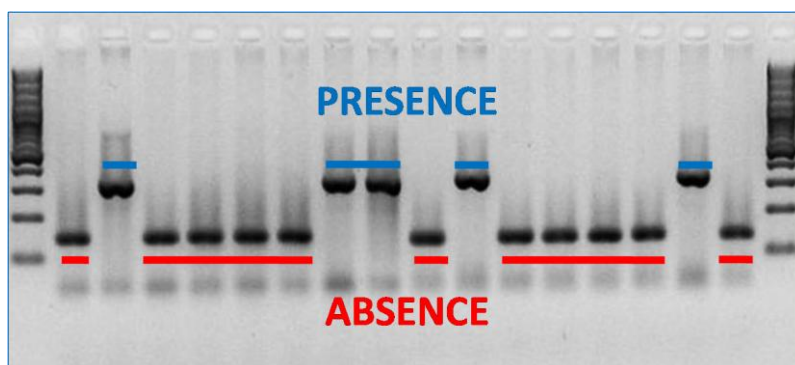
složky PCR směsi	objem 1 PCR (μl)	zdrojová koncentrace	koncová koncentrace
pufř (10x)	1.50	10x	
dNTPs	1.50	2.00 mM	0.20 mM
MgCl <sub>2</sub>	1.20	25.00 mM	2.00 mM
Primer F	1.00	0.15 mM	0.01 mM
Primer R	1.00	0.15 mM	0.01 mM
Taq	0.10	100.00 mM	0.67 mM
H <sub>2</sub> O	3.70		
DNA	5	10	3.33 ng/ul
Celkem 1 PCR	15		

**Tabulka 5:**

Teplota	Délka fáze	Počet cyklů
95 °C	3 min	1x
95 °C	45 s	35x
62 °C	45 s	
72 °C	45 s	
72 °C	10 min	1x

#### 4.3.5 HODNOCENÍ VÝSLEDKŮ ELEKTROFORÉZY

Po separaci amplifikovaných fragmentů pomocí elektroforézy bylo možno odečíst délku jednotlivých fragmentů a určit tak, na základě předchozí znalosti očekávané délky v případě přítomnosti/absence inserce, zdali je u daného kmene inserce přítomna (obrázek 9).



**Obrázek 9:** Znázornění výsledků elektroforézy. Proužky označeny modře jsou fragmenty obsahující inserci. Proužky označeny červeně jsou pak fragmenty bez inserce.

# 5. VÝSLEDKY

---

## 5.1 ANALÝZA KONSENSUÁLNÍCH SEKVENCÍ B1 PODRODIN

### 5.1.1 ANALÝZA KONSENSUÁLNÍCH SEKVENCÍ B1 PODRODIN Z DATABÁZE REPBASE

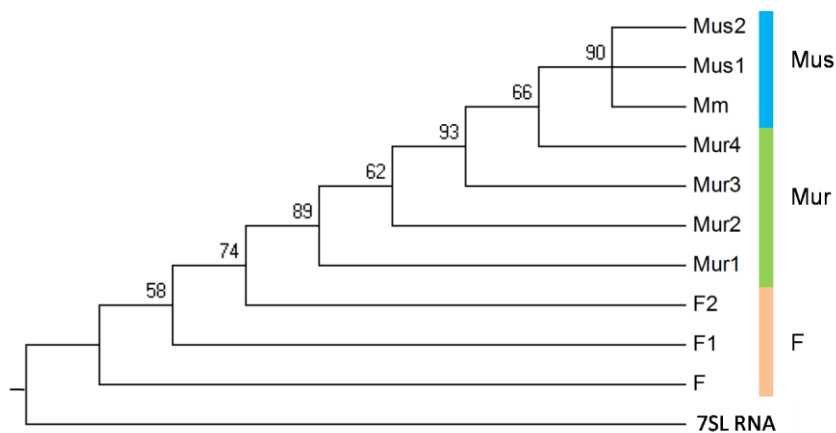
Konsensuální sekvence 10 B1 podrodin stažené z databáze RepBase (3.6.2010) byly rozděleny do 3 skupin na základě taxonomického zařazení v RepBase. Skupina Mus zahrnuje 3 podrodiny (Mm, Mus1, Mus2), které jsou podle taxonomického zařazení přítomny u rodu *Mus* bez jakékoliv další specifikace. Druhá skupina nazvaná Mur zahrnuje 4 podrodiny (Mur1, Mur2, Mur3, Mur4), které jsou přítomny u čeledě *Muridae*. Poslední je skupina F, která zahrnuje 3 podrodiny (F, F1, F2). Zástupci skupiny F evolučně předcházely skupině Mur, z informací uvedených v RepBase však není zřejmé, pro kterou taxonomickou jednotku v rámci hlodavců jsou charakteristické. Sekvence jednotlivých podrodin byly následně alignovány (příloha 5 a 6) a extrahovány posice, ve kterých se jednotlivé podrodiny lišily (tabulka 7). Celkem takto bylo extrahováno 36 posic včetně insercí a delecí. Konsensuální sekvence jednotlivých podrodin byly porovnávány k B1 konsensu odvozenému z těchto sekvencí. Jednotlivé báze v B1 konsensu odpovídaly vždy majoritně zastoupené (>50 %) bázi při srovnání sekvencí jednotlivých podrodin.

Z tabulky (tabulka 7) lze vyčíst, že skupiny Mus a F je možné velmi dobře charakterizovat. Skupina Mus se vyznačuje, ve srovnání s ostatními, 2 bodovými mutacemi (61T a 139G) a delecí dinukleotidu TG na posici 67/68. Skupina F je charakteristická 3 bodovými mutacemi (17A, 25G, 56A). V případě skupiny Mur nelze nalézt žádné exkluzivní mutace. Tuto skupinu lze tedy rozeznat jedině kombinací nukleotidů na všech posicích, kterými se liší od obou předchozích skupin. Zařazení B1 retrotransposonů do jednotlivých skupin lze tedy shrnout způsobem uvedeným v tabulce 6:

**Tabulka 6:** Mutace charakteristické pro skupiny podrodin

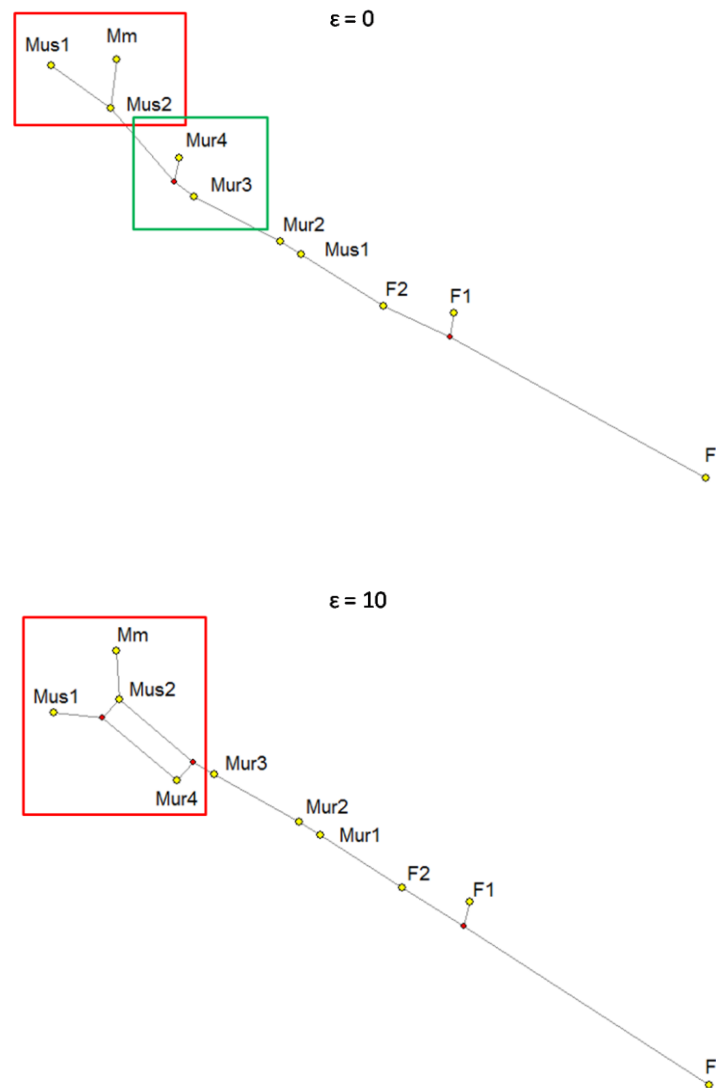
	Bodové mutace	Indel
Mus	17G, 25T, 56C, 61T, 139G	67-, 68-
Mur	17G, 25T, 56C, 61C, 139A	67T, 68G
F	17A, 25G, 56A, 61C, 139A	67T, 68G

Posice		4	8	9	11	17	21	25	39	40	47	56	61	62	63	64	65	67	68	73	76	77	87	93	96	0	3	7	3	5	6	9	0	2	5	8	9		
B1 konsensus*		G	G	-	G	G	G	T	C	G	A	C	C	T	-	C	T	T	G	C	G	G	T	T	T	T	C	A	A	G	G	A	C	C	A	G	A		
Mus	Mm	.	.	-	.	.	.	.	.	.	.	.	T	.	-	.	.	-	-	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	A	G
Mus	Mus2	.	.	-	.	.	.	.	.	.	.	.	T	.	-	.	.	-	-	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G
Mus	Mus1	.	A	G	.	.	.	.	T	.	.	.	T	.	-	.	.	-	-	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G
Mur	Mur4	.	.	G	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Mur	Mur3	.	.	-	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Mur	Mur2	.	.	-	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	T	T	.	.
Mur	Mur1	.	.	-	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	T	T	.	.
F	F2	.	.	-	.	A	.	G	.	.	.	A	.	.	-	C	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	T	T	.	.		
F	F1	.	.	-	.	A	.	G	.	.	T	A	.	G	-	C	.	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	T	T	.	.		
F	F	A	A	-	A	A	A	G	T	A	T	A	.	.	G	.	.	.	T	A	.	G	.	.	.	C	T	.	.	.	.	.	.	T	.	.	.		



**Tabulka 7:** Porovnání odlišností konsensuálních sekvencí podrodin B1 retrotransponů stažených z databáze RepBase.

**Obrázek 10:** Výsledek Maximux Parsimony. Jako outgroup byla použita 7SL RNA (Quentin 1992).



**Obrázek 11:** Výsledek Median-Joining Network. Nahoře – hladina tolerance 1, dole – hladina tolerance 2

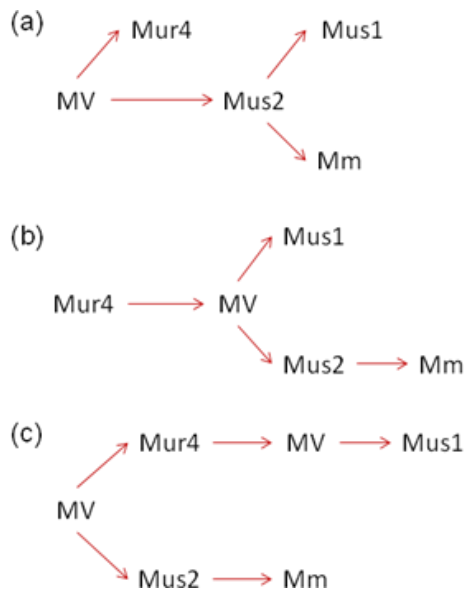
### 5.1.2 EVOLUČNÍ VZTAHY JEDNOTLIVÝCH PODRODIN B1

K určení evolučních vztahů B1 podrodin byly užity 2 metody. Výsledek metody Maximal Parsimony (viz kapitola Materiál a metodika) potvrdil fylogenezi jednotlivých skupin v pořadí F->Mur->Mus (obrázek 10). Jako outgroup byla použita sekvence 7SL RNA použitá v článku Quentin (1992). Analýza neumožňovala zahrnout inserce a delece. Až na evolučně nejmladší skupinu Mus bylo možné určit kladogenezi jednotlivých podrodin B1 retrotransposonů.

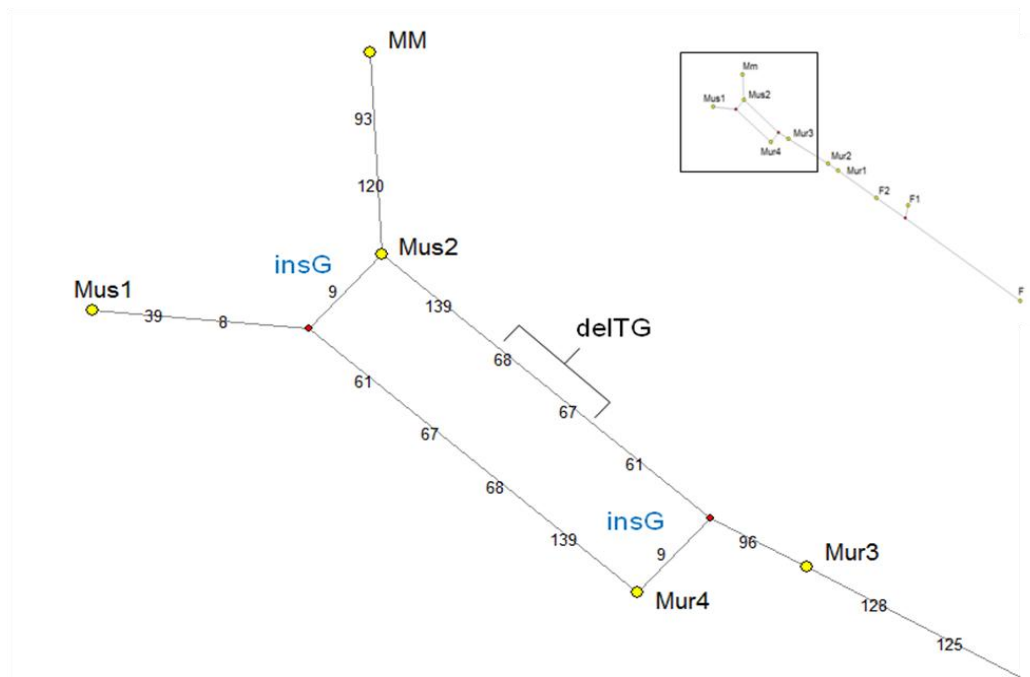
Vzhledem k nejasným evolučním vztahům v rámci skupiny Mus byla použita metoda Median-Joining Network (viz kapitola Materiál a metodika). Tato metoda

umožňuje velmi efektivně pracovat s daty, která nejsou evolučně příliš vzdálena a jsou tak v jednom čase přítomni jak potomci, tak i ancestrální stavy. Jednotlivé záměny byly váženy stejně – defoltní hodnotou 10 – bez ohledu na typ mutace. Byly testovány 2 hladiny tolerance ( $\epsilon$ ) při vyhledávání odpovídající sítě. Algoritmus nastavený na hladinu tolerance 1 ( $\epsilon = 0$ ) považoval každou mutaci za unikátní, na hladině tolerance 2 ( $\epsilon = 10$  – odpovídá váze jedné mutace) pak toleroval 1 mutaci (bez ohledu na typ mutace). Tedy v případě hladiny tolerance 2 algoritmus toleroval více mezistavů, které nebyly zaznamenány v případě hladiny tolerance 1. Z obrázku (obrázek 11) je zřejmé, že pro různé hladiny tolerance dostáváme pro mladší podrodiny 2 různé výsledky. Při hladině tolerance 1 je výsledkem jeden možný scénář evoluce B1 podrodin. Je jím jeden chybějící stav mezi podrodinami Mur3 a Mus2, od kterého byla odvozena podrodina Mur4 (viz zelený rámeček). Důvodem tohoto větvení je inserce G na posici 9, která vznikla mohem dříve než stejná inserce v případě podrodiny Mus1. Dále je z obrázku (obrázek 11 nahoře) zřejmá dichotomie se známým ancestrálním stavem. Tímto ancestrálním stavem je podrodina Mus2, ze které vznikly dvě podrodiny: Mus1 a Mm (červený rámeček). Otázkou je, zdali z Mus2 vznikla jedna z dvou nových podrodin mutací, či má každá svou vlastní zdrojovou kopii. Pro hladinu tolerance 2 (obrázek 11 dole), vytvořením dvou možných mezistavů, byly navrženy 3 možnosti evoluce 4 nejmladších rodin (obrázek 12). Vzhledem k faktu, že v případě (c) by muselo dojít paralelně dvakrát k deleci dinukleotidu TG na posicích 67 a 68, považuji za nejpravděpodobnější pouze možnosti (a) a (b) s tím, že možnost (a) je *de facto* možnost, která vyšla při užití hladiny tolerance 1. Tedy za druhou možnou alternativu lze považovat scénář (b), kde předkem všech 3 podrodin ze skupiny Mus je podrodina Mur4. Chybí tedy společný předek Mus1 a Mus2 mající inserci G na posici 9 a zároveň deleci výše zmíněného dinukleotidu.

Z obrázku je zřejmé, že nelze jednoznačně určit, které ze dvou navrhovaných schémat je pravděpodobnější. Jediný rozdíl mezi scénářem (a) a (b) je na posici 9. Jedná se o inserci G. V případě (a) by muselo dojít dvakrát nezávisle na sobě k dané inserci, zatímco v případě (b) nejprve k inserci a poté k deleci.



**Obrázek 12:** 3 hypotetické evoluční scénáře nejmladších podrodin B1. MV - chybějící mezistav („median vector“).



**Obrázek 13:** Detail sítě s nejmladšími podrodinami při hladině tolerance 2.

### 5.1.3 POROVNÁNÍ PODRODIN Z REPBASE A PODRODIN DEFINOVANÝCH QUENTINEM (1989)

Konsensuální sekvence podrodin z RepBase jsou vytvářeny na základě souboru myších B1 z celého genomu (<http://www.girinst.org/>). Zatímco podrodiny definované Quentinem (1989) byly ustanoveny dlouho před osekvenováním myšího genomu a jsou tedy definovány pouze na základě několika v té době známých sekvencí (viz Úvod). Aby bylo možno diskutovat mé výsledky s předchozími autory, bylo nutno porovnat tyto dva systémy klasifikace a pokusit se přiřadit navzájem jednotlivé podrodiny.

V tabulce (tabulka 8) byly podrodiny definované Quentinem (1989) přiřazovány na základě sekvenční podobnosti k podrodinám z databáze RepBase. K tomuto porovnání byly vyřazeny rychle mutující CpG páry. Z Quentinových (1989) podrodin bylo možno bez jediného rozdílu přiřadit pouze dvě podrodiny. Jednalo se o B1B a B1D podrodiny, které odpovídaly B1\_Mm a B1\_Mus2 z RepBase. Ostatní podrodiny se lišily vždy jednou posicí od konsensuálních sekvencí z RepBase. Podrodině B1A, která předcházela podrodině B1B, v podstatě nejvíce odpovídala stejně jako B1B podrodině B1\_Mm. Jediný rozdíl byl cytosin na posici 120 místo thyminu v případě B1\_Mm. Podrodina B1C víceméně odpovídala podrodině Mus1, lišila se od ní opět na posici 120. B1C místo cytosinu měla thymin. Podrodina B1E odpovídala podrodině Mur3 s rozdílem absence dinukleotidové inserce na posicích 67 a 68. Nejstarší Quentinova (1989) podrodina B1F se v mém přiřazení nejvíce shodovala s podrodinou Mur2. Rozdílem byl cytosin místo thyminu na posici 122. Detailněji bude tato problematika probrána v diskusi.

**Tabulka 8:** Srovnání klasifikačního systému podrodin B1 z databáze RepBase a systému vytvořeného Quentinem (1989).

	Posice podle RepBase	9	61	93	120	11	25	47	62	63	64	65	67	68	76	77	87	100	103	107	113	115	116	119	122	128
	Srovnání:																									
B1		-	T	N	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1_Mm	B1B, B1A(+120T)	-	T	A	T	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1_Mus2	B1D	-	T	G	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1_Mus1	B1C(+120C)	G	T	G	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1_Mur4		G	C	G	C	G	T	A	T	-	C	T	T	G	G	G	T	T	C	A	A	G	G	A	C	A
B1_Mur3	B1E(+ins67/68)	-	C	G	C	G	T	A	T	-	C	T	T	G	G	G	T	T	C	A	A	G	G	A	C	A
B1_Mur2	B1F(+122T, +ins67/68)	-	C	T	C	G	T	A	T	-	C	T	T	G	G	G	T	T	C	A	A	G	G	A	T	G
B1_Mur1		-	C	T	C	G	T	A	T	-	C	T	T	G	G	G	T	T	C	C	A	G	G	A	T	G
B1F2		-	C	T	C	G	G	A	T	-	C	C	T	G	G	G	T	T	C	C	A	G	G	A	T	G
B1F1		-	C	T	C	G	G	T	G	-	C	C	T	G	G	G	G	T	C	C	A	G	G	A	T	G
B1F			C	T	C	A	G	T	T	G	C	T	T	G	A	G	G	T	C	C	T	G	G	A	C	G
	Posice podle Quentin(1989)	9	60	89	116	10	24	46	61		62	63			72	73	83	96	99	103	109	111	112	115	118	124
B1		-	T	N	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1A		-	T	A	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1B		-	T	A	T	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1C		G	T	G	T	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1D		-	T	G	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1E		-	C	G	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A
B1F		-	C	T	C	G	T	A	T	-	C	T	-	-	G	G	T	T	C	A	A	G	G	A	C	A

Žlutě jsou vyznačeny posice definované Quentinem (1989) jako diagnostické. Modře jsou pak označeny posice, kterými se lišily jednotlivé konsenzuální sekvence podrodin B1 stažené z databáze RepBase.



## 5.2 SEKVENČNÍ ANALÝZA RECENTNĚ AKTIVNÍCH L1 A B1 RETROTRANSPOSONŮ

### 5.2.1 RECENTNĚ AKTIVNÍ PODRODINY

Na základě sekvenční podobnosti ke konsensu byly jednotlivé inserce pomocí programu RepeatMasker zařazeny do podrodin. Celkem 49 B1 retrotransposonů bylo zařazeno do 4 podrodin s tím, že nejpočetněji zastoupeni byli zástupci podrodin Mm a Mus2. Pouze 5 zástupců bylo nalezeno v případě podrodiny Mus1 a jen jeden zástupce podrodiny Mur4. V případě L1 retrotransposonů bylo nalezeno celkem 5 podrodin. Podrodina L1\_Mm, která nebyla definována autory Mears & Hutchison 2001, byla zastoupena pouze 1 kopií. Nejvíce byly zastoupeny podrodiny A a T s 43 a 41 kopiemi. Menší početní zastoupení měly podrodiny F2 a F3. Počty kopií pro jednotlivé podrodiny jsou uvedeny v tabulce (tabulka 9).

**Tabulka 9:**

	podrodina	počet
B1	B1_Mm	20
	B1_Mur4	1
	B1_Mus1	5
	B1_Mus2	23
L1	L1_Mm	1
	L1Md_A	43
	L1Md_F2	23
	L1Md_F3	8
	L1Md_T	41

### 5.2.2 MSM/Ms vs. C57BL/6J PŮVOD

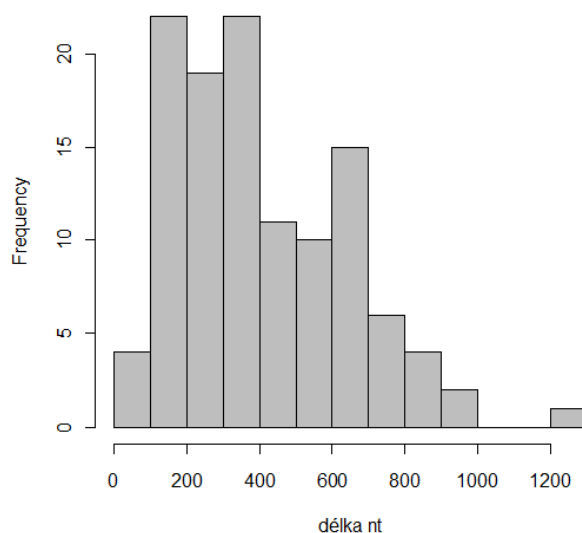
Počty insercí se lišily v závislosti na původu. Obecně bylo nalezeno mnohem více insercí z kmene C57BL/6J než z kmene MSM/Ms. Navíc byly zjištěny značné rozdíly mezi B1 a L1 retrotransposony. V případě L1 retrotransposonů jich bylo nalezeno jen 8 u MSM/Ms a 108 u C57BL/6J (tabulka 10). U rodiny B1 byly počty více vyrovnané: 17 u MSM/Ms a 32 u C57BL/6J. Tyto rozdíly vyšly signifikantně na hladině  $\alpha = 0,05$  při použití Fisherova exaktního testu ( $p\text{-value} = 1,82 \cdot 10^{-5}$ ).

**Tabulka 10:**

kmen	L1	B1
MSM/Ms	8	17
C57BL/6J	108	32

### 5.2.3 ANALÝZA SEKVENCÍ

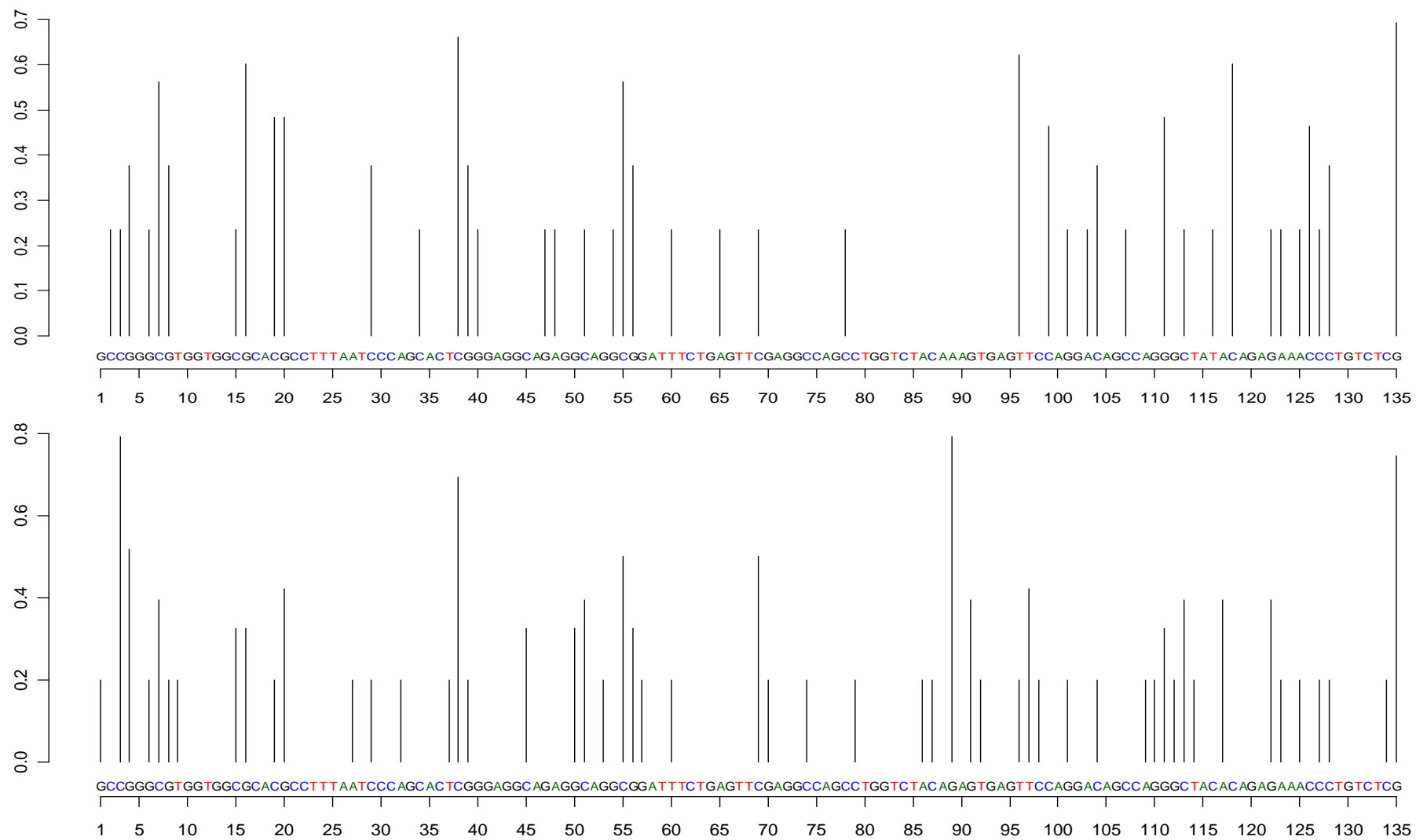
Pro následnou sekvenční analýzu byly vybrány pouze celé nezkrácené sekvence. Toto nebylo možné v případě sekvencí L1 retrotransposonů, protože všechny byly přirozeně zkráceny. Distribuce délek je znázorněna na grafu (obrázek 14) s průměrem 405 nt a mediánem 367 nt. Většina nalezených recentních L1 je kratších než 1/10 z délky konsensu. Sekvence B1 elementů ve většině případů dosahovaly plné délky konsensu, proto byla analyzována pouze tato rodina retrotransposonů.



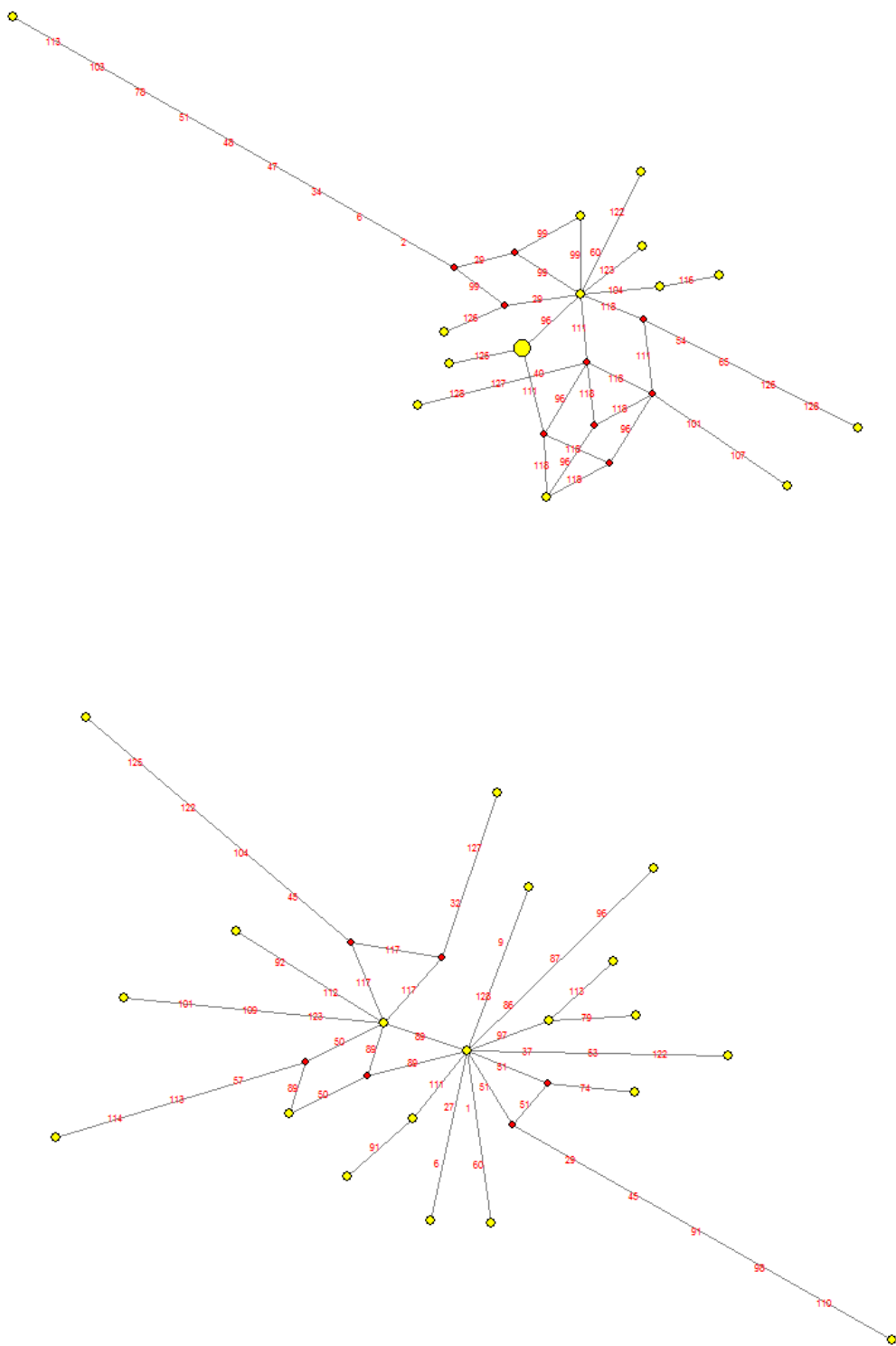
**Obrázek 14:** Distribuce délek 116 recentně aktivních L1 retrotransposonů. Na ose x vynesena délka jednotlivých insercí (nt), na ose y pak jejich frekvence.

### 5.2.4 SROVNÁNÍ AKTIVITY PODRODIN B1 RETROTRANSPONŮ

Sekvence B1 byly rozděleny na základě příslušnosti k jednotlivým podrodinám, následně pak byly alignovány a ručně upraveny na délku konsensuální sekvence dané podrodiny. Vzhledem k nízkému počtu kopií v případě podrodiny Mus1 a Mur4 byly dále analyzovány pouze podrodiny Mm a Mus2. Po vyloučení krátkých sekvencí se počet zástupců u podrodiny Mm zmenšil na 16 a u Mus2 na 20. Výpočtem entropie (viz



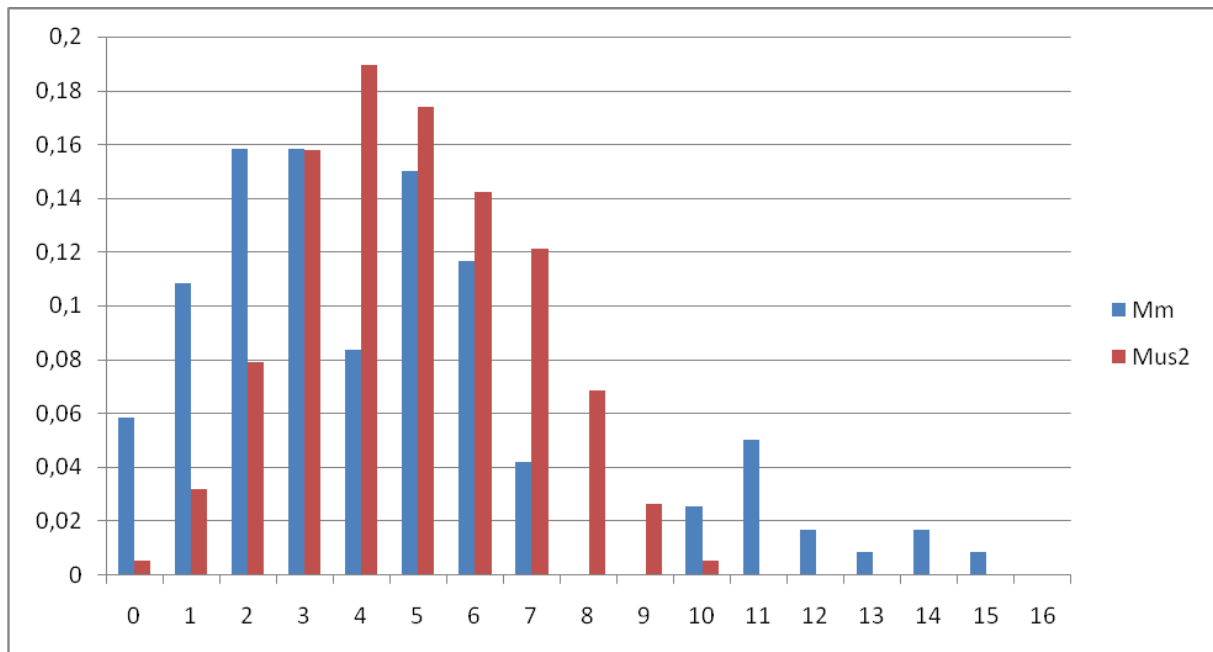
**Obrázek 15:** Hodnoty entropie vynesené na konsenzuální sekvenci dané podrodiny. Podrodina Mm nahoře, podrodina Mus2 dole.



**Obrázek 16:** Výsledky analýzy Median-Joining Networks ( $\epsilon = 0$ , váhy = 10).  
 (a) nahore – podrodina Mm, (b) dole – podrodina Mus2.

Materiál a metodika) byla zjištěna variabilita jednotlivých posic v sekvenci. Z obrázku (obrázek 15) je zřejmé, že nejvíce mutující posice jsou CpG páry. Někteří autoři uvádějí, že ve srovnání s ostatními posicemi, CpG páry mutují až 10x rychleji (Xing et al. 2004). Proto v případě SINE retroelementů doporučují tato místa před analýzou vyloučit. Sekvence jednotlivých podrodin byly tedy alignovány ke konsensu a všechny CpG páry vyloučeny. Následně byly sekvence analyzovány pomocí metody Median-Joining Networks s hladinou tolerance  $\epsilon = 0$ , všechny posice byly váženy stejně defoltní hodnotou 10. Cílem bylo zjistit pattern větvení jednotlivých sekvencí. V obou případech se jedná o recentní inserce, proto bylo předpokládáno, že všechny nalezené inserce by mohly mít jeden zdroj. Výsledné sítě jsou znázorněny na obrázku (obrázek 16). Z výsledků lze vidět, že obě sítě mají víceméně centrální bod popř. více bodů, od kterých se odvětvují ostatní. Přesto však se pravděpodobně nejedná o jeden „master gene“, který je zdrojem pro ostatní kopie. V případě sítě pro podrodinu Mm (obrázek 16 nahoře), lze vidět, že většina sekvencí je v centrální části, velmi často pak od ostatních oddělena pouze jednou mutací, lze nalézt i sekvence identické. Na druhou stranu nelze popřít, že i v rámci podrodiny Mm jsou sekvence velmi odvozené, které se od ostatních liší až 11 mutacemi. V případě podrodiny Mus2 (obrázek 16 dole) jsou vzdálenosti poněkud rovnoměrnější, více jsou zastoupeny sekvence, které jsou od sebe navzájem vzdáleny 2 i více mutacemi. U této podrodiny nebyly nalezeny elementy identické. Vzhledem k výsledku analýzy Median-Joining Networks byly sekvence dále analyzovány, aby bylo možno určit, zdali jsou mezi podrodinou Mm a Mus2 detekovatelné rozdíly v době aktivity. Pomocí programu RepeatMasker byla vypočítána divergence od celogenomového konsensu pro jednotlivé sekvence a pak průměr divergence pro jednotlivé podrodiny. Ukázalo se, že průměrná divergence od konsensu v rámci těchto dvou podrodin je *de facto* stejná. V případě podrodiny Mm je 4,1 a v případě podrodiny Mus2 je 4,17 (viz příloha 7). Vzhledem k faktu, že divergence je vypočítávána na základě celogenomového konsensu i mnohem starších zástupců dané podrodiny, může být zavádějící používat tuto míru kurčení rozdílu stáří mezi dvěma soubory velmi recentních kopií. Mnohem relevantnější je porovnávat jednotlivé recentní kopie mezi sebou. Proto byl vypočítán počet rozdílů pro páry sekvencí (nejsou zahrnuty CpG páry) v rámci podrodin a následně zobrazena distribuce těchto rozdílů („mismatch pair-wise distribution“). Výsledek je zobrazen na obrázku (obrázek 17). Z obrázku je patrné, že

jednotlivé sekvence v rámci podrodiny Mm jsou méně odlišné, než sekvence v rámci rodiny Mus2. Přesto však nejvyšších hodnot na ose x dosahují párová srovnání zástupců podrodiny Mm. Jedná se konkrétně o sekvenci 19\_37506476|MSMg01-237L20.TJ, která ve všech třech analýzách ukazuje značně odchýlené hodnoty. Od centrálního bodu v analýze Median-Joining Network je oddělena 11 mutacemi. Při výpočtu divergence od konsensu dosahuje velmi vysoké hodnoty 13 %.



**Obrázek 17:** Mismatch distribution. Distribuce párových srovnání počtu mutací mezi sekvencemi jednotlivých podrodin. Na ose x je vyneseno počet rozdílů, osa y znázorňuje frekvenci.

## 5.3 GENOTYPIZACE L1 A B1 INSERCÍ U INBREDNÍCH 8 KMENŮ

### 5.3.1 VÝSLEDKY GENOTYPIZACE

Ke zjištění variability nalezených retrotransposonů u všech poddruhů komplexu *M. musculus* a u dalších dvou druhů byly jednotlivé retrotransposony testovány na přítomnost/absenci u několika inbredních kmenů odvozených od zmiňovaných druhů myší (kompletní seznam viz Materiál a metodika). Kromě zjištění odlišné aktivity jednotlivých podrodin bylo cílem tohoto testování také nalézt možné diagnostické markery pro jednotlivé poddruhy *M. musculus*.

Z celkového počtu 49 B1 retrotransposonů a 116 L1 retrotransposonů bylo na přítomnost/absenci u 8 myších kmenů testováno pouze 30 B1 a 102 L1. Důvodem pro tento snížený počet byla neschopnost navrhnout primery kolem dané inserce kvůli vysokému obsahu repetitivních oblastí obklopujících studované inserce. V případě B1 došlo k výraznějšímu snížení než v případě L1. Testované retrotransposony tvoří 61 % (B1) a 88 % (L1). Následně byly vybrány pouze ty inserce, které bylo možné amplifikovat u všech testovaných kmenů. Celý proces výběru markerů je sumarizován v tabulce (tabulka 11).

**Tabulka 11:**

Krok	B1	Rel. počet B1	L1	Rel. počet L1
Před navržením primerů	49	1	116	1
Po navržení primerů	30	0,61	102	0,88
Genotypy u všech kmenů	23	0,47	82	0,71

V následujících dvou tabulkách (tabulka 12 a 13) jsou shrnuty výsledky laboratorní části. Jsou zde uvedeny všechny nalezené kombinace jednotlivých genotypů u všech studovaných myších kmenů. Počty insercí majících danou kombinaci jsou pak uvedeny zvlášť pro inserce nacházející se v referenčním genomu C57BL/6J (gen) a inserce nalezené v MSM/Ms (bac).

**Tabulka 12: B1 retrotransposony**

Kateg.	<i>M. m. domesticus</i>		<i>M. m. musculus</i>		<i>M. m. castaneus</i>		<i>M. macedonicus</i>	<i>M. spretus</i>	bac	gen
	Straas 2788	BZO	MPB	Stud 2802	CIM	CTP	XBS	SEG		
1&2	a	a	a	a	a	a	a	a	1	1
3&4	p	a	a	a	a	a	a	a	0	1
	a	h	a	a	a	a	a	a	0	1
	p	p	a	a	a	a	a	a	0	4
12	p	p	a	p	a	a	a	a	0	1
	p	p	h	p	a	a	a	a	0	1
	p	p	p	p	a	a	a	a	0	1
10	h	a	a	a	p	a	a	a	0	1
	p	p	a	a	p	p	a	a	0	1
11	a	a	p	p	p	a	a	a	1	0
	a	a	a	p	p	p	a	a	1	0
13	a	a	p	p	p	a	p	a	1	0
	p	p	a	p	p	a	a	p	0	1
	p	p	p	p	p	a	a	p	0	1
	p	p	a	a	a	a	p	p	0	2

	p	p	a	a	p	p	p	p	0	1
	a	a	p	p	p	p	p	p	1	0
14	p	p	p	p	p	p	p	p	0	1

**Tabulka 13: L1 retrotransposony**

Kateg	<i>M. m. domesticus</i>		<i>M. m. musculus</i>	<i>M. m. castaneus</i>		<i>M. macedonicus</i>	<i>M. spretus</i>	bac	gen
	Straas 2788	BZO	MPB	CIM	CTP	XBS	SEG		
1&2	a	a	a	a	a	a	a	2	10
3&4	h	a	a	a	a	a	a	0	1
	p	a	a	a	a	a	a	0	11
	p	h	a	a	a	a	a	0	2
	a	p	a	a	a	a	a	0	3
	p	p	a	a	a	a	a	0	23
12	a	a	p	a	a	a	a	2	0
	p	p	p	a	a	a	a	0	1
10	p	a	a	p	a	a	a	0	1
	a	p	a	p	a	a	a	0	1
	p	p	a	p	a	a	a	0	3
	p	p	a	a	p	a	a	0	1
	p	a	a	h	p	a	a	0	1
	a	p	a	p	p	a	a	0	1
	p	a	a	p	p	a	a	0	1
11	a	a	h	a	h	a	a	1	0
	a	a	a	a	p	a	a	1	0
9	p	h	p	p	h	a	a	0	1
	p	p	p	p	p	a	a	1	0
13	a	a	a	a	a	h	a	1	0
	p	p	p	a	h	h	h	0	1
	p	p	a	a	a	a	p	0	1
	p	h	a	a	a	p	p	0	1
	p	p	a	a	a	p	p	0	1
	a	a	a	p	a	p	p	0	1
	p	p	a	p	a	p	p	0	3
	p	p	h	p	a	p	p	0	1
	p	p	a	p	h	p	p	0	1
14	p	p	p	p	p	p	p	0	2

**Tabulka 12 a 13:** Shrnutí nalezených kombinací jednotlivých genotypů (a – absence, h – heterozygot, p – presence). Napravo jsou uvedeny počty insercí mající danou kombinaci (bac – inserce nalezené u MSM/Ms, gen – inserce nalezené u C57BL/6J). Tabulka 13: B1 retrotransposony (8 inbredních kmenů), tabulka 14: L1 retrotransposony (7 inbredních kmenů: Stud 2802 neamplifikoval). První sloupec odpovídá kategoriím v tabulce 16.



Inserce byly následně rozděleny do kategorií na základě přítomnosti/absence u studovaných taxonů (tabulka 14). Definice jednotlivých kategorií je pak v tabulce 15. Je možné vidět, že rozložení počtů recentních insercí se liší u skupin B1 a L1. V případě L1 retrotransposonů jsou významně zastoupeny inserce, které jsou přítomny pouze u poddruhu *M. m. domesticus* (kategorie 3 a 4). Jedná se tak o 49 % (kategorie č.3: 28 % + kategorie č.4: 21 %) insercí z celkového počtu 82 L1 insercí ze všech uvedených kategorií. V případě B1 je to jen 26 % (kategorie č.3: 17 % + kategorie č.4: 9 %). V obou případech, jak u L1, tak i u B1, také převládaly inserce sdílené více poddruhy než inserce charakteristické pro jednotlivé poddruhy *M. musculus* (nejsou-li počítány inserce charakteristické pouze pro *M. m. domesticus*). Vysoký počet insercí (10) obsahuje v případě L1 retroelementů kategorie číslo 10. Do této kategorie spadaly všechny inserce přítomné zároveň u poddruhu *M. m. domesticus* a *M. m. castaneus*. B1 jsou na rozdíl od L1 zastoupeny v kategoriích 10, 11 a 12 mnohem rovnoměrněji. Zajímavé bylo zjištění, že v případě obou rodin, byly nalezeny inserce přítomné také u sesterských taxonů *M. spretus* a *M. macedonicus*, přitom však polymorfní v rámci komplexu *M. musculus* (kategorie číslo 13), přičemž u B1 retrotransposonů takovéto inserce měly větší podíl (30 %) než v případě L1 retrotransposonů (13 %).

**Tabulka 14:**

	L1		B1	
	počet	(%)	Počet	(%)
Jen u C57BL/6j	10	12	1	4
Jen u MSM/Ms	2	2	1	4
<i>M. m. domesticus</i> specifické	23	28	4	17
<i>M. m. domesticus</i> polymorfní	17	21	2	9
<i>M. m. musculus</i> specifické	0	0	0	0
<i>M. m. musculus</i> polymorfní	2	2	0	0
<i>M. m. castaneus</i> specifické	0	0	0	0
<i>M. m. castaneus</i> polymorfní	1	1	0	0
<i>M. musculus</i>	2	2	0	0
<i>M. m. d.</i> & <i>M. m. c.</i>	10	12	2	9
<i>M. m. m.</i> & <i>M. m. c.</i>	1	1	2	9
<i>M. m. d.</i> & <i>M. m. m.</i>	1	1	3	13
<i>M. m.</i> & ( <i>M. s.</i> nebo <i>M. mac.</i> )	11	13	7	30
U všech testovaných kmenů	2	2	1	4
Celkem	82		23	

**Tabulka 15:**

Číslo kategorie	Jméno kategorie	Komentář
1	Jen u C57BL/6J	Inserce nalezené pouze v genomu C57BL/6J
2	Jen u MSM/Ms	Inserce nalezené pouze v genomu MSM/Ms
3	<i>M. m. domesticus</i> specifické	Inserce přítomné u obou studovaných inbredních kmenů odvozených od <i>M. m. domesticus</i>
4	<i>M. m. domesticus</i> polymorfní	Inserce přítomné u jednoho z testovaných kmenů odvozených od <i>M. m. domesticus</i>
5	<i>M. m. musculus</i> specifické	Inserce přítomné u obou studovaných inbredních kmenů odvozených od <i>M. m. musculus</i>
6	<i>M. m. musculus</i> polymorfní	Inserce přítomné u jednoho z testovaných kmenů odvozených od <i>M. m. musculus</i>
7	<i>M. m. castaneus</i> specifické	Inserce přítomné u obou studovaných inbredních kmenů odvozených od <i>M. m. castaneus</i>
8	<i>M. m. castaneus</i> polymorfní	Inserce přítomné u jednoho z testovaných kmenů odvozených od <i>M. m. castaneus</i>
9	<i>M. musculus</i>	Inserce přítomné u všech tří poddruhů, ale nepřítomné u <i>M. spretus</i> a <i>M. macedonicus</i>
10	<i>M. m. d.</i> & <i>M. m. c.</i>	Inserce přítomné vždy alespoň u jednoho kmene z poddruhů <i>M. m. domesticus</i> a <i>M. m. castaneus</i>
11	<i>M. m. m.</i> & <i>M. m. c.</i>	Inserce přítomné vždy alespoň u jednoho kmene z poddruhů <i>M. m. musculus</i> a <i>M. m. castaneus</i>
12	<i>M. m. d.</i> & <i>M. m. m.</i>	Inserce přítomné vždy alespoň u jednoho kmene z poddruhů <i>M. m. domesticus</i> a <i>M. m. musculus</i>
13	<i>M. m.</i> & ( <i>M. s.</i> nebo <i>M. mac.</i> )	Inserce přítomné kromě <i>M. musculus</i> i u sesterských druhů <i>M. spretus</i> a <i>M. macedonicus</i>
14	U všech testovaných kmenů	

### 5.3.2 ROZDÍLY MEZI PODRODINAMI NA ZÁKLADĚ ZASTOUPENÍ U TESTOVANÝCH KMENŮ

Přestože recentně aktivní inserce byly získány srovnáním dvou inbredních kmenů odvozených od poddruhů *M. musculus*, byly tyto polymorfní inserce nalezeny v několika případech i u sesterských taxonů *M. spretus* a *M. macedonicus*. Konkrétně se jednalo o 11 insercí L1 a 7 insercí B1 (tabulka 14). Bylo zjištěno, že pro jednotlivé podrodiny se zastoupení takovýchto insercí liší. Poměr insercí, které jsou přítomny alespoň v jednom z druhů *M. spretus* a *M. macedonicus* vzhledem k počtu všech insercí dané podrodiny, je uveden v tabulce (tabulka 17).

**Tabulka 16:**

podrodina	celkový počet insercí	inserce přítomné u <i>M. spretus</i> nebo <i>M. macedonicus</i>	Podíl (%)
B1			
B1_Mm	9	2	22
B1_Mus1	2	2	100
B1_Mus2	12	4	33
B1 celkem	23	8	35
L1			
L1_Mm	1	0	0
L1Md_A	31	4	13
L1Md_F2	18	5	28
L1Md_F3	6	3	50
L1Md_T	26	1	4
L1 celkem	82	13	16

V případě B1 retrotransposonů lze, vzhledem k nízkým počtům studovaných insercí, těžko hodnotit, zdali se jednotlivé podrodiny liší v přítomnosti mimo komplex *M. musculus* či nikoliv. Na základě výsledků Fisherova exaktního testu se počty mezi podrodinou Mm a Mus2 statisticky neliší ( $p$ -value = 0,659,  $\alpha$  = 0,05). Mnohem výraznější rozdíly jsou mezi podrodinami L1 retrotransposonů. Podrodina L1Md\_T se zdá být nejméně zastoupena u *M. spretus* a *M. macedonicus*. U ostatních podrodin zastoupení postupně narůstá. V případě podrodiny L1Md\_F3 je hodnota rovna číslu 0,5. Tedy 50 % mnou testovaných insercí bylo přítomno mimo komplex *M. musculus* (tabulka 18). Provedl jsem párová srovnání podrodin pomocí Fisherova exaktního testu a na hladině signifikance  $\alpha$  = 0,05 byly nalezeny pouze dvě srovnání jako signifikantně odlišná (viz tabulka 18). Jednalo se o dvojice podrodin T vs. F2 a T vs. F3.

**Tabulka 17:**

	L1Md_A	L1Md_F2	L1Md_F3	L1Md_T
L1Md_A				
L1Md_F2	0,259			
L1Md_F3	0,068	0,362		
L1Md_T	0,362	<b>0,034</b>	<b>0,015</b>	

P-value pro páry podrodin. Fisherův exaktní test ( $\alpha$  = 0,05). Červeně zvýrazněny signifikantní hodnoty.

Rozdíl mezi rodinami L1 a B1 je asi dvojnásobný. Přesto však výsledek Fisherova exaktního testu vyšel nesignifikantně ( $p$  = 0,07311,  $\alpha$  = 0,05).

### 5.3.3 MOŽNÉ VYUŽÍT INSERCÍ JAKO PODDRUHOVĚ DIAGNOSTICKÝCH MARKERŮ

Genotypizací insercí u inbredních kmenů byly nalezeny poddruhově diagnostické markery. Za takovéto markery byly považovány inserce, které byly nalezeny u obou testovaných kmenů z daného poddruhu. Všechny nalezené inserce, které se jevily jako poddruhově diagnostické byly diagnostické výhradně pro poddruh *M. m. domesticus*. Celkově tak bylo nalezeno 23 L1 insercí a pouze 4 B1 inserce.

## 6. DISKUSE

---

### 6.1 SROVNÁNÍ PODRODIN Z DATABÁZE REPBASE A PODRODIN DEFINOVANÝCH QUENTINEM (1989)

Po srovnání konsenzuálních sekvencí jednotlivých podrodin z databáze RepBase s podrodinami definovanými Quentinem (1989) se ukázalo, že lze víceméně jednotlivé podrodiny obou klasifikačních systémů porovnat. Podrodině Mm odpovídá bezvýhradně podrodina B1B a s jednou záměnou podrodina B1A, která v evoluci podrodině B1B předcházela. Doposud nebyla provedena celogenomová analýza jednotlivých podrodin B1 retrotransposonů, nelze tedy vyloučit, že by Mm mohla být rozdělena na více podrodin. Také podrodina B1C, která je charakteristická insercí G na posici 9, se až na jednu záměnu shodovala s podrodinou Mus1. Podrodina B1D, u které došlo podle Quentina (1989) k dichotomii vedoucí k B1C a B1A, odpovídá bezvýhrady podrodině Mus2. Srovnáme-li tedy evoluční schéma na obrázku 3 a jedno z možných schémat, které vyšlo při analýze Median-Joining Networks (obrázek 11 nahoře), zjistíme, že jsou *de facto* tato dvě schémata identická. Je obdivuhodné, že Quentin (1989), přestože měl k dispozici pouze několik desítek sekvencí, byl schopen velmi dobře charakterizovat to, co přinesla celogenomová analýza, která nešla nikterak do hloubky. Starší podrodiny již však tak dobře neodpovídají. Naprosto chybí ekvivalent podrodiny Mur4. Neměla-li by tato podrodina inserci dinukleotidu na pozicích 67 a 68, byla by na základě přítomnosti G na posici 9 velmi podobná, stejně jako podrodina Mus1, podrodině B1C. Podrodiny B1E a B1F, již tak dobře neodpovídají. V obou případech chybí opět inserce dinukleotidu na posici 67 a 68, která by je činila identickými s podrodinami Mur3 a Mur2. Podrodina B1F nikterak nesouvisí s podrodinou F uvedenou v databázi RepBase. Této diskrepanci si všimnul i Jurka (1999), když odesílal konsenzuální sekvenci podrodiny F do databáze RepBase (příloha 1). Nicméně vzhledem k omezenému množství dat v té době, lze zhodnotit Quentinovu (1989) klasifikaci jako velmi kvalitní.

## 6.2 ROZDÍLY V POČTECH NALEZENÝCH INSERCÍ MEZI KMENY C57BL/6J A MSM/Ms

Z výsledků je zřejmé, že se signifikantně liší poměr insercí nalezených u C57BL/6J a MSM/Ms mezi L1 a B1 retrotransposony. Obecně bylo nalezeno mnohem více kopií u kmene C57BL/6J než u kmene MSM/Ms s tím, že u L1 retrotransposonů byl tento rozdíl mnohem větší, než v případě B1 retrotransposonů. Velmi zajímavé se může zdát vysvětlení, že celková aktivita retrotransposonů obou rodin (L1 a B1) je mnohem vyšší u kmene C57BL/6J a tedy u poddruhu *M. m. domesticus*, ze kterého je tento kmen z větší části odvozen (Frazer et al. 2008), než u kmene MSM/Ms, jehož původ je u *M. m. molossinus* (Abe et al. 2004). Je ovšem třeba si uvědomit, že vlivem použité metody získání polymorfních insercí došlo ke vzniku zkreslení (bias). Způsob, jakým byly inserce získány, bylo srovnání dvou genomů. Byl alignován neúplný genom ve formě BACendů z MSM/Ms ke kompletnímu referenčnímu genomu C57BL/6J. Pravděpodobnost získání insercí z referenčního genomu byla mnohem vyšší než v opačném případě. Bylo to dáno faktem, že z kmene MSM/Ms byly použity pouze velmi krátké BACendy mající délku pouze několik stovek nukleotidů. V případě, že retrotransposon pokrýval značnou část sekvence BACendu od jejího začátku, nebyla možná jeho identifikace srovnáním s referenčním genomem. V případě, že se retrotransposon nacházel v referenčním genomu, nehrála jeho pozice roli. Počty nalezených insercí z MSM/Ms byly proto nižší než z C57BL/6J. Lze vidět také rozdíly v počtech mezi B1 a L1 retrotransposony. Tyto rozdíly do značné míry mohly být dány faktem, že přestože všechny L1 byly zkráceny, byly v průměru o několik stovek nukleotidů delší než B1. V případě L1 retrotransposonů byl bias pravděpodobně větší než u B1 retrotransposonů. Vliv biasu se stane ještě významnějším, pokud vezmeme v potaz výsledky, jež uvádí Stehlíková (2009). Autorka zkoumala recentní B2 retrotransposony, jejichž délka je okolo 200 nukleotidů, tedy co do délky je lze ve srovnání s B1 a L1 považovat za intermediární skupinu. Stejnou metodou, kterou byly nalezeny B1 a L1 polymorfní retrotransposony, bylo v případě B2 retrotransposonů nalezeno 45 insercí z C57BL/6J a jen 18 insercí z MSM/Ms. Tedy i tato čísla kladou B2 retrotransposony mezi B1 a L1. Domnívám se tedy, že rozdíly mezi počty získaných

insercí z různých rodin jsou dány spíše biasem vzniklým při získávání insercí, nežli rozdílnou aktivitou rodin retroelementů.

## 6.3 ZASTOUPENÍ INSERCÍ U PODDRUHŮ *M. MUSCULUS*

Zastoupení insercí se u jednotlivých poddruhů *M. musculus* liší v závislosti na tom, zdali daná inserce patří mezi L1 nebo B1 retrotransposony. Největší rozdíl byl mezi insercemi přítomnými pouze u poddruhu *M. m. domesticus*. V případě L1 se jednalo o 49 % testovaných insercí, v případě B1 pak jen o 26 % insercí. Tento rozdíl je možné vysvětlit biasem při získávání insercí. Vzhledem k faktu, že většina genomu laboratorního kmene C57BL/6J je odvozena od poddruhu *M. m. domesticus* (Yang et al. 2007, Frazer et al. 2008) a většina nalezených L1 insercí pocházela právě z tohoto kmene, lze očekávat, že většina jich bude specifická pro poddruh *M. m. domesticus*. V případě B1 již takový rozdíl mezi insercemi nalezenými u C57BL/6J a MSM/Ms není. Přesto však nelze říci, že by v případě této rodiny bylo zastoupení specifických insercí vyšší u poddruhu *M. m. castaneus* a *M. m. musculus*, ze kterých vznikl hybridizací *M. m. molossinus* a následně byl pak odvozen kmen MSM/Ms (Abe et al. 2004). Z obou skupin studovaných retrotransposonů byly pouze 2 L1 inserce přítomny pouze u *M. m. musculus*. Za zajímavý považuji rozdíl mezi L1 a B1 retrotransposony v případě sdílení insercí u dvojic poddruhů. U B1 retrotransposonů jsou jednotlivé kombinace zastoupeny rovnoměrně. *M. m. domesticus* a *M. m. musculus* sdílejí 3 inserce, zatímco tyto jednotlivé poddruhy sdílejí s *M. m. castaneus* jen 2 inserce. Jiná situace je ovšem v případě L1 retrotransposonů. Velmi výrazně převažují inserce sdílené dvojicí poddruhů *M. m. domesticus* a *M. m. castaneus*. Toto ovšem může být mírně zavádějící vzhledem k faktu, že na rozdíl od B1 retrotransposonů byl v případě L1 testován jen jeden kmen odvozený z poddruhu *M. m. musculus*. I přesto je však počet sdílených insercí mezi *M. m. domesticus* a *M. m. castaneus* relativně vysoký. Zajímavé je, že počty sdílených insercí mezi dvojicemi poddruhů odpovídají počtům zjištěným Stehlíkovou (2009) v případě B2 retrotransposonů. Konkrétně poddruhy *M. m. domesticus* a *M. m. castaneus* měly společných 9 insercí, poddruhy *M. m. domesticus* a *M. m. musculus* 2 inserce a *M. m. castaneus* a *M. m. musculus* 3 inserce. Na základě těchto dat lze tedy říci, že poddruhy *M. m. domesticus* a *M. m. castaneus* jsou si mnohem příbuznější, než kterýkoliv z těchto dvou poddruhů vzhledem k *M. m. musculus*. Toto schéma je konzistentní s

výsledky fylogenetické analýzy založené na 38 autosomálních protein-kódujících lokusech (Din et al. 1993). V ostatních pracích, které následovaly později, však schéma evoluce jednotlivých poddruhů vychází odlišně (Tucker et al. 2005, White et al. 2009). V obou studiích se jeví *M. m. domesticus* jako poddruh, který se odštěpil jako první, a *M. m. musculus* a *M. m. castaneus* jsou tedy příbuznější. Ovšem jak ukazuje studie White et al. (2009), různé části genomu vykazují odlišnou evoluční historii, která je navíc mnohdy smazána recentní introgresí v hybridní zóně (Macholán et al. 2007, Teeter et al. 2009). Není proto snadné určit, co vlastně sdílení insercí u *M. m. domesticus* a *M. m. castaneus* odráží. U B1 podrodiny toto však neodpovídá ani jednomu z výše uvedených schémat, vzhledem k celkově nižšímu počtu testovaných insercí je však možné, že se jedná čistě o náhodu.

## 6.4 PŘÍTOMNOST INSERCÍ U *M. SPRETUS* A *M. MACEDONICUS*

Zajímavým výsledkem laboratorní části bylo, že inserce, které byly nalezeny jako variabilní v rámci komplexu *M. musculus*, byly přítomny i u vzdálenějších taxonů jako *M. spretus* a *M. macedonicus*. Boursot et al. (1993) datuje vznik těchto dvou druhů do doby před 1 milionem let. Recentnější studie Chevret et al. (2005) datovala vznik *M. spretus* mnohem více do minulosti na asi 2 miliony let. Geraldés et al. (2008) odhadovali dobu vzniku poddruhů v rámci *M. musculus*. Podle výsledků této studie všechny tři poddruhy vznikly ve velmi krátké době v době asi před 500 000 lety. Je tedy dost pravděpodobné, že vzhledem k velmi krátké době, která dělí od sebe tyto dvě události, může být přítomnost insercí u *M. spretus* a *M. macedonicus* výsledkem přetrvávání tzv. ancestrálního polymorfismu vlivem nekompletního lineage sortingu. Tedy znamená to, že inserce vznikla už u společného předka těchto druhů a vlivem náhodných procesů (genetický drift) došlo k současnému náhodnému výskytu u studovaných druhů, který však v žádném případě neodráží fylogenezi těchto druhů (Shedlock et al. 2004). Recentní studie zaměřená na celogenomovou fylogenetickou analýzu tří poddruhů *M. musculus* prokázala, že různé části genomu vykazují různý evoluční původ právě díky nekompletnímu lineage sortingu a také recentní introgresi (White et al. 2009). Dalšími možnostmi, které mohou být zodpovědné za sledovaný pattern, jsou podle Ray et al. (2006): paralelní inserce, excise inserce a přítomnost paralogního lokusu. Všechny tři byly zmíněny v úvodní kapitole věnované problematice retrotransposonů jako markerů.



V souvislosti s mými výsledky bych se zastavil u možnosti, kdy retrotransposon byl přítomen u všech testovaných kmenů. Takovéto inserce byly nalezeny celkem 3 (1 B1 a 2 L1). Přestože tyto inserce nebyly nikterak dále testovány, lze na základě přítomnosti u všech testovaných kmenů říci, že by se mohlo jednat o excisi elementu. Je však zapotřebí následné testování, které by ověřilo, zdali se skutečně jedná o tento případ.

V ostatních případech, kdy byla inserce přítomna u *M. spretus* nebo *M. macedonicus*, byla zároveň její přítomnost variabilní u testovaných kmenů v rámci komplexu *M. musculus*. Zde by bylo možné také uvažovat všechny výše zmíněné možnosti. Lze se však domnívat, že se u většiny studovaných insercí jedná o první případ – přetrvávání ancestrálního polymorfismu – vzhledem k faktu, že zastoupení insercí, které jsou přítomny vně komplexu *M. musculus*, se liší mezi jednotlivými podrodinami (T vs. F2 a T vs. F3). Je dost málo pravděpodobné, že by se míra posledních tří zmiňovaných procesů lišila mezi podrodinami. Tento fakt spíše ukazuje na skutečnost, že jednotlivé podrodiny byly aktivní v různou dobu v průběhu evoluce zkoumaných myších druhů a sledovaný pattern ukazuje na neúplný lineage sorting u podrodin, které byly přítomny již u společného předka *M. musculus*, *M. spretus* a *M. macedonicus*. Dále tedy budu předpokládat, že sdílení insercí s druhy *M. spretus* a *M. macedonicus* znamená časnější aktivitu těchto insercí.

## 6.5 VYUŽITÍ L1 A B1 RETROTRANSPONŮ JAKO MARKERŮ V POPULAČNÍ GENETICE

Stejně jako Stehlíková (2009) i já jsem v mé diplomové práci hodnotil využitelnost retrotransponů v populační genetice myši domácí. Většina nalezených insercí byla polymorfní v rámci komplexu *M. musculus*. Řada z nich byla diagnostická jen pro jednotlivé poddruhy s tím, že nejvíce takovýchto insercí patřilo mezi L1 retrotransposony a bylo diagnostických jen pro poddruh *M. m. domesticus*. Bylo zjištěno, že je nutné takto nalezené inserce testovat i u blízce příbuzných druhů z důvodu vyloučení možnosti nekompletního lineage sortingu, jehož signál by smazal skutečnou fylogenezi. V případě L1 retrotransponů se jeví jako vhodnější markery pro populační genetiku myši domácí zástupci podrodin A a T, protože obsahují mnohem nižší zastoupení insercí sdílených s *M. spretus* a *M. macedonicus* a zároveň mnohem větší

zastoupení insercí poddruhově specifických. Rodina B1 měla insercí tohoto typu jen minimálně. Přes obtíže způsobené nekompletním lineage sortingem a dalšími jevy zmíněnými v úvodní kapitole věnované využití retrotransposonů jako markerů mají retrotransposony stále značné výhody oproti jiným typům markerů, které z nich činí stále velmi mocný nástroj při analýze populační struktury a fylogeneze různých organismů.

## 6.6 ZHODNOCENÍ AKTIVITY JEDNOTLIVÝCH PODRODIN B1 RETROTRANSPOSONŮ

Celkem bylo získáno 49 insercí, jež lze považovat za inserce, které vznikly v nedávné době aktivitou retrotransposonů. Tyto inserce patří s různým zastoupením do 4 podrodin. Na základě výsledků fylogeneze konsenzuálních sekvencí lze podrodiny Mus1 a Mm považovat za nejmladší. Přesto však je nejvíce zastoupena podrodina Mus2 s 23 kopiemi, pak Mm s 20 kopiemi a v případě podrodiny Mus1 bylo nalezeno pouze 5 insercí. Podrodina Mur4, která fylogeneticky vychází spíše jako starší podrodina, u níž se předpokládá aktivita v celé Muroidní linii (záznamy RepBase viz příloha 1), byla zastoupena pouze 1 insercí.

Vzhledem k nízkému počtu kopií v případě podrodin Mur4 a Mus1, nebylo možné se těmito skupinami zabývat hlouběji. Proto byla pozornost koncentrována spíše na podrodiny Mus2 a MM. Na základě analýzy Median-Joining Networks se mi jeví jako pravděpodobné, že podrodina Mus2 předcházela v evoluci podrodinu Mm, což ostatně předpokládal i Quentin (1989) v případě podrodin B1D, B1A a B1B. Otázkou by mohlo být, zdali každá z těchto dvou podrodin má v genomu svoji vlastní zdrojovou kopii, popř. jestli se Mm postupně vyvinula z Mus2 a obě měly tedy stejný zdroj, tak jak to uvádí Deininger et al. (1992). Z dostupných dat vyplývá, že v případě podrodin Mus2 a Mm se jedná s největší pravděpodobností o první hypotézu. Tedy o případ, kdy každá podrodina má svoji alespoň jednu zdrojovou kopii. Předpokládáme-li, že aktivita jednotlivých podrodin je víceméně stejná – tzn. že za určitou časovou jednotku vyprodukuje stejné (podobné) množství kopií – pak lze na základě víceméně stejného počtu nalezených insercí říci, že tyto dvě podrodiny byly (a jsou) aktivní ve stejnou dobu. Čemuž do určité míry odpovídá fakt, že Fisherův exaktní test srovnávající poměr

insercí přítomných uvnitř a vně komplexu *M. musculus* vyšel nesignifikantně. Na druhou stranu však párová srovnání rozdílů ukázala, že sekvence v rámci podrodiny Mm jsou méně odlišné, než v rámci podrodiny Mus2. Jen jediná sekvence 19\_37506476|MSMg01-237L20.TJ patřící do podrodiny Mm vycházela ve všech analýzách značně odchyleně. Tato sekvence by mohla být odvozena od jiné zdrojové kopie nebo mohla projít v nedávné době například genovou konverzí. Ta byla prokázána u Alu retrotransposonů ve studii Kass et al. (1995). Vyloučíme-li z našich úvah tuto konkrétní sekvenci, pak se podrodina Mm zdá být trochu mladší než podrodina Mus2. To *de facto* odpovídá výše zmiňované hypotéze, že podrodina Mm byla odvozena od podrodiny Mus2 s tím, že obě mají svoji vlastní zdrojovou kopii a jsou v genomu myši aktivní současně. Toto se shoduje i s výsledky Kass et al. (2000). Pro podrodiny B1A a B1B byla odhadnuta identická doba aktivity. V případě podrodiny B1C, která odpovídá Mus2, se doba aktivity překrývala s předchozími dvěma, avšak šla více do minulosti.

## 6.7 ZHODNOCENÍ AKTIVITY JEDNOTLIVÝCH PODRODIN L1 RETROTRANSPOSONŮ

V případě L1 retrotransposonů bylo získáno 116 insercí, které byly rozřazeny do 5 podrodin. Podrodina L1\_Mm, která nebyla definována ve studii Mears & Hutchison (2001) a byla zastoupena pouze 1 zástupcem, byla z analýz vyloučena. Ostatní podrodiny byly zastoupeny v různé míře, nejvíce však podrodiny A a T s 43 a 41 kopiemi, následně pak podrodina F2 s 23 a F3 s 8 kopiemi. Vzhledem k faktu, že byly dostupné pouze na 5' konci zkrácené sekvence L1 a zařazení do podrodin bylo provedeno pomocí programu RepeatMasker na základě sekvenční podobnosti ke konsensu, nebylo možné zařadit retrotransposony jemněji tak, jak je definoval Mears & Hutchison (2001). Podle těchto autorů jsou v genomu myši aktivní 3 podrodiny (T<sub>F2-II</sub>, T<sub>F1-I</sub> a A-I). RepeatMasker však na základě sekvenční podobnosti umožňuje zařazení pouze do podrodin A a T. Tyto dvě podrodiny mají největší zastoupení. Také poměr insercí přítomných uvnitř komplexu *M. musculus* a vně tohoto komplexu ukazuje, že A a T jsou ve srovnání F2 a F3 mnohem mladší. Podrodina T má zastoupení insercí přítomných u *M. spretus* a *M. macedonicus* minimální. Rozdíl mezi podrodinou T a podrodinami F2 a F3 vychází na základě Fisherova exaktního testu signifikantní.

V případě podrodiny A se hodnota pravděpodobnosti blíží hladině signifikance 0,05 při srovnání s podrodinou F3. Pokud bychom předpokládali, že všechny nalezené polymorfní podrodiny jsou aktivní, znamenalo by to, že kromě podrodin A a T jsou zároveň aktivní i staré podrodiny F2 a F3. Přesto však na základě mých výsledků, vzhledem k nízkým počtům a zastoupení u *M. spretus* a *M. macedonicus*, se jeví přítomnost těchto dvou podrodin spíše jako reliktní. Jedná se tedy o přetrvávající ancestrální polymorfismus, což v případě podrodiny F2 víceméně odpovídá datování podle Mears & Hutchison (2001), kde FII<sub>2</sub> má počátek aktivity před asi 1,4 milionem let, tedy zhruba v době, kdy vznikly studované druhy (Boursot et al. 1993, Chevret et al. 2005). Počátek aktivity podrodiny FIII (v databázi RepBase F3) je datován mnohem více do minulosti (asi před 4,1 milionem let). Otázka však je, jak dlouho její aktivita přetrvávala. Pokud je možné brát přítomnost F3 polymorfních kopií za nedávnou aktivitu této podrodiny, pak aktivita musela být přerušena ne příliš hluboko v minulosti. Vzhledem k výrazně nižšímu počtu nalezených polymorfních kopií a také vyššímu procentu sdílených kopií s *M. spretus* a *M. macedonicus*, než v případě podrodiny F2, lze říci, že poslední aktivita pravděpodobně spadá hlouběji do minulosti. Na základě mých výsledků však nelze určit přesnější datování.

## 6.8 POROVNÁNÍ AKTIVITY MEZI RODINAMI L1, B1 A B2 RETROTRANSPOSONŮ

Přestože počty insercí jsou značně ovlivněny zkrácením, které vzniklo metodou, jakou byly získány, lze na základě přítomnosti těchto insercí u jednotlivých testovaných kmenů srovnat recentní aktivitu mezi jednotlivými skupinami retroelementů. Z mých výsledků vyplývá, že obě nejvíce početné podrodiny B1 retrotransposonů jsou velmi výrazně zastoupeny u *M. spretus* a *M. macedonicus*. V případě L1 retrotransposonů podobné zastoupení u výše zmiňovaných druhů mají podrodiny, o kterých je možno na základě mých výsledků předpokládat, že jsou již neaktivní. Zároveň je také zajímavé, že zastoupení všech testovaných B1 a L1 vně komplexu *M. musculus* se na základě Fisherova exaktního testu blíží hodnotě signifikance. B1 a L1 retrotransposony se taky výrazně liší počty insercí specifických u *M. m. domesticus*. Tento rozdíl je dvojnásobný pokud neuvažují inserce nalezené pouze u kmene C57BL/6J, které nebyly přítomny u

ostatních testovaných kmenů. Jedná se s největší pravděpodobností o inserce, které jsou nové, a nestihly se rozšířit v populaci *M. m. domesticus*, popřípadě jsou unikátní jen u tohoto kmene. Takovýchto insercí nalezených pouze u C57BL/6J je v případě L1 výrazně větší množství než v případě B1. Na základě těchto výsledků se domnívám, že v současné době jsou L1 retrotransposony výrazně aktivnější než B1 retrotransposony. Na základě velkého zastoupení u *M. spretus* a *M. macedonicus* a jen minimálnímu počtu *M. m. domesticus* specifických insercí by mohlo docházet dokonce u *M. musculus* k poklesu aktivity B1 retrotransposonů. Mohlo by se jednat i o výskyt reliktní, tak jak je tomu v případě podrodin F2 a F3 u L1 retrotransposonů. Přesto však fakt, že v případě podrodiny Mm byly nalezeny 3 inserce bez jakýchkoliv záměn po vyloučení CpG párů, naznačuje, že byly aktivní v nedávné době. Proti reliktnímu výskytu B1 navíc svědčí výsledky několika srovnávacích studií ukazující, že ve srovnání s genomem potkana narostl počet B1 retroelementů v genomu myši dvojnásobně (Waterson et al. 2002, RGSPR 2005, Veniaminova et al. 2007b). Přesto však může docházet v současnosti u myši domácí k poklesu jejich aktivity.

Podobný pattern, jako byl nalezen u L1 retrotransposonů, nalezná i Stehlíková (2009) u B2 retrotransposonů. Velké množství insercí bylo přítomno pouze u *M. m. domesticus* popř. pouze u kmene C57BL/6J. V případě B2 retrotransposonů bylo velké množství přítomno i u poddruhu *M. m. musculus*. Otázkou je, zdali je tento výskyt dán faktem, že B2 jsou kratší než L1 a tedy bylo nalezeno větší množství B2 insercí u kmene MSM/Ms než L1 insercí, anebo je aktivita L1 koncentrována spíše u poddruhu *M. m. domesticus*, zatímco aktivita B2 je vysoká v rámci celého komplexu *M. musculus*. S přihlédnutím k výsledkům Stehlíkové (2009) lze říci, že rodiny L1 a B2 jsou v současnosti velmi aktivní skupiny retrotransposonů, což však nelze říci o B1, přestože jejich velmi recentní aktivitu na základě mých výsledků nelze popřít. Mé výsledky společně s výsledky Stehlíkové (2009) korespondují s výsledky studie Akagi et al. (2008). Na základě této studie, která srovnávala genomy několika laboratorních inbredních kmenů s cílem najít recentní inserce, bylo zjištěno, že právě rodiny L1 a B2 početně velmi převažují nad ostatními skupinami retrotransposonů. Rodina B1 má spíše minoritní zastoupení (obrázek 7). Už Kass et al. (1997) srovnáním počtů kopií patřící do rodiny B2 u různých druhů hlodavců zjistili, že se aktivita těchto elementů značně zvýšila v linii vedoucí směrem k rodu *Mus*, což odpovídá výsledkům Stehlíkové (2009).

V současné době jsou tedy neaktivnější v genomu myši 2 rodiny retrotransponů – jsou to L1 retrotransposony kódující si vlastní enzymatický aparát a skupina krátkých B2 retrotransponů, které zřejmě využívají vysoké aktivity L1 k vlastní propagaci. Rodina B1 retrotransponů s největší pravděpodobností aktivní je, přesto však zřejmě není schopná obstát v konkurenci s B2 retrotransposony. V případě, že by si B2 retrotransposony dokázaly lépe podmanit enzymatický aparát L1 a zároveň by narostla jejich transkripční aktivita, došlo by s největší pravděpodobností k potlačení nárůstu nových insercí B1 retrotransponů v genomu, přestože jejich zdrojová kopie by neustále produkovala nové B1 transkripty. Ovšem toto je jen hypotetický scénář, který bude nutno dále testovat.

## 7. SHRNU TÍ

---

Cílem práce bylo nalezení polymorfních insercí vzniklých recentní aktivitou dvou rodin non-LTR retrotransposonů (L1 a B1) a následné posouzení jejich aktivity společně s daty Stehlíkové (2009), která se ve své diplomové práci zaměřila na rodinu B2. Srovnání tohoto typu doposud nebylo žádným autorem přede mnou provedeno. Mé výsledky ukázaly, že v rámci obou rodin – B1 i L1 retrotransposonů – bylo v nedávné době popř. stále je aktivních několik podrodin. V případě L1 retrotransposonů byly nalezeny polymorfní inserce patřící do 5 podrodin, z toho pouze 2 podrodiny (T a A) byly předchozími autory shledány jako v současnosti aktivní. Přes polymorfní přítomnost podrodin F2 a F3 v rámci komplexu *M. musculus* nasvědčovala i má data, že spíše nežli o skutečnou aktivitu se jedná o relikty, které přetrvávají vlivem ancestrálního polymorfismu už z doby před vznikem *M. spretus* a *M. macedonicus*. V případě nových podrodin T a A byly inserce zase naopak přítomny pouze u poddruhu *M. m. domesticus*, což potvrzuje jejich současnou aktivitu. Zástupců podrodiny B1 bylo nalezeno mnohem méně. Nejvíce zastoupeny byly dvě podrodiny, jejichž zástupci byli rovnoměrně zastoupeni jak u *M. spretus* a *M. macedonicus*, tak i pouze u *M. m. domesticus*. Z tohoto lze vyvodit, že aktivita spadá do doby před vznikem druhů *M. spretus* a *M. macedonicus*, ale také, že některé inserce vznikly teprve nedávno. Vzhledem k celkově nižšímu počtu kopií B1, se domnívám, že přestože aktivita jednotlivých podrodin trvala po delší dobu, než v případě L1 retrotransposonů, celková aktivita L1 je výrazně vyšší, narůstající u jednotlivých poddruhů *M. musculus*. Při srovnání s daty Stehlíkové (2009) mohu zkonstatovat, že spolu s aktivitou L1 retrotransposonů, narůstala i aktivita B2 retrotransposonů. Vzhledem k faktu, že B1 a B2 retrotransposony využívají stejný mechanismus vlastní propagace, je dost pravděpodobné, že B2 obstály v konkurenci o L1 enzymatický aparát lépe než B1 retrotransposony, a původcem současného nízkého počtu B1 kopií může být právě nárůst aktivity B2 elementů.

# PŘEHLED LITERATURY

---

Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K., & Shiroishi, T. 2004. Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis. *Genome Res.*, 14(12): 2439-2447.

Adey, N. B., Schichman, S. A., Graham, D. K., Peterson, S. N., Edgell, M. H., & Hutchison, C. A., III 1994. Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Mol.Biol.Evol.*, 11(5): 778-789.

Adkins, R. M., Walton, A. H., & Honeycutt, R. L. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol.Phylogenet.Evol.*, 26(3): 409-420.

Akagi, K., Li, J., Stephens, R. M., Volfovsky, N., & Symer, D. E. 2008. Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Res.*, 18(6): 869-880.

Bamshad, M. J., Wooding, S., Watkins, W. S., Ostler, C. T., Batzer, M. A., & Jorde, L. B. 2003. Human population genetic structure and inference of group membership. *Am.J.Hum.Genet.*, 72(3): 578-589.

Bandelt, H. J., Forster, P., & Rohl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol.Biol.Evol.*, 16(1): 37-48.

Batzer, M. A. & Deininger, P. L. 1991. A human-specific subfamily of Alu sequences. *Genomics*, 9(3): 481-487.

Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J., & . 1994. African origin of human-specific polymorphic Alu insertions. *Proc.Natl.Acad.Sci.U.S.A.*, 91(25): 12288-12292.

Boursot, P., Auffray, J.-C., Britton-Davidian, J., & Bonhomme, F. 1993. The evolution of house mice. *Annual Review of Ecology and Systematics*, 24: 119-152.

Burt, A. & Trivers, R. 2006. *Genes in conflict*. Cambridge, Massachusetts London, England : The Belknap Press of Harvard University Press.

Casavant, N. C. & Hardies, S. C. 1994. The dynamics of murine LINE-1 subfamily amplification. *J.Mol.Biol.*, 241(3): 390-397.

Chevret, P., Veyrunes, F., & Britton-Davidian, J. 2005. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biological Journal of the Linnean Society*, 84: 417-427.



- Churakov, G., Sadasivuni, M. K., Rosenbloom, K. R., Huchon, D., Brosius, J., & Schmitz, J. 2010. Rodent evolution: back to the root. *Mol.Biol.Evol.*, 27(6): 1315-1326.
- Cordaux, R., Hedges, D. J., & Batzer, M. A. 2004. Retrotransposition of Alu elements: how many sources? *Trends Genet.*, 20(10): 464-467.
- Daniels, G. R. & Deininger, P. L. 1985. Repeat sequence families derived from mammalian tRNA genes. *Nature*, 317(6040): 819-822.
- Deininger, P. L., Moran, J. V., Batzer, M. A., & Kazazian, H. H., Jr. 2003. Mobile elements and mammalian genome evolution. *Curr.Opin.Genet.Dev.*, 13(6): 651-658.
- Din, W., Anand, R., Boursot, P., Darviche, D., Dod, B., Jouvin-Marche, E., Orth, A., Talwar, G. P., Cazenave, P.-A., & Bonhomme, F. 1996. Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology*, 9(5): 519-539.
- Dobigny, G., Ozouf-Costaz, C., Waters, P. D., Bonillo, C., Coutanceau, J. P., & Volobouev, V. 2004. LINE-1 amplification accompanies explosive genome repatterning in rodents. *Chromosome.Res.*, 12(8): 787-793.
- Duncan, C. H., Jagadeeswaran, P., Wang, R. R., & Weissman, S. M. 1981. Structural analysis of templates and RNA polymerase III transcripts of Alu family sequences interspersed among the human beta-like globin genes. *Gene*, 13(2): 185-196.
- Elder, J. T., Pan, J., Duncan, C. H., & Weissman, S. M. 1981. Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA. *Nucleic Acids Res.*, 9(5): 1171-1189.
- Feng, Q., Moran, J. V., Kazazian, H. H., Jr., & Boeke, J. D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, 87(5): 905-916.
- Frazer, K. A., Eskin, E., Kang, H. M., Bogue, M. A., Hinds, D. A., Beilharz, E. J., Gupta, R. V., Montgomery, J., Morenzoni, M. M., Nilsen, G. B., Pethiyagoda, C. L., Stuve, L. L., Johnson, F. M., Daly, M. J., Wade, C. M., & Cox, D. R. 2007. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*, 448(7157): 1050-1053.
- Fuhrman, S. A., Deininger, P. L., LaPorte, P., Friedmann, T., & Geiduschek, E. P. 1981. Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res.*, 9(23): 6439-6456.
- Geraldes, A., Basset, P., Gibson, B., Smith, K. L., Harr, B., Yu, H. T., Bulatova, N., Ziv, Y., & Nachman, M. W. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol.Ecol.*, 17(24): 5349-5363.
- Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., Scott, G., Steffen, D., Worley, K. C., Burch, P. E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Celera, Holt, R. A., Adams, M. D., Amanatides, P. G., Baden-Tillson, H., Barnstead, M., Chin, S., Evans, C. A., Ferriera, S., Fosler, C., Glodek, A., Gu, Z., Jennings, D., Kraft, C. L., Nguyen, T., Pfannkoch, C. M., Sitter, C., Sutton, G. G., Venter, J. C., Woodage, T., Smith, D., Lee, H. M., Gustafson, E.,

Cahill, P., Kana, A., Doucette-Stamm, L., Weinstock, K., Fecht, K., Weiss, R. B., Dunn, D. M., Green, E. D., Blakesley, R. W., Bouffard, G. G., De Jong, P. J., Osoegawa, K., Zhu, B., Marra, M., Schein, J., Bosdet, I., Fjell, C., Jones, S., Krzywinski, M., Mathewson, C., Siddiqui, A., Wye, N., McPherson, J., Zhao, S., Fraser, C. M., Shetty, J., Shatsman, S., Geer, K., Chen, Y., Abramzon, S., Nierman, W. C., Havlak, P. H., Chen, R., Durbin, K. J., Egan, A., Ren, Y., Song, X. Z., Li, B., Liu, Y., Qin, X., Cawley, S., Worley, K. C., Cooney, A. J., D'Souza, L. M., Martin, K., Wu, J. Q., Gonzalez-Garay, M. L., Jackson, A. R., Kalafus, K. J., McLeod, M. P., Milosavljevic, A., Virk, D., Volkov, A., Wheeler, D. A., Zhang, Z., Bailey, J. A., Eichler, E. E., Tuzun, E., Birney, E., Mongin, E., Ureta-Vidal, A., Woodwark, C., Zdobnov, E., Bork, P., Suyama, M., Torrents, D., Alexandersson, M., Trask, B. J., Young, J. M., Huang, H., Wang, H., Xing, H., Daniels, S., Gietzen, D., Schmidt, J., Stevens, K., Vitt, U., Wingrove, J., Camara, F., Mar, A. M., Abril, J. F., Guigo, R., Smit, A., Dubchak, I., Rubin, E. M., Couronne, O., Poliakov, A., Hubner, N., Ganten, D., Goesele, C., Hummel, O., Kreitler, T., Lee, Y. A., Monti, J., Schulz, H., Zimdahl, H., Himmelbauer, H., Lehrach, H., Jacob, H. J., Bromberg, S., Gullings-Handley, J., Jensen-Seaman, M. I., Kwitek, A. E., Lazar, J., Pasko, D., Tonellato, P. J., Twigger, S., Ponting, C. P., Duarte, J. M., Rice, S., Goodstadt, L., Beatson, S. A., Emes, R. D., Winter, E. E., Webber, C., Brandt, P., Nyakatura, G., Adetobi, M., Chiaromonte, F., Elnitski, L., Eswara, P., Hardison, R. C., Hou, M., Kolbe, D., Makova, K., Miller, W., Nekrutenko, A., Riemer, C., Schwartz, S., Taylor, J., Yang, S., Zhang, Y., Lindpaintner, K., Andrews, T. D., Caccamo, M., Clamp, M., Clarke, L., Curwen, V., Durbin, R., Eyrales, E., Searle, S. M., Cooper, G. M., Batzoglu, S., Brudno, M., Sidow, A., Stone, E. A., Venter, J. C., Payseur, B. A., Bourque, G., Lopez-Otin, C., Puente, X. S., Chakrabarti, K., Chatterji, S., Dewey, C., Pachter, L., Bray, N., Yap, V. B., Caspi, A., Tesler, G., Pevzner, P. A., Haussler, D., Roskin, K. M., Baertsch, R., Clawson, H., Furey, T. S., Hinrichs, A. S., Karolchik, D., Kent, W. J., Rosenbloom, K. R., Trumbower, H., Weirauch, M., Cooper, D. N., Stenson, P. D., Ma, B., Brent, M., Arumugam, M., Shteynberg, D., Copley, R. R., Taylor, M. S., Riethman, H., Mudunuri, U., Peterson, J., Guyer, M., Felsenfeld, A., Old, S., Mockrin, S., & Collins, F. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428(6982): 493-521.

Goodier, J. L., Ostertag, E. M., Du, K., & Kazazian, H. H., Jr. 2001. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.*, 11(10): 1677-1685.

Gregory, T. R. 2005. *The Evolution of the Genome*. London, UK: Elsevier Academic Press.

Han, K., Xing, J., Wang, H., Hedges, D. J., Garber, R. K., Cordaux, R., & Batzer, M. A. 2005. Under the genomic radar: the stealth model of Alu amplification. *Genome Res.*, 15(5): 655-664.

Hirose, Y. & Manley, J. L. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*, 395(6697): 93-96.

Hurst, G. D. & Werren, J. H. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nat.Rev.Genet.*, 2(8): 597-606.

Ichiyanagi, K., Nishihara, H., Duvernell, D. D., & Okada, N. 2007. Acquisition of endonuclease specificity during evolution of L1 retrotransposon. *Mol.Biol.Evol.*, 24(9): 2009-2015.

- International Human Genome Sequencing Consortium 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011): 931-945.
- Kass, D. H., Batzer, M. A., & Deininger, P. L. 1995. Gene conversion as a secondary mechanism short interspersed element (SINE) evolution. *Mol. Cel. Evol.* 15(1): 19-25
- Kass, D. H., Kim, J., Rao, A., & Deininger, P. L. 1997. Evolution of B2 repeats: the muroid explosion. *Genetica*, 99(1): 1-13.
- Kass, D. H., Raynor, M. E., & Williams, T. M. 2000. Evolutionary history of B1 retroposons in the genus *Mus*. *J.Mol.Evol.*, 51(3): 256-264.
- Kass, D. H. & Jamison, N. 2007. Identification of an active ID-like group of SINEs in the mouse. *Genomics*, 90(3): 416-420.
- Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P., & Georgiev, G. P. 1979. Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. *Nucleic Acids Res.*, 6(2): 697-713.
- Kurose, K., Hata, K., Hattori, M., & Sakaki, Y. 1995. RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. *Nucleic Acids Res.*, 23(18): 3704-3709.
- Lewin, B. 2004. *Genes VIII*. Pearson Prentice Hall.
- Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, 72(4): 595-605.
- Macholan, M., Munclinger, P., Sugerkova, M., Dufkova, P., Bimova, B., Bozikova, E., Zima, J., & Pialek, J. 2007. Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, 61(4): 746-771.
- Martin, S. L. 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol.Cell Biol.*, 11(9): 4804-4807.
- Mathias, S. L., Scott, A. F., Kazazian, H. H., Jr., Boeke, J. D., & Gabriel, A. 1991. Reverse transcriptase encoded by a human transposable element. *Science*, 254(5039): 1808-1810.
- Mears, M. L. & Hutchison, C. A., III 2001. The evolution of modern lineages of mouse L1 elements. *J.Mol.Evol.*, 52(1): 51-62.
- Mizrokhi, L. J., Georgieva, S. G., & Ilyin, Y. V. 1988. jockey, a mobile *Drosophila* element similar to mammalian LINES, is transcribed from the internal promoter by RNA polymerase II. *Cell*, 54(5): 685-691.

- Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., & Kazazian, H. H., Jr. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell*, 87(5): 917-927.
- Moran, J. V., DeBerardinis, R. J., & Kazazian, H. H., Jr. 1999. Exon shuffling by L1 retrotransposition. *Science*, 283(5407): 1530-1534.
- Munclinger, P., Boursot, P., & Dod, B. 2003. B1 insertions as easy markers for mouse population studies. *Mamm.Genome*, 14(6): 359-366.
- Ohshima, K., Hamada, M., Terai, Y., & Okada, N. 1996. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Mol.Cell Biol.*, 16(7): 3756-3764.
- Osheim, Y. N., Proudfoot, N. J., & Beyer, A. L. 1999. EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol.Cell*, 3(3): 379-387.
- Ostertag, E. M. & Kazazian, H. H., Jr. 2001. Biology of mammalian L1 retrotransposons. *Annu.Rev.Genet.*, 35: 501-538.
- Padgett, R. W., Hutchison, C. A., III, & Edgell, M. H. 1988. The F-type 5' motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. *Nucleic Acids Res.*, 16(2): 739-749.
- Pascale, E., Valle, E., & Furano, A. V. 1990. Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. *Proc.Natl.Acad.Sci.U.S.A.*, 87(23): 9481-9485.
- Quentin, Y. 1989. Successive waves of fixation of B1 variants in rodent lineage history. *J.Mol.Evol.*, 28(4): 299-305.
- Quentin, Y. 1992. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.*, 20(3): 487-493.
- Quentin, Y. 1992. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.*, 20(3): 487-493.
- Quentin, Y. 1994. A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res.*, 22(12): 2222-2227.
- Ray, D. A., Xing, J., Salem, A. H., & Batzer, M. A. 2006. SINEs of a nearly perfect character. *Syst.Biol.*, 55(6): 928-935.
- Rebuzzini, P., Castiglia, R., Nergadze, S. G., Mitsainas, G., Munclinger, P., Zuccotti, M., Capanna, E., Redi, C. A., & Garagna, S. 2009. Quantitative variation of LINE-1 sequences in five species and three subspecies of the subgenus *Mus* and in five Robertsonian races of *Mus musculus domesticus*. *Chromosome.Res.*, 17(1): 65-76.

- Rikke, B. A., Garvin, L. D., & Hardies, S. C. 1991. Systematic identification of LINE-1 repetitive DNA sequence differences having species specificity between *Mus spretus* and *Mus domesticus*. **J.Mol.Biol.**, 219(4): 635-643.
- Rogers, J. H. 1985. The origin and evolution of retroposons. **Int.Rev.Cytol.**, 93: 187-279.
- Sarich, V. M. 1985. Rodent macromolecular systematics. *Evolutionary relationships among rodents*: 423-462. New York: Plenum Press.
- Saxton, J. A. & Martin, S. L. 1998. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. **J.Mol.Biol.**, 280(4): 611-622.
- Serdobova, I. M. & Kramerov, D. A. 1998. Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. **J.Mol.Evol.**, 46(2): 202-214.
- Shedlock, A. M. & Okada, N. 2000. SINE insertions: powerful tools for molecular systematics. **Bioessays**, 22(2): 148-160.
- Sherry, S. T., Harpending, H. C., Batzer, M. A., & Stoneking, M. 1997. Alu evolution in human populations: using the coalescent to estimate effective population size. **Genetics**, 147(4): 1977-1982.
- Smit, A. F., Toth, G., Riggs, A. D., & Jurka, J. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. **J.Mol.Biol.**, 246(3): 401-417.
- Smit, A. F. 1996. The origin of interspersed repeats in the human genome. **Curr.Opin.Genet.Dev.**, 6(6): 743-748.
- Stehlíková, J. 2009. *Vlastnosti recentně aktivních B2 inzercí v myším genomu*. Katedra zoologie, PřF UK v Praze.
- Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., Jenkins, T., Tahir, M. A., Deininger, P. L., & Batzer, M. A. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. **Genome Res.**, 7(11): 1061-1071.
- Swergold, G. D. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. **Mol.Cell Biol.**, 10(12): 6718-6729.
- Tarlinton, R. E., Meers, J., & Young, P. R. 2006. Retroviral invasion of the koala genome. **Nature**, 442(7098): 79-81.
- Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E., Krenz, J. G., Sans-Fuentes, M. A., Nachman, M. W., & Tucker, P. K. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. **Genome Res.**, 18(1): 67-76.
- Tucker, P. K., Sandstedt, S. A., & Lundrigan, B. L. 2005. Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): examining gene trees and species trees. **Biological Journal of the Linnean Society**, 84: 653-662.

Veniaminova, N. A., Vasetskii, N. S., Lavrechenko, L. A., Popov, S. V., & Kramerov, D. A. 2007a. Phylogeny of the order Rodentia inferred from structural analysis of short retrotransposon B1. *Genetika*, 43(7): 916-929.

Veniaminova, N. A., Vassetzky, N. S., & Kramerov, D. A. 2007b. B1 SINES in different rodent families. *Genomics*, 89(6): 678-686.

Veniaminova, N. A., Gogolevskii, K. P., Vasetskii, N. S., & Kramerov, D. A. 2007c. Comparative analysis of copy number of ID and B1 short retroposons in rodent genomes. *Mol.Biol.(Mosk)*, 41(6): 1081-1084.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyraas, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C., & Lander, E. S. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915): 520-562.

Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A., Harpending, H. C., Rogers, A. R., & Jorde, L. B. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am.J.Hum.Genet.*, 68(3): 738-752.

Watkins, W. S., Rogers, A. R., Ostler, C. T., Wooding, S., Bamshad, M. J., Brassington, A. M., Carroll, M. L., Nguyen, S. V., Walker, J. A., Prasad, B. V., Reddy, P. G., Das, P. K., Batzer, M. A., & Jorde, L. B. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res.*, 13(7): 1607-1618.

Weiner, A. M. 2002. SINEs and LINEs: the art of biting the hand that feeds you. *Curr.Opin.Cell Biol.*, 14(3): 343-350.

White, M. A., Ane, C., Dewey, C. N., Larget, B. R., & Payseur, B. A. 2009. Fine-scale phylogenetic discordance across the house mouse genome. *PLoS.Genet.*, 5(11): e1000729.

Xing, J., Hedges, D. J., Han, K., Wang, H., Cordaux, R., & Batzer, M. A. 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. *J.Mol.Biol.*, 344(3): 675-682.

Yang, H., Bell, T. A., Churchill, G. A., & Pardo-Manuel, d., V 2007. On the subspecific origin of the laboratory mouse. *Nat.Genet.*, 39(9): 1100-1107.

# PŘÍLOHY

---

## PŘÍLOHA 1: ORIGINÁLNÍ ZÁZNAMY B1 PODRODIN Z REPBASE

```
ID B1F2      rebase;      DNA;      ROD; 149 BP.
XX
AC .
XX
DT 16-MAR-2006 (Rel. 11.03, Created)
DT 10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE SINE1 from Muridae.
XX
KW SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW B1F; B1F2; SINE1.
XX
OS Muridae
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Sciurognathi; Muroidea.
XX
RN [1]
RP 1-149
RA Smit A.F.;
RT "B1F2 - a subfamily of SINEs from Muridae.";
RL Direct Submission to Rebase Update (11-NOV-2005).
XX
DR [1] (Consensus)
XX
SQ Sequence 149 BP; 39 A; 41 C; 46 G; 23 T; 0 other;
agccggggcgt ggtggcacac gcctgtaatc ccagcactcg ggaggcagag gcaggaggat      60
ctccgtgagt tcgaggccag cctggtctac atagcgagtt ccaggccagc cagggctaca      120
tagtgagacc ctgtctcaaa aaaacaaaaa                                     149

//
ID B1F1      rebase;      DNA;      ROD; 149 BP.
XX
AC .
XX
DT 16-MAR-2006 (Rel. 11.03, Created)
DT 10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE SINE1 from Muridae.
XX
KW SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW B1F; B1F1; SINE1.
XX
OS Muridae
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Sciurognathi; Muroidea.
XX
RN [1]
RP 1-149
RA Smit A.F.;
RT "B1F1 - a subfamily of SINEs from Muridae.";
RL Direct Submission to Rebase Update (11-NOV-2005).
XX
DR [1] (Consensus)
XX
SQ Sequence 149 BP; 38 A; 40 C; 48 G; 23 T; 0 other;
```



```

agccgggcgt ggtggcacac gcctgtaatc ccagcactcg ggaggctgag gcaggaggat      60
cgccgtgagt tcgaggccag cctgggctac atagtgagtt ccaggccagc cagggtaca      120
tagtgagacc ctgtctcaaa aaaacaaaaa                                     149
//
ID B1F          rebase;      DNA;      ROD; 151 BP.
XX
AC .
XX
DT 20-JUL-1999 (Rel. 4.06, Created)
DT 20-JUL-1999 (Rel. 4.06, Last updated, Version 1)
XX
DE B1F repetitive element - an old subfamily consensus.
XX
KW SINE; Non-LTR Retrotransposon; Transposable Element; B1; B1F;
KW Repetitive sequence.
XX
OS Mus musculus
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Sciurognathi; Muroidea; Muridae; Murinae; Mus.
XX
RN [1]
RA Quentin Y.;
RT "Successive waves of fixation of B1 variants in rodent lineage
RT history.";
RL J.Mol.Evol 28, 299-305 (1989).
XX
RN [2]
RP 1-151
RA Jurka J.;
RT "B1F.";
RL Direct Submission to Repbase Update (JUL-1999).
XX
DR [2] (Consensus)
XX
CC This consensus differs slightly from Quentin's consensus.
XX
SQ Sequence 151 BP; 44 A; 38 C; 42 G; 27 T; 0 other;
gccaggcata gtggcacaca cctgtaatcc cagcacttag gaggctgagg caggaggatc      60
tgctgtgagt ttgaagccag cctgggctac atagtgagtt ccaggccagc ctgggctaca      120
cagtgagacc ctgtctcaaa aaacaaaaaa a                                     151
//
ID B1_Mur1      rebase;      DNA;      ROD; 149 BP.
XX
AC .
XX
DT 16-MAR-2006 (Rel. 11.03, Created)
DT 10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE SINE1 from Muridae.
XX
KW SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW B1; B1_Mur1; SINE1.
XX
OS Muridae
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Sciurognathi; Muroidea.
XX
RN [1]
RP 1-149
RA Smit A.F.;
RT "B1_Mur1 - a subfamily of SINEs from Muridae.";
RL Direct Submission to Repbase Update (11-NOV-2005).
XX
DR [1] (Consensus)
XX

```

```

SQ   Sequence 149 BP; 37 A; 41 C; 46 G; 25 T; 0 other;
      agccgggCGT ggtggcgcac gcctttaatc ccagcactcg ggaggcagag gcaggcggat      60
      ctctgtgagt tcgaggccag cctggtctac atagcgagtt ccaggccagc cagggctaca      120
      tagtgagacc ctgtctcaaa aaaacaaaa                                     149

//
ID   B1_Mur2      repbase;      DNA;      ROD; 149 BP.
XX
AC   .
XX
DT   16-MAR-2006 (Rel. 11.03, Created)
DT   10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE   SINE1 from Muridae.
XX
KW   SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW   B1; B1_Mur2; SINE1.
XX
OS   Muridae
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC   Sciurognathi; Muroidea.
XX
RN   [1]
RP   1-149
RA   Smit A.F.;
RT   "B1_Mur2 - a subfamily of SINEs from Muridae.";
RL   Direct Submission to Repbase Update (11-NOV-2005).
XX
DR   [1] (Consensus)
XX
SQ   Sequence 149 BP; 38 A; 40 C; 46 G; 25 T; 0 other;
      agccgggCGT ggtggcgcac gcctttaatc ccagcactcg ggaggcagag gcaggcggat      60
      ctctgtgagt tcgaggccag cctggtctac atagcgagtt ccaggacagc cagggctaca      120
      tagtgagacc ctgtctcaaa aaaacaaaa                                     149

//
ID   B1_Mur3      repbase;      DNA;      ROD; 149 BP.
XX
AC   .
XX
DT   16-MAR-2006 (Rel. 11.03, Created)
DT   10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE   SINE1 from Muridae.
XX
KW   SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW   B1; B1_Mur3; SINE1.
XX
OS   Muridae
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC   Sciurognathi; Muroidea.
XX
RN   [1]
RP   1-149
RA   Smit A.F.;
RT   "B1_Mur3 - a subfamily of SINEs from Muridae.";
RL   Direct Submission to Repbase Update (11-NOV-2005).
XX
DR   [1] (Consensus)
XX
SQ   Sequence 149 BP; 40 A; 41 C; 46 G; 22 T; 0 other;
      agccgggCGT ggtggcgcac gcctttaatc ccagcactcg ggaggcagag gcaggcggat      60
      ctctgtgagt tcgaggccag cctggtctac agagcgagtt ccaggacagc cagggctaca      120
      cagagaaacc ctgtctcaaa aaaacaaaa                                     149

//
ID   B1_Mur4      repbase;      DNA;      ROD; 150 BP.
XX

```

```

AC      .
XX
DT      16-MAR-2006 (Rel. 11.03, Created)
DT      10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE      SINE1 from Muridae.
XX
KW      SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW      B1; B1_Mur4; SINE1.
XX
OS      Muridae
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC      Sciurognathi; Muroidea.
XX
RN      [1]
RP      1-150
RA      Smit A.F.;
RT      "B1_Mur4 - a subfamily of SINEs from Muridae.";
RL      Direct Submission to Repbase Update (11-NOV-2005).
XX
DR      [1] (Consensus)
XX
SQ      Sequence 150 BP; 40 A; 40 C; 47 G; 23 T; 0 other;
          agccgggcg tgggtggcgca cgcctttaat cccagcactc gggaggcaga ggcaggcgga           60
          tctctgtgag ttcgaggcca gcctggtcta cagagtgagt tccaggacag ccagggtctac           120
          acagagaaac cctgtctcaa aaaaacaaaa                                         150

//
ID      B1_Mus1      repbase;      DNA;      ROD; 148 BP.
XX
AC      .
XX
DT      16-MAR-2006 (Rel. 11.03, Created)
DT      10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE      SINE1 from mouse.
XX
KW      SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW      B1; B1_Mus1; SINE1.
XX
OS      Mus
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC      Sciurognathi; Muroidea; Muridae; Murinae.
XX
RN      [1]
RP      1-148
RA      Smit A.F.;
RT      "B1_Mus1 - a subfamily of SINEs from mouse.";
RL      Direct Submission to Repbase Update (11-NOV-2005).
XX
DR      [1] (Consensus)
XX
SQ      Sequence 148 BP; 40 A; 38 C; 46 G; 24 T; 0 other;
          agccgggagc tgggtggcgca cgcctttaat cccagcactt gggaggcaga ggcaggcgga           60
          tttctgagtt cgaggccagc ctggtctaca gagtgagttc caggacagcc agggctacac           120
          agagaaacc  tgtctcgaaa aaacaaaa                                         148

//
ID      B1_Mus2      repbase;      DNA;      ROD; 147 BP.
XX
AC      .
XX
DT      16-MAR-2006 (Rel. 11.03, Created)
DT      10-APR-2007 (Rel. 11.03, Last updated, Version 1)
XX
DE      SINE1 from mouse.
XX

```

```

KW SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW B1; B1_Mus2; SINE1.
XX
OS Mus
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Sciurognathi; Muroidea; Muridae; Murinae.
XX
RN [1]
RP 1-147
RA Smit A.F.;
RT "B1_Mus2 - a subfamily of SINE1 SINEs from mouse.";
RL Direct Submission to Repbase Update (11-NOV-2005).
XX
DR [1] (Consensus)
XX
SQ Sequence 147 BP; 39 A; 39 C; 46 G; 23 T; 0 other;
agccgggCGT ggtggcgCAC gcctttaaTC ccagcactCG ggaggcagAG gcaggcggAT      60
ttctgagttC gaggccagCC tggTctacAG agtgagttCC aggacagcCA gggctacaCA      120
gagaaaccCT gtctcgaaaa acaaaaa                                     147

//
ID B1_Mm      repbase;      DNA;      ROD; 147 BP.
XX
AC .
XX
DT 16-MAR-2006 (Rel. 11.03, Created)
DT 19-APR-2006 (Rel. 11.03, Last updated, Version 1)
XX
DE SINE1 from mouse.
XX
KW SINE1/7SL; SINE; Non-LTR Retrotransposon; Transposable Element;
KW B1; B1_Mm; SINE1.
XX
OS Mus
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Sciurognathi; Muroidea; Muridae; Murinae.
XX
RN [1]
RP 1-147
RA Smit A.F.;
RT "B1_Mm - SINE1 subfamily from mouse.";
RL Direct Submission to Repbase Update (11-NOV-2005).
XX
DR [1] (Consensus)
XX
SQ Sequence 147 BP; 40 A; 38 C; 45 G; 24 T; 0 other;
agccgggCGT ggtggcgCAC gcctttaaTC ccagcactCG ggaggcagAG gcaggcggAT      60
ttctgagttC gaggccagCC tggTctacAA agtgagttCC aggacagcCA gggctataCA      120
gagaaaccCT gtctcgaaaa acaaaaa                                     147

//

```

## PŘÍLOHA 2: SEZNAM NALEZENÝCH INSERCÍ S KOORDINÁTAMI

Koordináty odpovídají assembly NCBI37/mm9

**Tabulka 1: B1 retrotransposony.**

ID elementu	ID BACend	ID	podrodina	chromosom	start (genom)	konec (genom)	gen bac	start (BACend)	konec (BACend)
1_176524044	MSMg01-530H05.T7	1_176524044 MSMg01-530H05.T7	B1_Mus2	1	173444424	173444650	gen		
1_184636837	MSMg01-346K09.TJ	1_184636837 MSMg01-346K09.TJ	B1_Mus2	1	181344812	181345014	gen		
2_28969957	MSMg01-365D03.TJ	2_28969957 MSMg01-365D03.TJ	B1_Mus2	2	28965944	28966132	gen		
2_144035101	MSMg01-447A02.T7	2_144035101 MSMg01-447A02.T7	B1_Mur4	2	145663660	145663827	gen		
3_33680452.1	MSMg01-236H23.T7	3_33680452.1 MSMg01-236H23.T7	B1_Mm	3	32628709	32628892	gen		
3_41093123.1	MSMg01-062M24.TJ	3_41093123.1 MSMg01-062M24.TJ	B1_Mus2	3	39951898	39952083	gen		
3_78899579	MSMg01-122C21.TJ	3_78899579 MSMg01-122C21.TJ	B1_Mus2	3	78161357	78161358	bac	635	804
3_92725614	MSMg01-360D06.TJ	3_92725614 MSMg01-360D06.TJ	B1_Mus2	3	90116643	90116644	bac	173	368
3_97825501	MSMg01-510H06.T7	3_97825501 MSMg01-510H06.T7	B1_Mus2	3	95469737	95469945	gen		
4_35055668	MSMg01-178D10.T7	4_35055668 MSMg01-178D10.T7	B1_Mm	4	34839506	34839692	gen		
4_46980783	MSMg01-291K18.TJ	4_46980783 MSMg01-291K18.TJ	B1_Mm	4	46546301	46546484	gen		
4_88704012.2	MSMg01-078J24.TJ	4_88704012.2 MSMg01-078J24.TJ	B1_Mm	4	87562356	87562360	bac	243	439
4_95244799	MSMg01-264O17.TJ	4_95244799 MSMg01-264O17.TJ	B1_Mus2	4	94133259	94133421	gen		
5_4619638	MSMg01-447G16.TJ	5_4619638 MSMg01-447G16.TJ	B1_Mus2	5	9892042	9892217	gen		
5_14316129	MSMg01-169M08.T7	5_14316129 MSMg01-169M08.T7	B1_Mm	5	14931584	14931585	bac	127	303
5_56513557	MSMg01-158L16.T7	5_56513557 MSMg01-158L16.T7	B1_Mm	5	55986692	55986839	gen		
5_63811851	MSMg01-403N07.T7	5_63811851 MSMg01-403N07.T7	B1_Mus2	5	63673735	63673736	bac	343	524
5_146871706	MSMg01-328A18.T7	5_146871706 MSMg01-328A18.T7	B1_Mus2	5	145740448	145740650	gen		
6_50311353	MSMg01-051J13.TJ	6_50311353 MSMg01-051J13.TJ	B1_Mm	6	49908006	49908119	gen		
6_58823397	MSMg01-184B18.T7	6_58823397 MSMg01-184B18.T7	B1_Mm	6	58557163	58557352	gen		
6_60498547	MSMg01-516O01.T7	6_60498547 MSMg01-516O01.T7	B1_Mm	6	60229347	60229348	bac	159	342
8_14430657	MSMg01-412M06.TJ	8_14430657 MSMg01-412M06.TJ	B1_Mm	8	13776559	13776747	gen		

8_59211105	MSMg01-092P07.TJ	8_59211105 MSMg01-092P07.TJ	B1_Mus2	8	58220573	58220791	gen		
9_4094488	MSMg01-231H09.T7	9_4094488 MSMg01-231H09.T7	B1_Mus1	9	4103538	4103539	bac	222	357
9_72337842	MSMg01-433H24.TJ	9_72337842 MSMg01-433H24.TJ	B1_Mus2	9	70103847	70104024	gen		
9_89499640	MSMg01-354D22.T7	9_89499640 MSMg01-354D22.T7	B1_Mm	9	86555569	86555747	gen		
9_111556280	MSMg01-323K23.T7	9_111556280 MSMg01-323K23.T7	B1_Mm	9	109052814	109052815	bac	320	475
10_119699620	MSMg01-423N15.T7	10_119699620 MSMg01-423N15.T7	B1_Mm	10	116270581	116270582	bac	72	244
11_36548853	MSMg01-051K21.T7	11_36548853 MSMg01-051K21.T7	B1_Mus1	11	37231328	37231463	gen		
11_56305209	MSMg01-355C22.T7	11_56305209 MSMg01-355C22.T7	B1_Mus2	11	57314248	57314416	gen		
11_75720483	MSMg01-143C03.T7	11_75720483 MSMg01-143C03.T7	B1_Mus2	11	76934509	76934510	bac	290	447
11_77724284	MSMg01-520024.T7	11_77724284 MSMg01-520024.T7	B1_Mm	11	78938309	78938572	gen		
11_95901520	MSMg01-166H03.T7	11_95901520 MSMg01-166H03.T7	B1_Mm	11	97157734	97157915	gen		
13_57491756	MSMg01-340G10.TJ	13_57491756 MSMg01-340G10.TJ	B1_Mus2	13	56856093	56856278	gen		
13_88893392.1	MSMg01-076E17.TJ	13_88893392.1 MSMg01-076E17.TJ	B1_Mm	13	87096032	87096218	gen		
13_91614534	MSMg01-304J16.TJ	13_91614534 MSMg01-304J16.TJ	B1_Mus2	13	89806272	89806438	gen		
14_65639239.1	MSMg01-347C01.TJ	14_65639239.1 MSMg01-347C01.TJ	B1_Mm	14	65140426	65140427	bac	160	356
15_55878401	MSMg01-275E20.TJ	15_55878401 MSMg01-275E20.TJ	B1_Mus2	15	54978313	54978539	gen		
16_10849613	MSMg01-369E23.TJ	16_10849613 MSMg01-369E23.TJ	B1_Mus2	16	11286960	11287132	gen		
16_14595666	MSMg01-141015.TJ	16_14595666 MSMg01-141015.TJ	B1_Mus2	16	14509634	14509807	gen		
16_28210146	MSMg01-405G12.T7	16_28210146 MSMg01-405G12.T7	B1_Mus2	16	28260855	28260856	bac	154	332
16_59705297	MSMg01-509C16.T7	16_59705297 MSMg01-509C16.T7	B1_Mus1	16	59540202	59540203	bac	184	358
17_65594026	MSMg01-528I10.TJ	17_65594026 MSMg01-528I10.TJ	B1_Mus1	17	63924315	63924316	bac	306	483
18_34472190	MSMg01-432O14.TJ	18_34472190 MSMg01-432O14.TJ	B1_Mm	18	34522749	34522751	bac	222	402
18_42374186	MSMg01-541H12.T7	18_42374186 MSMg01-541H12.T7	B1_Mus2	18	42113168	42113350	gen		
18_72042872	MSMg01-519H18.TJ	18_72042872 MSMg01-519H18.TJ	B1_Mm	18	72481064	72481065	bac	352	525
19_4482934	MSMg01-486E07.T7	19_4482934 MSMg01-486E07.T7	B1_Mus1	19	6154116	6154117	bac	341	521
19_37506476	MSMg01-237L20.TJ	19_37506476 MSMg01-237L20.TJ	B1_Mm	19	36732316	36732498	gen		
19_41388940	MSMg01-386E09.TJ	19_41388940 MSMg01-386E09.TJ	B1_Mus2	19	40493365	40493545	gen		

**Tabulka 2: L1 retrotransposony.**

ID elementu	ID BACend	ID	podrodina	chromosom	start (genom)	end (genom)	gen bac	start (BACend)	end (BACend)
1_8329451	MSMg01-325P17.T7	1_8329451 MSMg01-325P17.T7	L1Md_A	1	8267002	8267648	bac		
1_17678674	MSMg01-047F05.TJ	1_17678674 MSMg01-047F05.TJ	L1Md_F2	1	17785968	17786193	bac		
1_31389386	MSMg01-268A02.T7	1_31389386.1 MSMg01-268A02.T7	L1Md_A	1	31274723	31275439	bac		
1_39623985	MSMg01-458C03.TJ	1_39623985 MSMg01-458C03.TJ	L1Md_F2	1	39359577	39359918	bac		
1_70843897	MSMg01-457L10.T7	1_70843897 MSMg01-457L10.T7	L1Md_T	1	69929547	69930097	bac		
1_71165456	MSMg01-464I14.TJ	1_71165456 MSMg01-464I14.TJ	L1Md_A	1	70253578	70253843	bac		
1_82036590	MSMg01-364I24.TJ	1_82036590 MSMg01-364I24.TJ	L1Md_A	1	81052577	81053004	bac		
1_99787920	MSMg01-471I03.T7	1_99787920 MSMg01-471I03.T7	L1Md_A	1	98173931	98174208	bac		
1_110594102	MSMg01-378O22.T7	1_110594102 MSMg01-378O22.T7	L1Md_A	1	108589235	108588912	bac		
1_115940924	MSMg01-522E22.T7	1_115940924.2 MSMg01-522E22.T7	L1Md_F3	1	113791908	113791527	bac		
1_126011684	MSMg01-269D13.T7	1_126011684 MSMg01-269D13.T7	L1Md_F2	1	125279851	125280769	bac		
1_134739382	MSMg01-060I15.T7	1_134739382 MSMg01-060I15.T7	L1Md_F3	1	132175764	132175765	gen	321	538
1_147931066	MSMg01-119B23.T7	1_147931066 MSMg01-119B23.T7	L1Md_A	1	145267320	145268130	bac		
1_169920251	MSMg01-534L01.TJ	1_169920251 MSMg01-534L01.TJ	L1Md_F2	1	167036866	167036510	bac		
1_171605415	MSMg01-067J08.T7	1_171605415 MSMg01-067J08.T7	L1Md_A	1	168625138	168624587	bac		
1_172498001	MSMg01-177F12.T7	1_172498001 MSMg01-177F12.T7	L1Md_A	1	169533731	169533919	bac		
2_13219396	MSMg01-502F14.T7	2_13219396.2 MSMg01-502F14.T7	L1Md_A	2	13209744	13210624	bac		
2_15303085	MSMg01-480A19.T7	2_15303085 MSMg01-480A19.T7	L1Md_A	2	15297279	15298253	bac		
2_74335778	MSMg01-307E11.T7	2_74335778 MSMg01-307E11.T7	L1Md_T	2	75111840	75112321	bac		
2_102613148	MSMg01-546N18.TJ	2_102613148 MSMg01-546N18.TJ	L1Md_T	2	103828562	103828823	bac		
2_104944842	MSMg01-444H15.TJ	2_104944842 MSMg01-444H15.TJ	L1Md_F3	2	106184662	106184770	bac		
2_126218541	MSMg01-468F12.TJ	2_126218541 MSMg01-468F12.TJ	L1Md_A	2	127612512	127612880	bac		
3_6390028	MSMg01-164F16.TJ	3_6390028.2 MSMg01-164F16.TJ	L1Md_T	3	6336568	6336777	bac		
3_7111342	MSMg01-161G09.TJ	3_7111342.1 MSMg01-161G09.TJ	L1Md_A	3	7069681	7070362	bac		

3_10138001	MSMg01-492E08.T7	3_10138001 MSMg01-492E08.T7	L1Md_F2	3	10057242	10057606	bac		
3_12698563	MSMg01-318K24.TJ	3_12698563 MSMg01-318K24.TJ	L1Md_T	3	12606810	12607211	bac		
3_26002004	MSMg01-270B18.T7	3_26002004 MSMg01-270B18.T7	L1Md_A	3	25024979	25025625	bac		
3_30262943	MSMg01-442B12.T7	3_30262943.2 MSMg01-442B12.T7	L1Md_T	3	29213710	29214169	bac		
3_73893616	MSMg01-159B05.TJ	3_73893616 MSMg01-159B05.TJ	L1Md_A	3	73222417	73222589	bac		
3_84234575	MSMg01-265M14.T7	3_84234575 MSMg01-265M14.T7	L1Md_T	3	83329491	83330059	bac		
3_85186470	MSMg01-274E05.T7	3_85186470 MSMg01-274E05.T7	L1Md_T	3	84339357	84339613	bac		
3_121804640	MSMg01-501M03.TJ	3_121804640.1 MSMg01-501M03.TJ	L1Md_A	3	118895062	118895255	bac		
3_124317857	MSMg01-486C09.T7	3_124317857 MSMg01-486C09.T7	L1Md_T	3	121445848	121446551	bac		
3_131484992	MSMg01-290B17.TJ	3_131484992 MSMg01-290B17.TJ	L1Md_T	3	128535735	128536089	bac		
3_136518998	MSMg01-064G09.T7	3_136518998 MSMg01-064G09.T7	L1Md_T	3	133392144	133392565	bac		
3_161286348	MSMg01-496I04.TJ	3_161286348 MSMg01-496I04.TJ	L1Md_A	3	156789091	156789092	gen	370	485
4_81736049	MSMg01-541I09.TJ	4_81736049 MSMg01-541I09.TJ	L1Md_A	4	80598074	80598240	bac		
4_82559210	MSMg01-464P04.T7	4_82559210 MSMg01-464P04.T7	L1Md_F2	4	81579572	81579131	bac		
4_124184313	MSMg01-139F13.TJ	4_124184313.1 MSMg01-139F13.TJ	L1Md_A	4	123244341	123244744	bac		
5_3910210	MSMg01-092I20.T7	5_3910210 MSMg01-092I20.T7	L1Md_T	5	10665650	10665972	bac		
5_3298726	MSMg01-231P23.TJ	5_3298726 MSMg01-231P23.TJ	L1Md_T	5	11997671	11997672	gen	183	324
5_59948493	MSMg01-048N19.TJ	5_59948493.2 MSMg01-048N19.TJ	L1Md_T	5	59570420	59570634	bac		
5_58682582	MSMg01-093E12.TJ	5_58682582 MSMg01-093E12.TJ	L1Md_T	5	59656721	59657474	bac		
5_71704925	MSMg01-414G03.T7	5_71704925 MSMg01-414G03.T7	L1Md_F2	5	71311709	71311880	bac		
5_71821629	MSMg01-412I10.TJ	5_71821629 MSMg01-412I10.TJ	L1Md_T	5	71426233	71426861	bac		
5_92950676	MSMg01-151N15.T7	5_92950676 MSMg01-151N15.T7	L1Md_T	5	91867666	91867804	bac		
6_7875345	MSMg01-411B12.T7	6_7875345 MSMg01-411B12.T7	L1Md_A	6	7892228	7892653	bac		
6_13642581	MSMg01-274G02.T7	6_13642581 MSMg01-274G02.T7	L1Md_T	6	13621584	13622259	bac		
6_16484548	MSMg01-313G04.TJ	6_16484548 MSMg01-313G04.TJ	L1Md_A	6	16507379	16507976	bac		
6_22837330	MSMg01-270K09.TJ	6_22837330 MSMg01-270K09.TJ	L1Md_A	6	22875442	22875448	gen	208	357
6_26625388	MSMg01-085P02.TJ	6_26625388 MSMg01-085P02.TJ	L1Md_T	6	26685840	26686217	bac		



6_28072861	MSMg01-529A01.T7	6_28072861 MSMg01-529A01.T7	L1Md_T	6	28149653	28150284	bac		
6_32965288	MSMg01-071J05.TJ	6_32965288.2 MSMg01-071J05.TJ	L1Md_A	6	32997255	32997586	bac		
6_44961499	MSMg01-141D11.T7	6_44961499 MSMg01-141D11.T7	L1Md_F3	6	44838198	44838413	bac		
6_59069827	MSMg01-357J23.TJ	6_59069827.1 MSMg01-357J23.TJ	L1Md_A	6	58898532	58898774	bac		
6_64098349	MSMg01-089L09.TJ	6_64098349 MSMg01-089L09.TJ	L1Md_F2	6	63830002	63830739	bac		
6_79515770	MSMg01-291H14.T7	6_79515770 MSMg01-291H14.T7	L1Md_F2	6	78881235	78881842	bac		
6_80910728	MSMg01-214B24.T7	6_80910728 MSMg01-214B24.T7	L1Md_A	6	80282670	80283492	bac		
7_51783755	MSMg01-470J23.T7	7_51783755 MSMg01-470J23.T7	L1Md_A	7	47854312	47854892	bac		
7_62957382	MSMg01-092A06.T7	7_62957382 MSMg01-092A06.T7	L1Md_F2	7	58870559	58870813	bac		
7_140845894	MSMg01-362H09.T7	7_140845894 MSMg01-362H09.T7	L1Md_A	7	137541794	137541796	gen	472	637
8_64555191	MSMg01-425N19.TJ	8_64555191 MSMg01-425N19.TJ	L1Md_T	8	63632545	63632712	bac		
9_8466020	MSMg01-450F10.TJ	9_8466020.1 MSMg01-450F10.TJ	L1Md_A	9	8445245	8445416	bac		
9_15785513	MSMg01-234C22.T7	9_15785513 MSMg01-234C22.T7	L1_Mm	9	15432966	15433333	bac		
9_17650740	MSMg01-149K02.T7	9_17650740.1 MSMg01-149K02.T7	L1Md_F3	9	17306513	17306781	bac		
9_27223449	MSMg01-489O20.TJ	9_27223449 MSMg01-489O20.TJ	L1Md_F2	9	26475939	26476130	bac		
9_81399318	MSMg01-326M07.TJ	9_81399318.1 MSMg01-326M07.TJ	L1Md_A	9	79066467	79067106	bac		
9_84292630	MSMg01-057O06.T7	9_84292630 MSMg01-057O06.T7	L1Md_A	9	79066467	79067106	bac		
10_8875109	MSMg01-348J18.T7	10_8875109 MSMg01-348J18.T7	L1Md_T	10	8843283	8843679	bac		
10_43887393	MSMg01-493O16.T7	10_43887393 MSMg01-493O16.T7	L1Md_T	10	42125945	42126136	bac		
10_52466401	MSMg01-167G17.T7	10_52466401 MSMg01-167G17.T7	L1Md_T	10	50470772	50471291	bac		
10_98345022	MSMg01-112F18.TJ	10_98345022 MSMg01-112F18.TJ	L1Md_A	10	95369428	95369897	bac		
10_99657349	MSMg01-326H18.T7	10_99657349 MSMg01-326H18.T7	L1Md_A	10	96659933	96660243	bac		
10_108435506	MSMg01-453A02.TJ	10_108435506 MSMg01-453A02.TJ	L1Md_T	10	105368263	105369093	bac		
11_47298780	MSMg01-301B21.TJ	11_47298780 MSMg01-301B21.TJ	L1Md_F2	11	48201615	48202120	bac		
11_47376024	MSMg01-345L05.TJ	11_47376024 MSMg01-345L05.TJ	L1Md_T	11	48279272	48278874	bac		
11_91562780	MSMg01-403I07.T7	11_91562780 MSMg01-403I07.T7	L1Md_T	11	92819100	92819871	bac		
11_97140118	MSMg01-398O08.T7	11_97140118 MSMg01-398O08.T7	L1Md_T	11	98396330	98396452	bac		
11_97894539	MSMg01-455L06.T7	11_97894539 MSMg01-455L06.T7	L1Md_F3	11	99150755	99151177	bac		

11_111042012	MSMg01-440E22.T7	11_111042012 MSMg01-440E22.T7	L1Md_A	11	112226905	112227195	bac		
12_15791523	MSMg01-490C07.T7	12_15791523 MSMg01-490C07.T7	L1Md_T	12	15511341	15511784	bac		
12_64051351	MSMg01-532N01.TJ	12_64051351 MSMg01-532N01.TJ	L1Md_F2	12	62981905	62981242	bac		
12_83874715	MSMg01-279F10.T7	12_83874715.3 MSMg01-279F10.T7	L1Md_A	12	82537504	82537928	bac		
13_13280616	MSMg01-471O15.T7	13_13280616 MSMg01-471O15.T7	L1Md_F2	13	13688359	13688796	bac		
13_17601292	MSMg01-454I02.T7	13_17601292 MSMg01-454I02.T7	L1Md_A	13	17095241	17095402	bac		
13_29779817	MSMg01-352F11.TJ	13_29779817.1 MSMg01-352F11.TJ	L1Md_A	13	29337432	29337761	bac		
13_82239999	MSMg01-507K19.TJ	13_82239999 MSMg01-507K19.TJ	L1Md_T	13	80407576	80407870	bac		
13_95998175	MSMg01-374K19.TJ	13_95998175 MSMg01-374K19.TJ	L1Md_T	13	94125406	94125595	bac		
13_106371884	MSMg01-260M16.TJ	13_106371884 MSMg01-260M16.TJ	L1Md_F2	13	104439907	104440716	bac		
14_11331875	MSMg01-377D05.TJ	14_11331875 MSMg01-377D05.TJ	L1Md_T	14	11332759	11333486	bac		
14_28665790	MSMg01-125C02.TJ	14_28665790 MSMg01-125C02.TJ	L1Md_A	14	28850262	28850918	bac		
14_33348676	MSMg01-167E12.T7	14_33348676 MSMg01-167E12.T7	L1Md_T	14	39657932	39658502	bac		
14_81471965	MSMg01-295N06.TJ	14_81471965.1 MSMg01-295N06.TJ	L1Md_F2	14	80399492	80399674	bac		
14_104768340	MSMg01-379I16.T7	14_104768340 MSMg01-379I16.T7	L1Md_T	14	102969874	102970487	bac		
14_104906806	MSMg01-045J17.T7	14_104906806 MSMg01-045J17.T7	L1Md_F2	14	103108525	103109213	bac		
14_117280613	MSMg01-280P16.TJ	14_117280613.1 MSMg01-280P16.TJ	L1Md_T	14	115266859	115267194	bac		
15_15528705	MSMg01-536D05.T7	15_15528705 MSMg01-536D05.T7	L1Md_T	15	15299682	15300212	bac		
15_23441603	MSMg01-161A07.TJ	15_23441603 MSMg01-161A07.TJ	L1Md_T	15	23047239	23047961	bac		
15_25643563	MSMg01-267K06.TJ	15_25643563 MSMg01-267K06.TJ	L1Md_A	15	25004158	25004159	gen	442	578
15_27111053	MSMg01-055D13.T7	15_27111053 MSMg01-055D13.T7	L1Md_A	15	26324117	26324342	bac		
15_27482753	MSMg01-304P08.TJ	15_27482753 MSMg01-304P08.TJ	L1Md_F2	15	26712091	26712643	bac		
15_36065904	MSMg01-492D19.T7	15_36065904 MSMg01-492D19.T7	L1Md_T	15	35628024	35628218	bac		
15_48174104	MSMg01-164P07.T7	15_48174104 MSMg01-164P07.T7	L1Md_F2	15	47436417	47436579	bac		
16_50550325	MSMg01-507F14.T7	16_50550325 MSMg01-507F14.T7	L1Md_F2	16	50297370	50297651	bac		
16_53096182	MSMg01-448F13.T7	16_53096182 MSMg01-448F13.T7	L1Md_F3	16	52757213	52757602	bac		
16_54871030	MSMg01-501H13.TJ	16_54871030 MSMg01-501H13.TJ	L1Md_A	16	54688487	54688727	bac		
16_55654112	MSMg01-162F05.TJ	16_55654112 MSMg01-162F05.TJ	L1Md_F2	16	55477223	55477358	bac		

18_8655834	MSMg01-438L15.T7	18_8655834 MSMg01-438L15.T7	L1Md_T	18	8371709	8372443	bac		
18_40480595	MSMg01-375A09.TJ	18_40480595 MSMg01-375A09.TJ	L1Md_T	18	40553380	40554019	bac		
18_43502515	MSMg01-134A21.TJ	18_43502515 MSMg01-134A21.TJ	L1Md_F2	18	43089137	43089365	bac		
19_13563297	MSMg01-207J15.TJ	19_13563297.2 MSMg01-207J15.TJ	L1Md_F3	19	12908951	12908953	gen	243	380
19_14939555	MSMg01-388G17.T7	19_14939555 MSMg01-388G17.T7	L1Md_A	19	14388317	14388580	bac		
X_61904809	MSMg01-351D08.T7	X_61904809 MSMg01-351D08.T7	L1Md_A	X	56465114	56465465	bac		
X_62061500	MSMg01-509N14.TJ	X_62061500 MSMg01-509N14.TJ	L1Md_T	X	56621806	56622758	bac		
X_65938750	MSMg01-343I02.TJ	X_65938750 MSMg01-343I02.TJ	L1Md_A	X	60527704	60528276	bac		
X_106770124	MSMg01-323P19.TJ	X_106770124 MSMg01-323P19.TJ	L1Md_F2	X	102037447	102037448	gen	241	558

## PŘÍLOHA 3: SEKVENCE NAVRŽENÝCH PRIMERŮ

Tabulka 1: Primery B1

ID elementu	F	R
1_184636837	ATCGCACCACCTTGGAAATGT	TTGCACACGGAGTTGCTTAG
11_36548853	TTCAAAGGCACAGCTTTGTAGA	CAGAATTGGCCACTCCAAGT
11_75720483	TTTGCAGCATCCAACAATTT	CACCCAGGTATGCTCCCTAA
11_77724284	AAAGGGAAGGTGCCAAGAAG	AAAACCCCCAAAACAAAAC
11_95901520	CCGACCCTCTAGGAAGCAAT	GTCCTCAGCTGTCAGGAAGG
13_57491756	ATGGTTTTGTGGGAGGTGTC	CTTCCTGGTCGCAGTTCTTC
13_88893392.1	ATGGGGGACGGTATTGAAA	CAGTTTTGGTTATACATCACAGCA
13_91614534	TTCACCTTCTCCCAATGCAC	CCAACTCCTCCAGACACTC
14_65639239.1	TTCAGGCTAAACACCTATGCAC	TGCACCTTGCTTTCTTGTC
16_14595666	TGGAAAATGACTTCTCAGACCA	GGAAAATATCTATGAAACAAAATGGAA
17_65594026	TTTCCCAATTTGGCATCTGT	CAGGATCGCTGAAATGTCAAT
18_42374186	TGCATCATCAGCACCCATAC	GACAAAGCTGTGGGAATGGT
18_72042872	TTGCTTTATGGATTAAGCCTCAG	CATTTGTTTTTAGCAGATGTTGA
19_41388940	CCAAGACCAACTTCCCAGAA	AGGCTGTTTCAATCCCACAC
19_4482934	AAAAGCCAGAGGAGGAGAGG	AGCTATGACGTCCCAACCAG
2_28969957	GCTCCTCCACAGCTTCTCTG	CCCTTGGATCTTGGTTCTCA
3_33680452.1	CTTTTGTCGTTTAGGGGATGG	TGGAGAAGGGACATCTTTGG
3_41093123.1	GTGGCATGCTTCATAGTGGA	TCCAGGAAAGGGAATATTCAA
3_78899579	TTTTGCAATGTTGTCTGAAAATG	ATGCCCTTCTACATGGACA
3_97825501	AGCCACAAAGCCAGCTTAAA	ATCTGGCTAAAGCCATGCTG
4_35055668	CGTGCTGACTTTGGTTGAAA	GCAAGTTGGTCTGCTCCTTC
4_46980783	TGCCTGGCTCCTATTTCTGT	CTACCCCTTGGAGGGAAGAG
4_95244799	TCAAGACATTGAAAGGGAGA	CACACACACTGTAGCAAGGACA
5_14316129	TCCCTGATGGCAAATCCTAC	CCATGGAGATCAGCAAGGTT
5_146871706	TGTTTCGTTTGGGGAGAGAG	GGTGTGCAAAGCAAGTTTCA
5_56513557	TCCCTCTCTTTGATCCCTGA	CCTATCATTGGCATGCAGGT
8_14430657	GCAGATGACTGCTGCAAGTT	CAGCAGAATCCACACCCTCT
8_59211105	GCACACTTTCAAGCAGGATG	TTGAGCTATCAAAGATGTGGAAA
9_72337842	CAAATGGTGTGCAAATGGA	CGGCAGAACCCTCGAAAGTTA
9_89499640	TCTGGCCCTCTTCTGGTCTA	TTGCACCCATGGATTACATTT

**Tabulka 2: Primery L1**

<b>ID elementu</b>	<b>F</b>	<b>R</b>
1_110594102	TGTTTGCACAATAGCCTTTGG	GAGGAAGATGCAGAGTCAGG
1_115940924.2	ACAAATGTAGGTTTGGGCAT	AGCCATCATATCTCCCTGAA
1_134739382	GAGAGAAAGGCAGTGCAGAG	TAGCTGGAGTAGCTTGACCT
1_147931066	TTTGGAACACTTGGAGGAGG	CAGTAGAGCCAAAGTTGAACG
1_169920251	ATGAGAAGGCTGAACACTGG	CTCCGGTCTGACATACACAG
1_172498001	GTTCAGTCAGGCAAAGCATC	CGCGTGCTTTCTTAAGAGTC
1_17678674	CCCATGAACTCCATAGTGTAGA	TGTCAAATTCAATTAAGGACTATCCC
1_31389386.1	GCTCATAACAAATTGAGGGCT	CATTCCATTTCTTCTCTCCTGC
1_39623985	GGTTAGCTTCTGGTCTCTGG	CAGCTGAAAGTTGGAAGCAG
1_70843897	ACCCTGAGATTTCACTTTCAACT	GGATCTGATGACCAGGGTTC
1_71165456	CTCAGGATTGAGGTCTGCTC	GCTCATGGACATGTTGTTGC
1_82036590	AGACATTTTGTAGCATAAAAACAGAT	CCTATTGATGTACCTACATTCTAAA
1_8329451	TCTGAGCATTGAAACGTAACAG	ACACTAAGAGAAGACACCAAAAAC
1_99787920	ACCAGTCCTACCTCCTGATA	TCTTGAGCTTCTCTTTCCCA
10_108435506	GGTATCCCTAGAGCTGGAGT	CAAAGGATCCGTAAGGGGAGG
10_43887393	TCAAAATGCTATGCTCCGAGA	ACCCGATTTCTGCTTAGGAT
10_52466401	AACCTCTTCTACTCTCCCC	AACCACCAACTTTTGGGACT
10_8875109	TAGCCTCCCCTCTGCATTAT	AAGAAAAACAGAGCAACGGC
10_98345022	GGGGTGGGAGGGTAATAAGA	GAAAACCATGTAGGCAAGGC
10_99657349	AGCCCCCTAGACACATCATTT	ACTTTCACTTGTCTTTTTGTCA
11_111042012	CTCAAGATATCCCGGCCAAT	TTACTGTCAGGACAGCCAAG
11_47298780	AGCAGAGTCAACAGCATGTA	TTGGGTCCCAGGAAAAGAAA
11_47376024	CAAGCACTTTGGGCTTTCTC	TGCTAACAGAACCCTCTCT
11_97140118	GCACTGTCTCACACGACTAA	ACCTCCAAGTAACCCGAGTA
11_97894539	CTTCATACTCCCAGTGCAT	TTAGCCTCCTGCACAGATTC
12_15791523	TCCTGTGACTTTTCTCCCTTG	GCCCCATTGATCTCTCAG
12_64051351	AATGAAACAGTCTAAACACTATCG	CCCACCTGTACTTGAAAATAGAT
12_83874715.3	AACCAACTGGACGTGATTGA	AAATGGGAACAGGAAGCAGT
13_13280616	GTGGAAAAGGGGTACACAT	CCAGGAACGGCAGCTATATT
13_17601292	TGCAACCTGCCATAACTGAT	TGTGCCAAGTATGTTTTGTGT
13_29779817.1	AGCCACACTGAAAGCAAAG	TGGAAATAGCATCTCACTCTTGA
13_82239999	CACAAGGACAACACCGAAAG	GAGTGCTCTCACCTGGAATC
14_104768340	TGTAAGGAGCTTGTGGTGTT	CAAAGGAAGGAGGTAGTGTGT
14_104906806	TCTGGGATCAAGGATCCACT	TTCTTGTCTGTTGGGCAAT
14_11331875	TGCAAGCTACGAGTATTGCT	GGGCCTCAAATTAGACCCAT
14_117280613.1	GCTGTGGACATACTGAGCTT	CTACAGGGGAGAGTACCA
14_28665790	TCGTGTTCAGGTTTTATCTAGCA	TACCGCTGATGTTGTCATGT
14_81471965.1	ATATAAGGCAAGGAGCAGGC	AGGCTCCTAGGGGAAAAA
15_15528705	GTAACCCTAACAGCAGGGTC	CCCCTCAAATGGAAAGAGT
15_23441603	AGGCTCTTCTTACAAAGTGTCA	TCACTCAAACCATTACATCTACAGT
15_25643563	GACTGTACTCTCCTGGGGAT	CCCTGAGCATTCTTGCATTT
15_27482753	GTGTTCCACCCAAGTCTAA	GTGGTAGAGGTGTGTACCTG
15_36065904	TCATGGAAAGGCAAAGGAGA	TGCAAACTGAACTAAGGCAA
15_48174104	GGCCATTTGGAGGTCACATA	TCCCTCTTTGCCAGAATGC
16_50550325	AGCACTACCTTAAGAGTTGCAT	AGACTCCTGAATATATGAGAAACACT
16_53096182	CTGGAATTCCCTCTGATCGGG	GGAACAGACAGACCAACACT
16_54871030	ATGCAAGAACTTTGCAGACC	GCATCCATCCTCCCTTAGTG

**Tabulka 2: Primery L1**

ID elementu	F	R
16_55654112	TTTCCTGGGTCTGTTTAGGC	TGTGATTCCTCATCGCACAT
18_40480595	GCAGCTAATTGCACAGTTCC	CACGGTTTGATCCATGAGGA
18_43502515	TGTCATCATCTCTTTCTTCATTCTT	GGCTGGGTAGAAAATAAAACCA
18_8655834	GCAGTGGCTGCATTATATGTT	GCTATTGGAATCCATGAATGGTC
19_13563297.2	AGAAGCTAGCAAGGCTTTTAGA	TCCATGAGGCAGTTTTGAGT
19_14939555	TCTGGATCATAAGAAAACAGTGGA	TAGGATGCTTAGGGGGATGT
2_102613148	TTGACTGGGGAAATGTGTGT	AGAGAGCAGCCTCATCTAGT
2_104944842	GGCAAGAGTATGTGCATCTG	AGACTTTGACTGAGGGTGTG
2_126218541	TGTGGTGCTAGATCACATATTTA	GCCGTTACTGGATTGTTTCAT
2_74335778	TGAGGCTTAACAAGGCTGAG	ATTTGACCTCAAGGTGAGACA
3_10138001	TCCCACTCTTGAGAGCAAAG	ACCACATGGCTCATTATGTCT
3_121804640.1	TGCAATCACACCTATAAGAGAACT	ACTTCTAAGTGGCTTTCAAGACT
3_124317857	ACATCCCAAGTCTAAGATGCTC	AGAAAGTGCAAATAATGAGCC
3_12698563	CCTTTGTCAGATATTCACTTTACCT	ACCCAGGAATTCTAAAACCTTTTCT
3_131484992	CAGACTCTAAGCCCCAAAA	AAGATCAGCCACATTCCGGAG
3_161286348	ACTGAGCAGCATTTCAGCTAT	CAGGCACTGACATCAAAAAGC
3_26002004	ACGTCTGAACTTAACTCAGGG	ACATGCCATTGCAAAACACAA
3_30262943.2	ATGCCAACATTTCTAGGCCA	AGGAATATGAAGATGATATCCAGACA
3_6390028.2	TTATACAAGCGCAAATCGGG	CCTTATGTGTATCCCAGTGTCA
3_7111342.1	AGGGAATGAATGTTTTGTTGTGT	TGCATAACAAATGATTTCTGTGAAGA
3_73893616	ACCAACTGGTGCTTAAATGC	CCATCGATTCTTTATTGTCTCAC
3_84234575	TGACCATTAGGAGCAACCTT	AAATGCGCTTAACTCTGTGG
3_85186470	TGTTTGGTATCATAGTAAGTGGAA	CTAGTAGCAGAGTGGACAGA
4_124184313.1	ACAGTACAGAAAACCTGAGCCA	TTGTCTCACAAGAGAGGACC
4_81736049	CTTGCTGAGTTCATACCCTT	AGAAAGAAAACAGTTCCTACTCT
4_82559210	CGGGCATGGTATCTTCACAT	GGAGATTGGTTTTCCCTCTGG
5_3298726	TGGCTTGGGCAAAATCTGTA	CTCAGTGAGCAGAGTGAGTC
5_3910210	AGGTGAATGTTTTCTAACTGATGG	GGGACAGCCAAAGAAATGAC
5_58682582	CCAGGGCTTAATGACAACCT	TCTCTGGACTACCTTAAACAATCA
5_59948493.2	CAGAGTCAGTGGGAAAACCT	TCCCTGGCCTTTAACTAGGA
5_71704925	CAACTAACAGATTCAGTGGTCA	AGGCTAGTAAAAACAAAGGTGT
5_71821629	TGCCAAGAAGTATTTGACATTCAT	CCCAACCATCCCAAAATGAC
5_92950676	AAAGCATTGCTATTGCTCTCC	AGCAAGATCTATAAGGAGCACAT
6_13642581	GCTACAAACACCACCTCTGA	TGGAGCTCCATTTTTAGTAGT
6_16484548	AACACCATATCTTTGTGGTAATTGT	TGAAGGATGTGTGATGTGTGA
6_22837330	TGTCCAATACAGCCTTTCCC	GAAGGAGACATGGCCAGAAA
6_26625388	ATGTTTGCCACAAAGACTGC	ACTGCTGCCTCCTGATTATG
6_28072861	AAATTCAGGTCAAGTAGCGT	GGGGTGTAATTTGTATGAATCTTATT
6_32965288.2	GGTCGTGCAAACCCATAAAA	GCTTTCTTCAGGCAAACACT
6_44961499	GTACTGCTGTGCTATTGAGC	CTCATGCCAGGGATCTTCAA
6_59069827.1	ATATTGCCCTGTGTGGCTTT	TGTTGCTAACAGCAGGAAGT
6_64098349	CTGCAAGAAGAATGTAAATAAGCTG	GCTTTGTCCATAAAGCAGGT
6_7875345	AGCCTCTTCCTGGACTCTTA	TCTTGCCCCATCTTTCTGAC
6_80910728	AGACACATTCATCTTATAGTGTGAG	CCTGTCTAATCCTCCTTGTGA
7_140845894	AAGATGGTTGTCACTGGACC	GGTCTCTCAGGCTTCTGATG
7_51783755	GTGCCTAGAGGTGCTACTA	TTGGCCATAAAACAACTCTTCA
7_62957382	AGGGATGAGGCCTTCTATGA	ACCTGTCAGCCAACAGTTTAT
9_15785513	TTCTTTCTCTTTCAGCCAC	TCATGTGTGCTTTCTATGTCA
9_17650740.1	TTGTAATCAGCCAGGGTGTG	CCATGGAGTGGGTAAAGGG

**Tabulka 2: Primery L1**

<b>ID elementu</b>	<b>F</b>	<b>R</b>
9_27223449	GCATTGTCTTCTTCCTTCATGG	GTTATGGGGGCCAGAGAATG
9_81399318.1	TTGGCTGCTTTGATTCATGT	GCTTATCATTCTCATTCTGG
9_84292630	GGAAGAGTGCAAGTGAGTGA	TGATGTCCCAACATCACAG
9_8466020.1	CATTAAGGGGCGGGCTA	GTCACAGATACACTGATCTCCTT
X_106770124	GCTCAACTTCCCAGTCTTGC	AAGAGCATGCACACGAACTG
X_61904809	GGAATCTGACAACCGAGAA	TCCTTCCAGCCCTAAACTGA

## PŘÍLOHA 4: GENOTYPIZACE B1 A L1 RETROTRANSPONŮ.

(a) – absence inserce, (h) – heterozygot, (p) – přítomnost inserce, (NA) – daný kmen neamplifikoval fragment, (-) – kolem inserce nebylo možné navrhnout primery.

**Tabulka 1: B1 retrotransposony**

ID	BZO	CIM	CTP	MPB	SEG	Straas 2788	Stud 2802	XBS
1_176524044 MSMg01-530H05.T7	-	-	-	-	-	-	-	-
1_184636837 MSMg01-346K09.TJ	a	a	a	a	a	p	a	a
2_28969957 MSMg01-365D03.TJ	p	a	a	a	NA	a	a	a
2_144035101 MSMg01-447A02.T7	-	-	-	-	-	-	-	-
3_33680452.1 MSMg01-236H23.T7	h	a	a	a	a	a	a	a
3_41093123.1 MSMg01-062M24.TJ	a	p	a	a	a	h	a	a
3_78899579 MSMg01-122C21.TJ	NA	NA	NA	NA	p	a	NA	p
3_92725614 MSMg01-360D06.TJ	-	-	-	-	-	-	-	-
3_97825501 MSMg01-510H06.T7	p	a	a	p	a	p	p	a
4_35055668 MSMg01-178D10.T7	p	a	a	a	p	p	a	p
4_46980783 MSMg01-291K18.TJ	p	a	a	a	a	p	a	a
4_88704012.2 MSMg01-078]24.TJ	-	-	-	-	-	-	-	-
4_95244799 MSMg01-264O17.TJ	p	a	a	a	a	p	a	a
5_4619638 MSMg01-447G16.TJ	-	-	-	-	-	-	-	-
5_14316129 MSMg01-169M08.T7	a	a	a	a	a	a	a	a
5_56513557 MSMg01-158L16.T7	p	p	p	p	p	p	p	p
5_63811851 MSMg01-403N07.T7	-	-	-	-	-	-	-	-
5_146871706 MSMg01-328A18.T7	p	a	a	a	a	p	a	a
6_50311353 MSMg01-051]13.TJ	-	-	-	-	-	-	-	-
6_58823397 MSMg01-184B18.T7	-	-	-	-	-	-	-	-
6_60498547 MSMg01-516O01.T7	-	-	-	-	-	-	-	-
8_14430657 MSMg01-412M06.TJ	p	a	a	a	a	p	a	a



8_59211105 MSMg01-092P07.TJ	p	a	a	h	a	p	p	a
9_4094488 MSMg01-231H09.T7	-	-	-	-	-	-	-	-
9_72337842 MSMg01-433H24.TJ	p	p	p	a	p	p	a	p
9_89499640 MSMg01-354D22.T7	p		a	a	p	p	NA	NA
9_111556280 MSMg01-323K23.T7	-	-	-	-	-	-	-	-
10_119699620 MSMg01-423N15.T7	-	-	-	-	-	-	-	-
11_36548853 MSMg01-051K21.T7	p	p	a	a	p	p	p	a
11_56305209 MSMg01-355C22.T7	-	-	-	-	-	-	-	-
11_75720483 MSMg01-143C03.T7	a	p	p	p	p	a	p	p
11_77724284 MSMg01-520O24.T7	p	p	p	p	p	p	p	NA
11_95901520 MSMg01-166H03.T7	a	a	a	a	a	a	a	a
13_57491756 MSMg01-340G10.TJ	p	a	a	a	p	p	a	p
13_88893392.1 MSMg01-076E17.TJ	p	a	NA	a	p	p	NA	a
13_91614534 MSMg01-304 16.TJ	a	a	a	a	p	a	a	NA
14_65639239.1 MSMg01-347C01.TJ	a	p	a	p	a	a	p	a
15_55878401 MSMg01-275E20.TJ	-	-	-	-	-	-	-	-
16_10849613 MSMg01-369E23.TJ	-	-	-	-	-	-	-	-
16_14595666 MSMg01-141O15.TJ	p	a	a	a	a	p	p	a
16_28210146 MSMg01-405G12.T7	-	-	-	-	-	-	-	-
16_59705297 MSMg01-509C16.T7	-	-	-	-	-	-	-	-
17_65594026 MSMg01-528I10.TJ	a	p	a	p	a	a	p	p
18_34472190 MSMg01-432O14.TJ	-	-	-	-	-	-	-	-
18_42374186 MSMg01-541H12.T7	p	p	a	p	p	p	p	a
18_72042872 MSMg01-519H18.TJ	a	p	p	a	a	a	p	a
19_4482934 MSMg01-486E07.T7	NA	p	p	p	NA	NA	p	NA
19_37506476 MSMg01-237L20.TJ	-	-	-	-	-	-	-	-
19_41388940 MSMg01-386E09.TJ	p	p	p	a	a	p	a	a

**Tabulka 2: L1 retrotransposony**

ID	BZO	CIM	CTP	MPB	SEG	Straas 2788	XBS
1_8329451 MSMg01-325P17.T7	a	a	a	a	a	p	a
1_17678674 MSMg01-047F05.TJ	p	a	p	a	a	p	a
1_31389386.1 MSMg01-268A02.T7	a	a	a	a	a	p	a
1_39623985 MSMg01-458C03.TJ	p	a	a	a	a	p	a
1_70843897 MSMg01-457L10.T7	p	a	a	a	a	p	NA
1_71165456 MSMg01-464I14.TJ	p	a	a	a	a	p	a
1_82036590 MSMg01-364I24.TJ	p	a	a	a	a	NA	a
1_99787920 MSMg01-471I03.T7	p	a	a	a	a	p	a
1_110594102 MSMg01-378O22.T7	p	a	a	a	a	p	a
1_115940924.2 MSMg01-522E22.T7	p	p	a	h	p	p	p
1_126011684 MSMg01-269D13.T7	-	-	-	-	-	-	-
1_134739382 MSMg01-060I15.T7	p	p	p	p	a	p	a
1_147931066 MSMg01-119B23.T7	a	a	a	a	a	p	a
1_169920251 MSMg01-534L01.TJ	p	a	a	a	a	p	a
1_171605415 MSMg01-067J08.T7	-	-	-	-	-	-	-
1_172498001 MSMg01-177F12.T7	p	a	a	a	a	a	a
2_13219396.2 MSMg01-502F14.T7	-	-	-	-	-	-	-
2_15303085 MSMg01-480A19.T7	-	-	-	-	-	-	-
2_74335778 MSMg01-307E11.T7	p	a	a	a	a	p	a
2_102613148 MSMg01-546N18.TJ	p	p	p	NA	NA	NA	NA
2_104944842 MSMg01-444H15.TJ	p	p	a	a	p	p	p
2_126218541 MSMg01-468F12.TJ	h	a	a	a	a	p	a
3_6390028.2 MSMg01-164F16.TJ	p	a	a	a	a	p	a
3_7111342.1 MSMg01-161G09.TJ	p	a	a	a	a	p	a

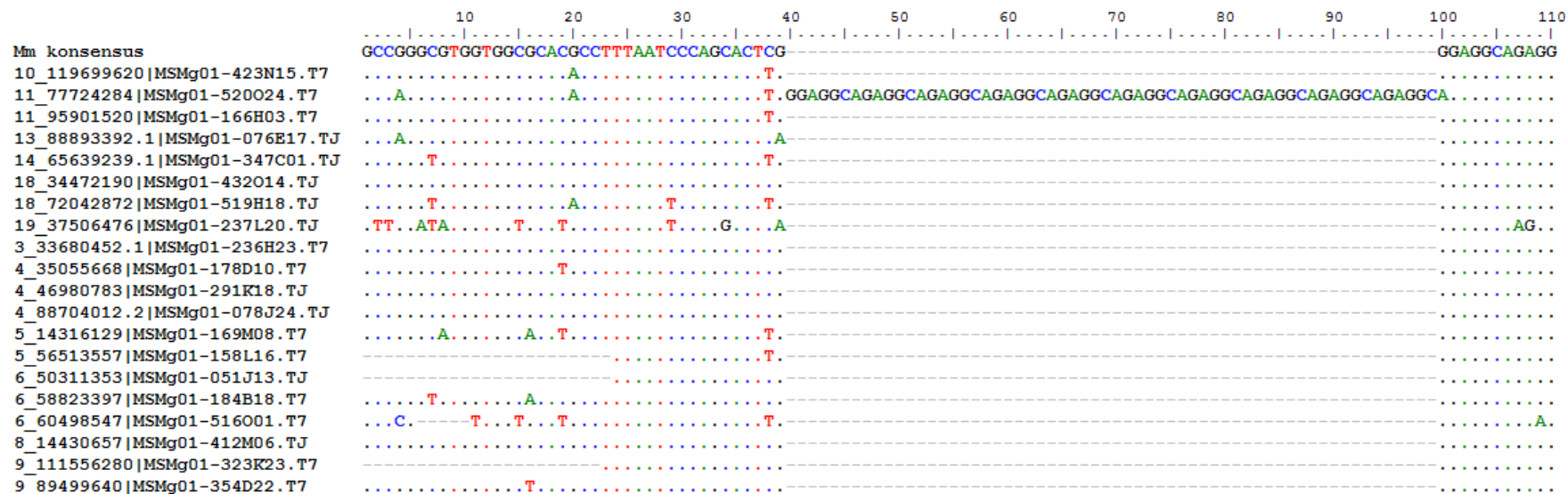
3_10138001 MSMg01-492E08.T7	p	a	a	a	a	p	a
3_12698563 MSMg01-318K24.TJ	a	a	a	a	a	a	a
3_26002004 MSMg01-270B18.T7	p	a	a	a	a	p	a
3_30262943.2 MSMg01-442B12.T7	p	a	a	a	a	p	a
3_73893616 MSMg01-159B05.TJ	p	p	p	a	a	p	NA
3_84234575 MSMg01-265M14.T7	p	p	a	a	a	p	a
3_85186470 MSMg01-274E05.T7	NA	a	a	a	a	a	a
3_121804640.1 MSMg01-501M03.TJ	a	p	a	a	a	p	a
3_124317857 MSMg01-486C09.T7	p	a	a	a	a	p	a
3_131484992 MSMg01-290B17.TJ	a	a	a	a	a	a	a
3_136518998 MSMg01-064G09.T7	-	-	-	-	-	-	-
3_161286348 MSMg01-496I04.TJ	a	a	a	p	a	a	a
4_81736049 MSMg01-541I09.TJ	a	a	a	a	a	a	a
4_82559210 MSMg01-464P04.T7	p	a	a	a	a	a	a
4_124184313.1 MSMg01-139F13.TJ	a	a	a	a	NA	a	a
5_3910210 MSMg01-092I20.T7	a	a	a	a	a	a	a
5_3298726 MSMg01-231P23.TJ	a	a	h	h	a	a	a
5_59948493.2 MSMg01-048N19.TJ	p	p	a	a	a	a	a
5_58682582 MSMg01-093E12.TJ	NA	a	NA	a	a	a	a
5_71704925 MSMg01-414G03.T7	a	h	p	a	a	p	a
5_71821629 MSMg01-412I10.TJ	p	a	a	a	a	p	a
5_92950676 MSMg01-151N15.T7	NA	NA	NA	NA	a	a	a
6_7875345 MSMg01-411B12.T7	p	a	a	a	a	p	a
6_13642581 MSMg01-274G02.T7	a	a	a	a	a	h	a
6_16484548 MSMg01-313G04.TJ	NA	a	a	a	a	p	a
6_22837330 MSMg01-270K09.TJ	a	a	p	a	a	a	a
6_26625388 MSMg01-085P02.TJ	p	a	a	a	a	p	a
6_28072861 MSMg01-529A01.T7	p	p	a	a	a	p	a

6_32965288.2 MSMg01-071 05.TJ	p	NA	a	a	a	a	a
6_44961499 MSMg01-141D11.T7	p	a	a	a	a	p	a
6_59069827.1 MSMg01-357 23.TJ	a	a	a	a	a	p	a
6_64098349 MSMg01-089L09.TJ	p	p	a	a	p	p	p
6_79515770 MSMg01-291H14.T7	-	-	-	-	-	-	-
6_80910728 MSMg01-214B24.T7	a	a	a	a	a	p	a
7_51783755 MSMg01-470 23.T7	a	a	a	a	a	p	a
7_62957382 MSMg01-092A06.T7	p	a	a	p	a	p	a
7_140845894 MSMg01-362H09.T7	a	a	a	a	a	a	a
8_64555191 MSMg01-425N19.TJ	-	-	-	-	-	-	-
9_8466020.1 MSMg01-450F10.TJ	p	a	h	p	h	p	h
9_15785513 MSMg01-234C22.T7	a	a	a	a	a	a	a
9_17650740.1 MSMg01-149K02.T7	p	NA	NA	NA	p	NA	p
9_27223449 MSMg01-489O20.TJ	a	p	a	a	p	a	p
9_81399318.1 MSMg01-326M07.TJ	p	a	a	a	a	p	a
9_84292630 MSMg01-057O06.T7	p	a	NA	a	a	a	a
10_8875109 MSMg01-348 18.T7	p	a	a	a	a	NA	a
10_43887393 MSMg01-493O16.T7	p	a	a	a	a	p	a
10_52466401 MSMg01-167G17.T7	a	a	a	a	a	p	a
10_98345022 MSMg01-112F18.TJ	p	a	a	a	a	p	a
10_99657349 MSMg01-326H18.T7	a	a	a	a	a	a	a
10_108435506 MSMg01-453A02.TJ	NA	a	a	a	a	p	a
11_47298780 MSMg01-301B21.TJ	p	a	a	a	p	NA	p
11_47376024 MSMg01-345L05.TJ	p	a	a	a	a	a	a
11_91562780 MSMg01-403I07.T7	-	-	-	-	-	-	-
11_97140118 MSMg01-398O08.T7	h	a	a	a	a	p	a
11_97894539 MSMg01-455L06.T7	h	a	a	a	p	p	p
11_111042012 MSMg01-440E22.T7	NA	a	a	a	a	p	a

12_15791523 MSMg01-490C07.T7	p	p	p	p	p	p	p
12_64051351 MSMg01-532N01.TJ	p	p	p	p	a	NA	a
12_83874715.3 MSMg01-279F10.T7	p	p	p	p	p	p	p
13_13280616 MSMg01-471O15.T7	h	p	h	p	a	p	a
13_17601292 MSMg01-454I02.T7	a	a	a	a	a	a	a
13_29779817.1 MSMg01-352F11.TJ	a	p	p	a	a	p	a
13_82239999 MSMg01-507K19.TJ	a	a	a	a	a	a	a
13_95998175 MSMg01-374K19.TJ	-	-	-	-	-	-	-
13_106371884 MSMg01-260M16.TJ	-	-	-	-	-	-	-
14_11331875 MSMg01-377D05.TJ	a	a	a	a	a	p	a
14_28665790 MSMg01-125C02.TJ	p	a	a	a	a	p	a
14_33348676 MSMg01-167E12.T7	-	-	-	-	-	-	-
14_81471965.1 MSMg01-295N06.TJ	p	a	a	a	p	p	a
14_104768340 MSMg01-379I16.T7	p	a	a	a	a	p	a
14_104906806 MSMg01-045J17.T7	p	a	h	a	a	p	a
14_117280613.1 MSMg01-280P16.TJ	p	a	a	a	a	p	a
15_15528705 MSMg01-536D05.T7	a	a	a	a	a	a	a
15_23441603 MSMg01-161A07.TJ	p	a	a	a	a	p	a
15_25643563 MSMg01-267K06.TJ	a	a	a	a	a	a	h
15_27111053 MSMg01-055D13.T7	-	-	-	-	-	-	-
15_27482753 MSMg01-304P08.TJ	p	p	a	a	a	p	a
15_36065904 MSMg01-492D19.T7	a	a	a	a	a	a	a
15_48174104 MSMg01-164P07.T7	p	p	a	a	p	p	p
16_50550325 MSMg01-507F14.T7	p	a	a	a	p	p	p
16_53096182 MSMg01-448F13.T7	p	h	p	a	a	p	NA
16_54871030 MSMg01-501H13.TJ	p	p	h	a	p	p	p
16_55654112 MSMg01-162F05.TJ	p	p	p	a	a	a	a
18_8655834 MSMg01-438L15.T7	p	a	p	a	NA	p	a

18_40480595 MSMg01-375A09.TJ	NA	NA	NA	NA	NA	p	NA
18_43502515 MSMg01-134A21.TJ	a	a	a	a	a	p	a
19_13563297.2 MSMg01-207]15.TJ	a	a	a	a	a	a	a
19_14939555 MSMg01-388G17.T7	a	a	a	a	a	p	a
X_61904809 MSMg01-351D08.T7	a	a	a	a	a	p	a
X_62061500 MSMg01-509N14.TJ	-	-	-	-	-	-	-
X_65938750 MSMg01-343I02.TJ	-	-	-	-	-	-	-
X_106770124 MSMg01-323P19.TJ	a	a	a	p	a	a	a

# PŘÍLOHA 5: ALIGNMENT SEKVENCÍ PODRODINY MM



	120	130	140	150	160	170	180	190	200
Mm konsensus	CAGGCGGA	TTTCTGAG	TTCGAGGCCAG	CCCTGGTCTACAAAG	TGAGTTC	CCAGGACAGCCAG	GGCTATACAGAGAA	CCCTGTCTCG	
10_119699620 MSMg01-423N15.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
11_77724284 MSMg01-520024.T7	.....	.....G.....	.....	.....	.....	.....	.....A.....	.....TT.....	.....
11_95901520 MSMg01-166H03.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
13_88893392.1 MSMg01-076E17.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
14_65639239.1 MSMg01-347C01.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
18_34472190 MSMg01-432014.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
18_72042872 MSMg01-519H18.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
19_37506476 MSMg01-237L20.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
3_33680452.1 MSMg01-236H23.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
4_35055668 MSMg01-178D10.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
4_46980783 MSMg01-291K18.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
4_88704012.2 MSMg01-078J24.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
5_14316129 MSMg01-169M08.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
5_56513557 MSMg01-158L16.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
6_50311353 MSMg01-051J13.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
6_58823397 MSMg01-184B18.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
6_60498547 MSMg01-516001.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
8_14430657 MSMg01-412M06.TJ	.....	.....	.....	.....	.....	.....	.....	.....	.....
9_111556280 MSMg01-323K23.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....
9_89499640 MSMg01-354D22.T7	.....	.....	.....	.....	.....	.....	.....	.....	.....



## PŘÍLOHA 6: ALIGNMENT SEKVENCÍ PODRODINY MUS2

	10	20	30	40	50	60	70	80	90	100	110	
Mus2 konsensus	G	C	C	G	G	G	C	G	C	A	C	C
1_176524044 MSMg01-530H05.T7	.	T	.	A	.	T	.	T	.	T	.	A
1_184636837 MSMg01-346K09.TJ	.	.	.	.	.	.	.	.	.	GA	A	.
11_56305209 MSMg01-355C22.T7	.	.	.	.	.	.	.	.	.	A	.	G
11_75720483 MSMg01-143C03.T7	.	T	.	.	.	T	.	T	.	.	.	G
13_57491756 MSMg01-340G10.TJ	.	.	.	A	.	.	.	.	.	.	A	.
13_91614534 MSMg01-304J16.TJ	.	TC	AT	.	.	G	.	.	.	T	.	A
15_55878401 MSMg01-275E20.TJ	.	T	.	A	.	.	.	A	.	.	T	.
16_10849613 MSMg01-369E23.TJ	.	A	.	A	.	.	T	.	.	A	.	T
16_14595666 MSMg01-141O15.TJ	.	.	.	.	.	T	.	.	.	A	.	T
16_28210146 MSMg01-405G12.T7	---	.	C	.	.	.	C	.	C	.	.	A
18_42374186 MSMg01-541H12.T7	.	A	.	A	.	G	.	.	A	.	T	.
19_41388940 MSMg01-386E09.TJ	.	A	.	.	.	CT	.	.	A	.	T	.
2_28969957 MSMg01-365D03.TJ	.	.	.	.	.	.	.	.	.	.	A	.
3_41093123.1 MSMg01-062M24.TJ	.	.	.	G	.	T	.	.	.	.	.	.
3_78899579 MSMg01-122C21.TJ	.	.	.	.	T	.	.	.	A	.	---	.
3_92725614 MSMg01-360D06.TJ	.	T	.	.	.	.	T	.	.	.	.	A
3_97825501 MSMg01-510H06.T7	.	.	.	.	.	.	.	.	.	.	.	.
4_95244799 MSMg01-264O17.TJ	---	.	C	.	.	A	.	T	.	.	A	.
5_146871706 MSMg01-328A18.T7	.	.	.	.	.	.	.	T	.	.	T	.
5_4619638 MSMg01-447G16.TJ	.	T	.	.	.	.	.	T	.	.	A	.
5_63811851 MSMg01-403N07.T7	A	.	T	.	.	.	.	T	.	C	.	.
8_59211105 MSMg01-092P07.TJ	.	.	.	.	T	.	.	T	.	.	A	.
9_72337842 MSMg01-433H24.TJ	.	.	T	.	A	.	.	T	.	A	.	.

```

                                120      130
Mus2 konsensus      . . . . | . . . . | . . . . | . . . . | . .
GGCTACACA--GAGAAACCCGTCTCG
1_176524044|MSMg01-530H05.T7  . . . . .-- . . . . . . . . . . . . . . . . T
1_184636837|MSMg01-346K09.TJ  . . . . .-- . . . . . . . . . . . . . . . .
11_56305209|MSMg01-355C22.T7  . . A . . . . .-- . . . . . . . . . . . . . . . .
11_75720483|MSMg01-143C03.T7  . . . . .C . . . . .C . C . . . . . . . . . . T
13_57491756|MSMg01-340G10.TJ  . . . . .-- . . G . . . . . . . . . . . . . . . .
13_91614534|MSMg01-304J16.TJ  . . . . .-- . . . . . . . . . . . . . . . . A
15_55878401|MSMg01-275E20.TJ  . . . . .-- . . . . . . . . . . . . . . . .
16_10849613|MSMg01-369E23.TJ  A . . . . .-- . . . . . . . . . . . . . . . .
16_14595666|MSMg01-141O15.TJ  . A . . . . .-- . . . . . . . . . . . . . . . . A
16_28210146|MSMg01-405G12.T7  . . . . .-- . . . . . . . . . . . . . . . .
18_42374186|MSMg01-541H12.T7  . . . . .-- . . . . . . . . . . . T . . . . .
19_41388940|MSMg01-386E09.TJ  . . . . .-- . . A . . . . . . . . . . . . . . . .
2_28969957|MSMg01-365D03.TJ   . . . . .-- . . . . . . . . . . . . . . . .
3_41093123.1|MSMg01-062M24.TJ  . . . . .G . . . . .-- . . . . . T . . . . .
3_78899579|MSMg01-122C21.TJ   . . . . .T . . . . .-- . . . . . . . . . . . . . . . .
3_92725614|MSMg01-360D06.TJ   A . . . . .-- . . . . . . . . . . . . . . . . A
3_97825501|MSMg01-510H06.T7   . . . . .-- . . . . . . . . . . . . . . . . A
4_95244799|MSMg01-264O17.TJ   . . . . .-- . . . . . . . . . . . . . . . .
5_146871706|MSMg01-328A18.T7  . . . . .CA . . . . .-- . . . . . . . . . . . . . . . .
5_4619638|MSMg01-447G16.TJ    . . . . .-- . . . . . . . . . . . . . . . .
5_63811851|MSMg01-403N07.T7  . . . . .-- . . . . . . . . . . . . . . . .
8_59211105|MSMg01-092P07.TJ  . . . . .-- . . . . . . . . . . . . . . . .
9_72337842|MSMg01-433H24.TJ  . . GA . . . . .-- . . . . . . . . . . . . . . . . A

```

## PŘÍLOHA 7: DIVERGENCE OD KONSENSU PODRODIN MM A MUS2

**Tabulka 1: Podrodina Mm**

ID	divergence (%)
10_119699620 MSMg01-423N15.T7	2
11_77724284 MSMg01-520024.T7	6.5
11_95901520 MSMg01-166H03.T7	2
13_88893392.1 MSMg01-076E17.TJ	4.1
14_65639239.1 MSMg01-347C01.TJ	5.5
18_34472190 MSMg01-432O14.TJ	2.7
18_72042872 MSMg01-519H18.TJ	5.4
19_37506476 MSMg01-237L20.TJ	13
3_33680452.1 MSMg01-236H23.T7	3.4
4_35055668 MSMg01-178D10.T7	4.8
4_46980783 MSMg01-291K18.TJ	1.4
4_88704012.2 MSMg01-078J24.TJ	2.1
5_14316129 MSMg01-169M08.T7	4.8
5_56513557 MSMg01-158L16.T7	4.7
6_50311353 MSMg01-051J13.TJ	4
6_58823397 MSMg01-184B18.T7	2.9
6_60498547 MSMg01-516O01.T7	7.5
8_14430657 MSMg01-412M06.TJ	1.4
9_11156280 MSMg01-323K23.T7	2.4
9_89499640 MSMg01-354D22.T7	1.4
<b>Průměr</b>	<b>4.1</b>

**Tabulka 2: Podrodina Mus2**

ID	divergence (%)
1_176524044 MSMg01-530H05.T7	9.7
1_184636837 MSMg01-346K09.TJ	2.9
11_56305209 MSMg01-355C22.T7	2.7
11_75720483 MSMg01-143C03.T7	6.1
13_57491756 MSMg01-340G10.TJ	3.4
13_91614534 MSMg01-304J16.TJ	6.1
15_55878401 MSMg01-275E20.TJ	2.7
16_10849613 MSMg01-369E23.TJ	5.5
16_14595666 MSMg01-141O15.TJ	3.4
16_28210146 MSMg01-405G12.T7	4.9
18_42374186 MSMg01-541H12.T7	5.4
19_41388940 MSMg01-386E09.TJ	6.3
2_28969957 MSMg01-365D03.TJ	1.4
3_41093123.1 MSMg01-062M24.TJ	2.8
3_78899579 MSMg01-122C21.TJ	2.2
3_92725614 MSMg01-360D06.TJ	4.8
3_97825501 MSMg01-510H06.T7	0
4_95244799 MSMg01-264O17.TJ	5.6
5_146871706 MSMg01-328A18.T7	3.4
5_4619638 MSMg01-447G16.TJ	4.8
5_63811851 MSMg01-403N07.T7	2.9
8_59211105 MSMg01-092P07.TJ	2.7
9_72337842 MSMg01-433H24.TJ	6.2
<b>Průměr</b>	<b>4.17</b>

## PŘÍLOHA 8: INSERCE VHODNÉ JAKO PODDRUHOVĚ DIAGNOSTICKÉ MARKERY.

### **L1 - diagnostické pro *M. m. domesticus***

1_110594102 MSMg01-378022.T7
1_71165456 MSMg01-464114.TJ
1_99787920 MSMg01-471103.T7
10_98345022 MSMg01-112F18.TJ
14_28665790 MSMg01-125C02.TJ
3_26002004 MSMg01-270B18.T7
3_7111342.1 MSMg01-161G09.TJ
6_7875345 MSMg01-411B12.T7
9_81399318.1 MSMg01-326M07.TJ
1_169920251 MSMg01-534L01.TJ
1_39623985 MSMg01-458C03.TJ
3_10138001 MSMg01-492E08.T7
6_44961499 MSMg01-141D11.T7
10_43887393 MSMg01-493O16.T7
14_104768340 MSMg01-379I16.T7
14_117280613.1 MSMg01-280P16.TJ
15_23441603 MSMg01-161A07.TJ
2_74335778 MSMg01-307E11.T7
3_124317857 MSMg01-486C09.T7
3_30262943.2 MSMg01-442B12.T7
3_6390028.2 MSMg01-164F16.TJ
5_71821629 MSMg01-412I10.TJ
6_26625388 MSMg01-085P02.TJ

### **B1 - diagnostické pro *M. m. domesticus***

4_46980783 MSMg01-291K18.TJ
8_14430657 MSMg01-412M06.TJ
4_95244799 MSMg01-264O17.TJ
5_146871706 MSMg01-328A18.T7