

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra biochemie



**Bioinformatická analýza distribuce transpozonů na
pohlavních chromozomech rostlin**

DIPLOMOVÁ PRÁCE

Autor: **Kateřina Jurásková**

Studijní program: B1406 Biochemie

Studijní obor: Bioinformatika

Forma studia: Prezenční

Vedoucí práce: **Zdeněk Kubát, Ph.D.**

Rok: 2020

Prohlašuji, že jsem diplomovou práci vypracovala samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním diplomové práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne

Ráda bych zde poděkovala vedoucímu diplomové práce Zdeňku Kubátovi, Ph.D., za odborné vedení, vstřícnost, časovou flexibilitu a pomoc při zpracování práce. Děkuji také Mgr. Markétě Bodlákové a Nicolasu Blavetovi, Ph.D., za trpělivost a ochotu, se kterou mi pomáhali. Mé poděkování patří také mé rodině a blízkým přátelům za pomoc a podporu během studia.

Bibliografická identifikace

Jméno a příjmení autora	Kateřina Jurásková
Název práce	Bioinformatická analýza distribuce transpozonů na pohlavních chromozomech rostlin
Typ práce	Diplomová
Pracoviště	Biofyzikální ústav AV ČR
Vedoucí práce	Zdeněk Kubát, Ph.D.
Rok obhajoby práce	2020

Abstrakt

Regulace transpozonů v zárodečných drahách rostlin je dosud otevřené téma. U modelového druhu *Arabidopsis thaliana* byly objeveny epigenetické regulační dráhy přítomné pouze v samčí zárodečné dráze (pyl). Biologická relevance a dlouhodobý vliv těchto regulačních drah na transpozici aktivitu transpozonů je ale u hermafroditní *A. thaliana* nsnadno studovatelný kvůli absenci pohlavně specifického chromatinu. U druhu *Rumex acetosa* s pohlavními chromozomy X a Y je známo, že většina rodin transpozonů vykazuje nerovnoměrné rozložení na chromozomech. Doposud užívané metody ale nemají dostatečné rozlišení k tomu, aby mohlo být prokázáno, že se transpozony u tohoto druhu šíří přednostně v jedné nebo druhé zárodečné dráze. Zde zvolená bioinformatická analýza dat chromozomů tříděných pomocí průtokové cytometrie má potenciál vnést do této problematiky nová zjištění. V této práci jsme analyzovali složení repetitivní frakce genomu *R. acetosa* pomocí RepeatExploreru, pokusili se o potvrzení modelu rozdílného šíření transpozonů u samčích a samičích zárodečných buněk. Očekávali jsme také objev indicií, že i samičí zárodečná dráha obsahuje dosud neobjevené epigenetické regulační mechanismy.

Klíčová slova	transponovatelné elementy, pohlavní chromozomy, satelity, distribuce transpozonů závislá na pohlaví, RepeatExplorer, NGS
Počet stran	59
Počet příloh	1 CD
Jazyk	Český

Bibliographical identification

Autor's first name and surname	Kateřina Jurásková
Title	Bioinformatic Analysis of Transposable Elements Distribution on Plant Sex Chromosomes
Type of thesis	Diploma
Department	Department of biochemistry
Supervisor	Zdeněk Kubát, Ph.D.
The year of presentation	2020

Abstract

Regulation of transposons in plant germline is still an open topic. In the model plant *Arabidopsis thaliana* were found regulatory epigenetic pathways that are specific only for the male germline (pollen). The biological relevance and long term effect of these regulatory pathways on the transposition activity of transposons in hermafroditic *A. thaliana* is difficult to study because of the absence of sex-specific chromatin. In *Rumex acetosa* with sex chromosomes X and Y is known, that most of transposon families show an uneven distribution on sex chromosomes. However, the methods used so far do not have sufficient resolution to demonstrate, that transposons in this species are propagated preferentially either in male or female germline. The bioinformatics analysis of flow sorted chromosomes we used here has the potential to bring new findings into this issue. In this thesis we analyzed repetitive fraction of *R. acetosa* genome and we tested the model of different transposons proliferation in male and female germline. We also expected the discovery of new indications that the female germline contains as yet undiscovered epigenetic regulatory mechanisms.

Keywords	transposable elements, sex chromosomes, satelites, sex-specific distribution of transposable elements, RepeatExplorer, NGS
Number of pages	59
Number of appendices	1 CD
Language	Czech

OBSAH

1. ÚVOD	8
2. SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	9
2.1. Transponovatelné elementy.....	9
2.1.1. Klasifikace.....	10
2.1.2. Retrotranspozony	11
2.1.3. DNA transpozony.....	17
2.2. Regulace aktivity TE	20
2.2.1. Epigenetika.....	20
2.2.1.1. Transkripční umlčování genů	20
2.2.1.2. Posttranskripční umlčování genů.....	22
2.2.2. Epigenetická regulace transpozonů u rostlin.....	23
2.3. TE u rostlin s pohlavními chromozomy	25
3. EXPERIMENTÁLNÍ ČÁST	29
3.1. RepeatExplorer	29
3.2. Klastrová (shluková) analýza	30
3.3. Programovací jazyk R	33
3.4. Metody.....	33
3.4.1. Bioinformatická analýza	33
3.4.2. Statistická analýza.....	34
4. VÝSLEDKY	38
4.1. Složení samčího genomu a jeho komponent	38
4.2. Rozdíly ve složení pohlavních chromozomů	40
4.3. Podíl jednotlivých typů repeticí na zvětšení pohlavních chromozomů.....	41
4.4. Ověřování modelu distribuce transpozonů.....	42
5. DISKUSE	47
6. ZÁVĚR	50
7. LITERATURA	51
8. SEZNAM POUŽITÝCH ZKRATEK	59

Cíle práce

Teoretická část

- Vypracování literární rešerše stavu poznání v oblasti transpozonů u rostlin s pohlavními chromozomy.
- Prostudování odborné literatury.

Praktická část

- Kontrola a příprava NGS dat, analýza repetitivních sekvencí pomocí RepeatExploreru a dalších bioinformatických nástrojů.
- Analýza složení samčího genomu a pohlavních chromozomů u dvoudomé rostliny *Rumex acetosa*
- Potvrzení modelu parentálně specifického šíření transpozonů pomocí bioinformatické analýzy NGS dat získaných ze sortovaných pohlavních chromozomů a autozomů *R. acetosa*.
- Příprava grafických výstupů pro publikaci pomocí veřejně dostupných a samostatně vytvořených bioinformatických nástrojů a algoritmů.

1. ÚVOD

Transpozony jsou mobilní genetické elementy, které tvoří dominantní složku většiny rostlinných genomů. Kvůli své transpoziční aktivitě jsou transpozony potenciálně nebezpečné, mohou narušovat strukturu nebo funkci genů. Rostliny se proto jejich aktivitě brání prostřednictvím epigenetických mechanismů založených na krátkých molekulách RNA, jež řídí transkripční a posttranskripční umlčení transpozonů. Přestože jsou již popsány základní rysy této epigenetické obrany, dosud se neví, do jaké míry se liší obranné mechanismy v samčích a samičích zárodečných drahách, v nichž je aktivita transpozonů nejvyšší.

Na základě předběžných dat získaných v naší laboratoři jsme vytvořili model, podle kterého se některé transpozony šíří přednostně v samčí a jiné v samičí zárodečné linii buněk. Očekává se potvrzení modelu rozdílného šíření transpozonů u samčích a samičích zárodečných buněk. V takovém případě budou vykazovat počty transpozonů dané rodiny na jednotlivých chromozomech určité poměry. Například rodina transpozonů, která je umlčená v samčí zárodečné dráze, bude mít minimální počet kopií na chromozomu Y, vysoký na chromozomu X a střední na autozomech. Rodina umlčená v samičí zárodečné dráze bude mít vysoký počet kopií na chromozomu Y, střední na autozomech a nejnižší na chromozomu X.

2. SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

2.1. Transponovatelné elementy

Transponovatelné elementy (TE, transpozony) jsou mobilní sekvence DNA, které tvoří většinu rostlinných genomů. Jsou schopny se integrovat do nových pozic v genomu, čímž v průběhu času navyšují svůj počet. Spoléhají přitom na jednu nebo více enzymatických funkcí, které si samy kódují nebo jim je poskytují podobné autonomní (plně funkční) elementy (Lisch, 2013). Transpozony mohou ovlivňovat evoluci regulací genové exprese a mutagenní transpozicí, a proto jsou buňkou umlčovány pomocí epigenetických modifikací (Chuong et al., 2017; Craig, 2015). TE byly poprvé popsány u kukuřice Barbarou McClintockovou na konci 40. let 20. století, a díky schopnosti regulovat genovou expresi, byly pojmenovány jako kontrolní prvky (McClintock, 1956).

TE můžeme rozdělit do dvou tříd na základě transpozičních mechanismů a strukturálních rysů - na retrotranspozony (třída I), které se replikují prostřednictvím meziproductů RNA a extrachromozomální komplementární DNA (cDNA), a DNA transpozony (třída II), které se replikují prostřednictvím meziproductu DNA (Craig, 2015). Četnost transpozonů se mezi jednotlivými genomy liší. Například, u kukuřice tvoří TE 75 % genomu a u člověka 40 % genomu (Slotkin a Martienssen, 2007). Faktory, na kterých frekvence zastoupení TE závisí, jsou složité a zahrnují četnost transkripce a transpozice, preference inzerčních míst, pozitivní a negativní selekci. Obecně ale platí, že frekvence zastoupení TE u eukaryot pozitivně koreluje s velikostí genomu (Craig, 2015; Lynch, 2007).

Až do nedávna byly TE považovány za „genomický odpad“. Teprve nedávné studie odhalily podstatný dopad TE na funkce a evoluci genomu, potenciál TE vytvářet funkční genomové struktury (například centromery nebo telomery) nebo regulační genové sítě. Na roli TE v genomech dosud nepanuje jednotný názor a lze na ni pohlížet z různých perspektiv. TE jsou chápány buď jako „sobecká DNA“ usilující pouze o svou vlastní expanzi (Orgel a Crick, 1980) nebo se o TE uvažuje jako o „kontrolních prvcích“ s důležitou úlohou v regulaci genomu. Nejpravděpodobnější je, že TE představují genomové parazity, kteří hrají důležitou pozitivní roli v evoluci organismů (zvyšují jejich variabilitu) a genomů (podílejí se na tvorbě funkčních struktur a regulačních mechanismů) (Mita a Boeke, 2016). Protože jsou TE exprimovány v genomu a podléhají stejným pravidlům dědičnosti jako geny jejich „hostitele“, mohou být transpozony

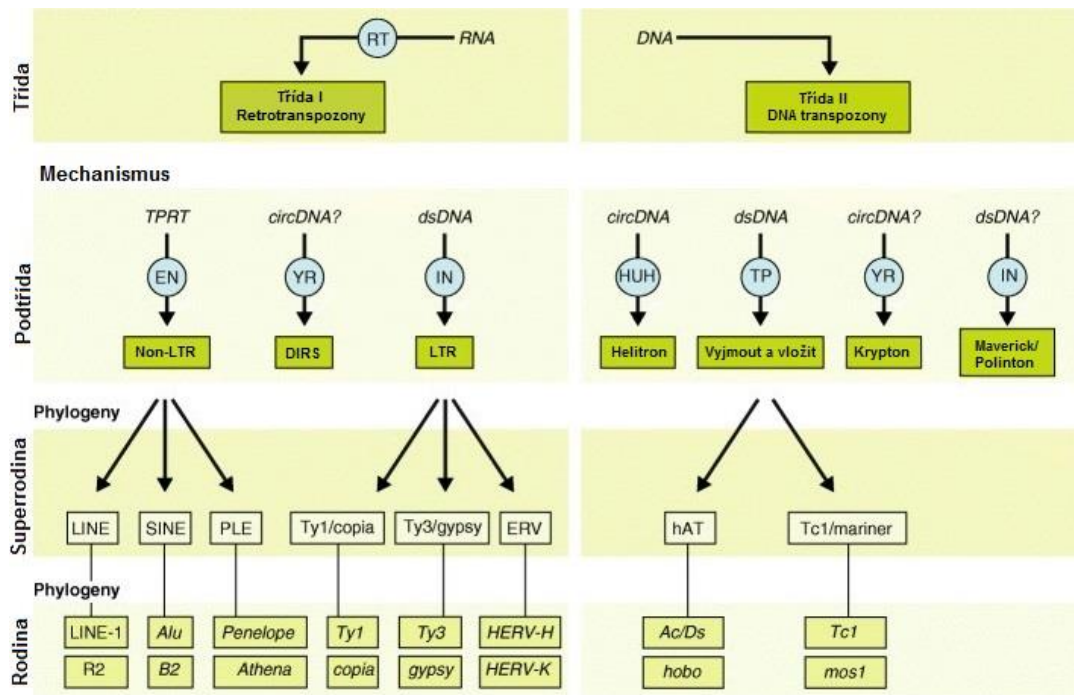
v biologických procesech využity ku prospěchu organismu (Kumar a Bennetzen, 1999; Pardue et al., 1997).

2.1.1. Klasifikace

Transpozony jsou sekvence DNA, které mají schopnost měnit svou polohu v genomu. V důsledku neustálé diverzifikace a evolučních tlaků vytvořily TE ohromující rozmanitost typů a podob. TE lze rozdělit do dvou hlavních tříd na základě jejich mechanismu transpozice, a každou třídu lze rozdělit do podtříd na základě mechanismu chromozomální integrace (Obr. 1) (Bourque et al., 2018). Podle mechanismu transpozice je dělíme na retrotranspozony, které používají proces „kopírovat a vložit“, a DNA transpozony, které používají mechanismus „vyjmout a vložit“ (Mita a Boeke, 2016).

Retrotranspozony a DNA transpozony existují jako autonomní nebo neautonomní elementy (Craig, 2015; Slotkin a Martienssen, 2007). Autonomní elementy kódují proteiny potřebné pro transpozici, ale pro úspěšný přesun potřebují i enzymy hostitele. Naproti tomu neautonomní elementy jsou nekompletní z hlediska proteinové výbavy, mají pouze regulační sekvence vyžadované pro transpozici a vyžadují příbuzný autonomní element, aby zajistil transpoziční faktory (enzymy) *in trans*. DNA transpozony pro transpozici nutně nevyžadují transkripci TE, jelikož si vystačí s proteinovými produkty příbuzných elementů (Craig, 2015). Naopak retrotranspozony přepsány být musí, protože k transpozici může dojít pouze prostřednictvím meziproduktů RNA (kopírování) a cDNA (vlození) (Underwood et al., 2017).

Stav jednotlivých kopií TE v hostitelském genomu se liší v závislosti na stáří a aktivitě dané rodiny TE. Autonomní TE kódují enzymy potřebné pro jejich mobilitu, zatímco mobilita neautonomních elementů závisí na enzymech dodávaných autonomními elementy patřícími stejnému nebo příbuznému TE. Aktivita neautonomních elementů je tak přímo závislá na přítomnosti aktivních autonomních kopií. V mnoha případech je počet neautonomních TE v genomu srovnatelný nebo dokonce vyšší než autonomních TE. TE sekvence v průběhu času akumulují mutace, které narušují schopnost těchto sekvencí produkovat proteiny a dále mobilizovat. Tento proces stárnutí vedl ve většině genomů k přítomnosti mnoha fosilních TE sekvencí vedle několika aktivních kopií (Kidwell a Lisch, 2001).



Obr. 1 Klasifikace eukaryotických TE. Schéma a příklady ukazující klíčové rysy a vztahy mezi třídami, podtřídami, superrodinami a rodinami transpozonů. Modré kruhy představují enzymy kódované TE; circDNA, kruhová DNA; DIRS, Dictyostelium opakující se sekvence; dsDNA, lineární dvouřetězcový DNA meziprodukt; EN, endonukleáza; IN, integráza; PLE, Penelope-like elementy; HUH, Rep / Helicase protein s HUH endonukleázovou aktivitou; RT, reverzní transkriptáza; TP, transposáza, TPRT cíl primovaná reverzní transkripce; YR, tyrosin rekombináza. Převzato z (Bourque et al., 2018).

2.1.2. Retrotranspozony

TE třídy I, také známé jako retrotranspozony, se mobilizují prostřednictvím mechanismu „kopírovat a vložit“, kdy se meziprodukt RNA reverzně přepisuje do cDNA, která je integrována jinde v genomu (Boeke et al., 1985). U retrotranspozonů s dlouhými terminálními repeticemi (LTR) probíhá zpětný přepis v cytoplazmě uvnitř částic podobných virům, které jsou tvořeny geny *gag* kódovanými retrotranspozonem. K integraci dochází štěpením cílového místa v DNA a následnou integrací retrotranspozonové cDNA, což je katalyzováno integrázou podobně jako u retrovirů (Brown et al., 1987). U retrotranspozonů bez LTR (non-LTR retrotranspozony), které zahrnují jak dlouhé, tak krátké rozptýlené jaderné prvky (LINE a SINE), je chromosomální integrace spojena s reverzní transkripcí prostřednictvím procesu označovaného jako reverzní transkripce zahajovaná cílovým místem (Luan et al., 1993).

Retrotranspozony byly odhaleny téměř ve všech eukaryotách z různých taxonomických skupin, jako jsou prvoci, houby, živočichové a rostliny (Kumar

a Bennetzen, 1999), většinou tvoří podstatnou složku genomů a hrají významnou roli v evoluci genů a genomů. Rostlinné retrotranspozony jsou strukturně a funkčně podobné ostatním eukaryotickým retrotranspozonomům a retrovirům, ale existují tu významné rozdíly v genomické organizaci retrotranspozonů, často vysokém počtu jejich kopií, značně heterogenních populacích a jejich distribuci na chromozomech (Kumar a Bennetzen, 1999). Mechanismus replikační transpozice může rychle zvýšit počet kopií elementů a tím výrazně zvětšit velikost genomu rostliny (Kumar, 1996; Sanmiguel a Bennetzen, 1998). Nárůst genomu kukuřice o 75 %, ke kterému došlo v posledních pěti milionech let, je výsledkem transpozice a rozšíření LTR retrotranspozonů (SanMiguel et al., 1998).

Mutace, ke kterým došlo v důsledku vložení retrotranspozonů, jsou na rozdíl od mutací způsobených DNA transpozony stabilní. DNA transpozony používají mechanismus transpozice vyjmutí a vložení, což je vratný proces vedoucí jen k pomalému zvětšování rostlinného genomu, zatímco u retrotranspozonů zůstane sekvence v místě inzerce zachována (Georgiev, 1984; Geyer et al., 1986; Peifer a Bender, 1988).

Retrotranspozony se dělí do tří skupin, které se liší ve své strukturální organizaci a transpozičním mechanismu. První skupina zahrnuje retrotranspozony s dlouhými koncovými repeticemi na obou koncích (LTR retrotranspozony). Druhá skupina představuje non-LTR retrotranspozony, které neobsahují LTR. Poslední skupina obsahuje prvky podobné DIRS (Dictyostelium opakující se sekvence), které obsahují gen tyrosin rekombinázy (YR) namísto genu integrázy (INT) (Obr. 1) (Sormacheva a Blinov, 2011).

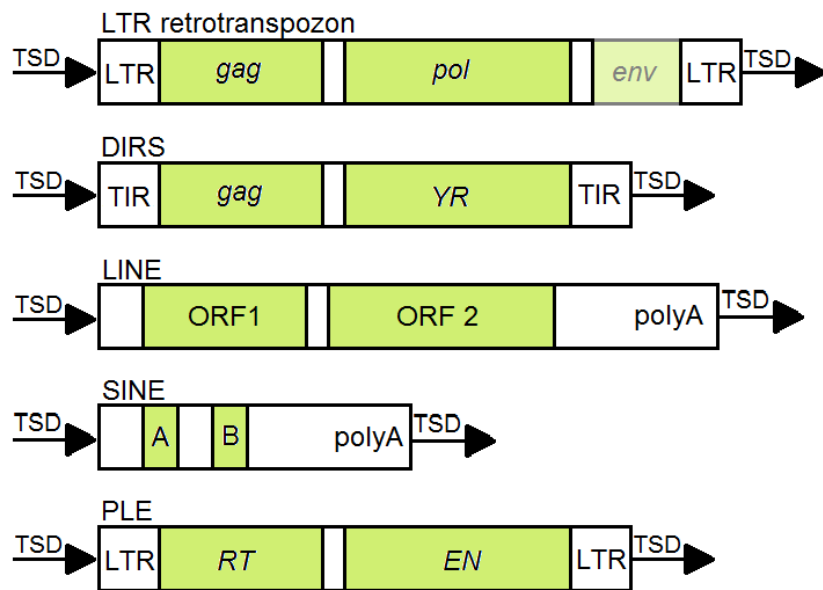
LTR retrotranspozony

LTR retrotranspozony jsou strukturálně podobné retrovirům. Mohou se pohybovat uvnitř hostitelského genomu z jedné chromozomální polohy do druhé přes meziproduct RNA mechanismem „kopírování a vkládání“ (Kumar a Bennetzen, 1999). LTR retrotranspozon má délku přibližně 4-20 tisíc párů bází a obsahuje otevřený čtecí rámec (ORF), který je lemovaný dvěma přímými LTR. ORF je rozdělen do dvou oblastí, *gag* a *pol* (Obr. 2) (Wicker et al., 2007). Gen *gag* kóduje RNA-vázající protein, který tvoří částici podobnou viru, ve které probíhá zpětná transkripce retrotranspozonového genomu. Gen *pol* kóduje několik proteinů, jako je proteáza (PR), integráza, reverzní transkriptáza (RT) a RNáza H, které jsou potřebné pro syntézu a integraci cDNA do hostitelského chromozomu

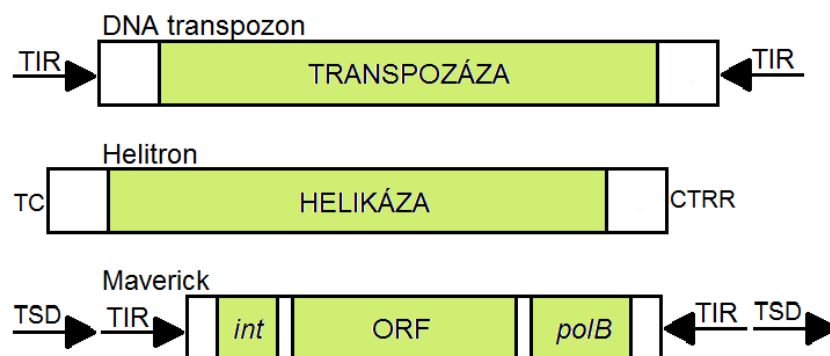
(Boeke a Corces, 1989). Geny *gag*, *pol* a *int* mohou být kódovány v jediném translačním ORF, ale také ve dvou nebo více ORF (Kumar a Bennetzen, 1999).

LTR retrotranspozony mají přímé dlouhé koncové repetice, jejichž délky se mohou pohybovat v rozmezí od několika stovek párů bází do více než 5 kbp. LTR nekódují žádné známé proteiny, jsou zakončeny krátkými obrácených repeticemi, obvykle 5'— TG — 3' a 5' — CA — 3' a obsahují promotory a terminátory spojené s transkripcí LTR retrotranspozonů.

Třída I



Třída II



Obr. 2 Struktura transponovatelných elementů. TSD, duplikace cílového místa; LTR, dlouhá koncová repetice; TIR, terminální invertovaná repetice; ORF, otevřený čtecí rámec; A a B, promotory RNA polymerázy III; polyA dlouhá sekvence adeninových nukleotidů; YR, tyrosin rekombináza; RT, reverzní transkriptáza; EN, endonukleáza; *polB*, DNA polymeráza Beta.

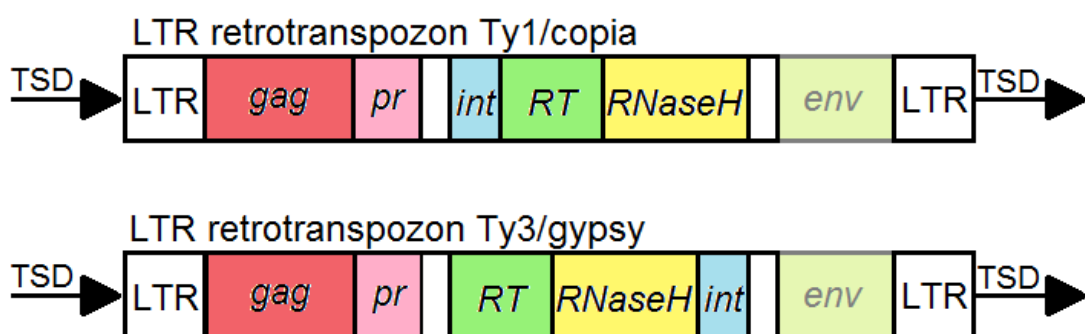
Dalším rysem LTR retrotranspozonů je přítomnost krátkých sekvencí známých jako místo vázající primer (PBS) a polypurinový trakt (PPT). PBS je umístěn za 5' LTR a je nezbytný pro syntézu prvního řetězce cDNA (Sormacheva a Blinov, 2011).

Na základě uspořádání domén v rámci proteinu *pol* a podobnosti sekvence jsou LTR retrotranspozony dále klasifikovány buď jako typ Ty1/copia (*Pseudoviridae*) nebo Ty3/gypsy (*Metaviridae*) (Kumar a Bennetzen, 1999). Oba typy retrotranspozonů se v genomech vyšších rostlin vyskytují ve vysokých počtech (od 5.5 % v genomu *Arabidopsis thaliana* do více než 50 % v genomu kukuřice) (Kaul et al., 2000; SanMiguel et al., 1996).

- Ty1/copia

Ty1/copia je superrodina retrovirů a LTR retrotranspozonů, do které patří mnoho známých retrotranspozonů. Běžně se vyskytují v genomech rostlin, hub, živočichů, řas a některých prvoků (Sormacheva a Blinov, 2011). Název byl vytvořen podle prvního objeveného zástupce Ty1 u kvasinek a copia u octomilek. Ty1/copia obsahuje LTR (dlouhé 100 - 1300 nukleotidů), ORF pro geny *gag* a *pol* a doménu IN, která je umístěna mezi doménami proteázy a reverzní transkriptázy (Obr. 3) (Eickbush a Jamburuthugoda, 2008). Může také obsahovat další ORF kódující retrovirový gen *env* a díky tomu mít status potenciálních retrovirů (Havecker et al., 2005).

Genomy rostlin obsahují obvykle jen několik aktivních kopií Ty1/copia retrotranspozonů, většina kopií je nefunkční (Flavell et al., 1992). Aktivita těchto prvků byla detekována pouze během určitých fází vývoje rostlin nebo za stresových podmínek, což naznačuje, že jejich transpozice je regulována na úrovni transkripce (Flavell et al., 1992).



Obr. 3 Rozdílné uspořádání domén u Ty1/copia a Ty3/gypsy.

- Ty3/gypsy

Ty3/gypsy je další velkou skupinou, do které patří většina známých retrotranspozonů. Stejně jako u Ty1/copia byla skupina pojmenována podle prvních objevených zástupců u kvasinek (Ty3) a octomilek (gypsy). Délka Ty3/gypsy retrotranspozonů se pohybuje od 4 do 25 tisíc párů bází (Sormacheva a Blinov, 2011) a LTR jsou dlouhé 100 - 5000 nukleotidů.

Významný rozdíl mezi Ty1/copia a Ty3/gypsy je v uspořádání domén genu *pol*, a stejně jako Ty1/copia, i Ty3/gypsy mohou obsahovat retrovirový gen *env*. Většina Ty3/gypsy se podobá struktuře retrovirů, *pol* genové domény jsou ve stejném pořadí [proteáza – RT – ribonukleáza H (RH) – integráza] (Obr. 3) (Xiong a Eickbush, 1990). V některých případech může být nejasné, zda je element, podobný typu Ty3, vlastně retrotranspozonem nebo retrovirem (Bennetzen, 1996; Kumar, 1998).

- ERV

Retroviry jsou charakterizovány replikačním cyklem, ve kterém je genom virové RNA reverzně transkribován a integrován do jaderného genomu hostitelské buňky (Vogt, 1997). Enzymy RT a IN umožňují přeměnu jednovláknové virové RNA na dvouvláknovou DNA, po které následuje integrace virové DNA do jaderného genomu infikované buňky. Tím se vytvoří „provirus“. Jako chromozomální inzerce má integrovaný provirus celoživotní vztah s infikovanou buňkou a přežívá tak dlouho, dokud žije buňka (nebo její potomstvo). Když dojde k integraci v zárodečné buňce, může být výsledný provirus vertikálně zděděn jako hostitelská alela. Takový provirus se nazývá endogenní retrovirus (ERV) (Gifford et al., 2018).

LTR retrotranspozony se liší od retrovirů pouze nepřítomností *env* genu kódujícího obalový protein virových částic. Dlouho se věřilo, že LTR retrotranspozony vznikly v důsledku „domestikace“ retrovirů, a jsou to tedy retroviry, které ztratily svůj *env* gen (Xiong a Eickbush, 1990). Fylogenetická analýza LTR retrotranspozonů a retrovirů však ukázala, že situace je zcela opačná. S velkou pravděpodobností jsou retroviry odvozeny od LTR retrotranspozonů. Získáním nového genu *env*, bylo umožněno protoretroviru vytvořit virovou částici s infekční schopností (Xiong a Eickbush, 1990).

DIRS

Dictyostelium opakující se sekvence (DIRS) byly objeveny na počátku 80. let (Zuker a Lodish, 1981) v genomu hlenky. DIRS jsou strukturně rozdělené skupiny TE, které mají

gen tyrosin rekombinázy místo integrázy. Jejich konce se podobají buď rozděleným přímým repetitívám (SDR) nebo obráceným repetitívám (TIR) (Makałowski et al., 2019).

ORF *pol* zahrnuje reverzní transkriptázu, RNázuH (RH) a doménu podobnou bakteriální a fágové DNA N-6-adenin-methyltransferáze (Piednoël et al., 2011). Ve srovnání s retroviry chybí v YR retrotranspozonech jak aspartátová proteáza, tak DDE integráza (Goodwin a Poulter, 2004).

Non-LTR retrotranspozony

Retrotranspozony bez dlouhých terminálních repetitív představují další skupinu. Nepřítomnost LTR je odlišuje strukturálně i mechanicky. Dlouhé rozptýlené jaderné elementy (LINE) a krátké rozptýlené jaderné elementy (SINE) se vyskytují v genomu jako součást opakující se DNA, která není uspořádána v tandemových repetitívách, ale je rozptýlena (Kumar a Bennetzen, 1999). LINE částečně kódují stejné proteiny jako LTR-retrotranspozony: *gag* a RT / RNaseH. Pravděpodobně je za inzerci do hostitelského genomu zodpovědná kódující sekvence pro endonukleázu místo integrázy. I přes částečně identické proteinové funkce je pro transpozici použit jiný mechanismus, než je tomu u LTR retrotranspozonů. Naznačují to terminální poly(A), chybějící LTR a doprovodné přímé obrácené repetice. Za velkou část repetitivní DNA u rostlin jsou zodpovědné LINE elementy (Becker et al., 2001).

- LINE

Dlouhé rozptýlené jaderné elementy jsou sice jednodušší než LTR retrotranspozony, ale stále kódují některé stejné proteiny. LINE mají geny *gag* a *pol*, ale postrádají integrázu. Místo za genem *gag* obvykle zaujímá protein s endonukleolytickou aktivitou, který je zapojen do integrace DNA LINE elementů do chromozomální DNA (Cost a Boeke, 1998). LINE se také mohou integrovat do chromozomální DNA v místě zlomů a tím je opravovat (Moore a Haber, 1996). Podle kritérií sekvenční diverzity se elementy LINE zdají být nejstarší třídou eukaryotických retrotranspozonů. Kladistické studie naznačují, že první LTR retrotranspozony mohly vzniknout získáním LTR (Xiong a Eickbush, 1990).

O mechanismech transkripce nebo integrace LINE elementů u rostlin je známo málo. Většina LINE, které byly u rostlin pozorovány, je zkrácena na 5' konci, stejně jako asi 95 % lidských prvků L1 (Schmid, 1998), což naznačuje podobné procesy transkripce,

replikace a integrace (Kumar a Bennetzen, 1999). Elementy LINE jsou autonomní (Mita a Boeke, 2016).

- SINE

Krátké rozptýlené jaderné elementy jsou malé retrotranspozony, které se od ostatních tříd liší tím, že nekódují žádné proteiny nezbytné pro transpozici, jsou tedy neautonomní. Proto se také někdy nazývají retrogeny (Becker et al., 2001). Proteiny nutné k mobilizaci musí být dodány *in trans* a obvykle jsou jejich zdrojem autonomní non-LTR retrotranspozony. Například lidské elementy Alu1 jsou mobilizovány proteinovým aparátem autonomních L1 elementů. Všechny známé SINE jsou odvozeny od produktů RNA polymerázy III (např. tRNA) (Boeke, 1997). V tomto ohledu jsou SINE nejpodobnější pseudogenům bez intronů, molekulám mRNA, které jsou občas reverzně transkribovány, a poté vloženy do eukaryotických genomů (Drouin a Dover, 1987; Loguercio a Wilkins, 1998). SINE jsou u rostlin nejběžnějšími neautonomními retrotranspozony (Lisch, 2013).

- PLE

Elementy podobné Penelope (PLE) jsou široce rozšířené, ale ne příliš studovanou třídou eukaryotických TE. Jsou charakterizované jedním ORF kódujícím RT, neobvyklou GIY-YIG EN doménou, která byla nalezena také v bakteriálních intronech skupiny I, a přítomností sestřihovaných intronů u několika typů (Arkhipova et al., 2003; Evgen'ev a Arkhipova, 2005). Ve fylogenezi retroelementů zauímají zvláštní místo tím, že sdílejí společného předka s TERT (telomerázové RT, geny, které syntetizují telomery eukaryot) (Arkhipova et al., 2003). PLE se vkládají do genomu náhodně, přičemž upřednostňují cíle bohaté na AT (Evgen'ev a Arkhipova, 2005). To je způsobeno slabou sekvenční preferencí konzervativních domén endonukleázy (Pyatkov et al., 2004).

2.1.3. DNA transpozony

Prvky třídy II, také známé jako DNA transpozony, jsou přesouvány pomocí meziprojektu DNA. Buď přímo prostřednictvím mechanismu „vyjmout a vložit“ (Greenblatt a Brink, 1963; Rubin et al., 1982) nebo v případě Helitronů „mechanismem otáčivé kružnice“, replikačním mechanismem zahrnujícím kruhový meziprojekt DNA (Grabundzija et al., 2016). U DNA transpozonů Krypton a Maverick/Polinton zatím nebyl mechanismus prozkoumán.

Pro TE II. třídy je charakteristická přítomnost koncových obrácených terminálních invertovaných repeticí (TIR) na obou koncích, které jsou dlouhé 9 až 40 párů bází, a jsou rozpoznávány transpozázou (Pray, 2008). Dále obsahují doprovodné přímé repetice. Přímé repetice nejsou ve skutečnosti součástí TE, ale jsou to pozůstatky po vložení TE. Jakmile jsou TE vyříznuty, zůstanou tyto repetice jako „stopy“. Občas tyto stopy ovlivňují genovou expresi i poté, co se jejich TE přesunula na jiné místo v genomu (Pray, 2008). Lidský genom se skládá převážně z retrotranspozonů, z DNA transpozonů jsou to méně než 2 % (Kazazian a Moran, 1998).

Helitrony

Helitrony jsou eukaryotické DNA transpozony. Mají unikátní strukturu a předpokládá se, že používají mechanismus otáčivé kružnice, podobný replikaci, prostřednictvím meziprojektu jednovláknové DNA (Feschotte a Wessler, 2001; Kapitonov a Jurka, 2001). Vyskytují se v různých eukaryotách, ale v největším zastoupení byly nalezeny u netopýra hnědavého a v kukuřici, kde tvoří několik procent genomu (Craig, 2015). Helitrony se akumulují blízko sebe, ale ne uvnitř sebe (Yang a Bennetzen, 2009).

Helitrony zachycují hostitelské geny nebo jejich fragmenty nacházející se v okolí. Může to být díky mechanismu otáčivé kružnice, který nemusí končit na 5' konci ale pokračovat dále. Když dojde k ukončení replikace, může mobilizovaná DNA nově zahrnovat sekvence, které byly umístěny až za 5' koncem Helitronu. Toto zachycování a následná duplikace genů může být velmi důležité pro evoluci druhů (Du et al., 2009; Yang a Bennetzen, 2009). Nicméně mechanismus transpozice otáčivé kružnice zůstává od doby zveřejnění dobře podporovanou, ale dosud nepotvrzenou hypotézou (Craig, 2015).

Helitrony byly první skupinou TE, která byla objevena výpočetní analýzou celých genomových sekvencí (Kapitonov a Jurka, 2001), a zahájily tak éru objevování TE za pomoci počítačových výpočtů. Zároveň s tím se věda posunula od studia jednotlivých elementů a rodin, ke studiu celých populací TE a dalších významných složek genomu. Výpočetní studie znamenaly revoluci v tom, co víme o složení genomů, distribuci a životním cyklu TE.

DNA transpozony „Vyjmout a vložit“

Tyto TE se transponují pomocí mechanismu „vyjmout a vložit“, ve kterém je TE fyzicky vyříznut z chromozomu a znovu začleněn do nového umístění. Proces katalyzuje enzym

transpozása, který je kódován transpozonom. Efektivní replikace je dosaženo tím, že hostitel opraví dvojřetězcový zlom způsobený vyříznutím TE. U rostlin jsou jedněmi z nejčastějších DNA elementů členové superrodiny hAT (hobo, Activator a Tam3) (Pray, 2008). Neautonomní verze těchto elementů nesou minimální sekvence nezbytné pro transpozici a jsou krátké jen několik stovek párů bází. Nejběžnějšími neautonomními elementy v rostlinných genomech jsou miniaturní invertované repetice (MITE) (Lisch, 2013).

- hAT

hAT transpozony jsou první objevené transpozony vůbec. Řadí se mezi transpozony třídy II. a jsou běžné v genomech rostlin, živočichů i hub. Prvním popsáným transpozonom byl Ac u kukuřice, který poprvé popsala Barbara McClintocková (McClintock, 1950). Název hAT je odvozen od tří členů této nadrodiny: hobo z octomilky, disociátoru (Ds) nebo aktivátoru (Ac) z kukuřice a elementu Tam3 z hledíku většího (Rubin et al., 2001). hAT elementy mají 2.5 – 5 kb s krátkými TIR a po integraci do cílové DNA sekvence vytvářejí duplikace cílového místa (TSD) dlouhé 8 párů bází (Atkinson et al., 1993; Coen et al., 1989).

hAT transpozony jsou sice distribuovány napříč eukaryotickými genomey, ale ne ve všech organismech jsou aktivní. Příkladem jsou savčí genomey (včetně lidského), které obsahují neaktivní hAT (Arensburger et al., 2011). hAT jsou velmi staré a předpokládá se, že byly přítomny už v genomu předků obratlovců (Chalopin et al., 2015). Tak jako byl element Ac prvním hAT transpozonom nalezeným u rostlin, element hobo byl jako první nalezen u živočichů a zůstává jako jeden z mála aktivní (Arensburger et al., 2011).

- Tc1/ mariner

Tc1/ mariner elementy patří do skupiny DNA transpozonů „vyjmout a vložit“. Byly poprvé pozorovány u hmyzu a hlístic, Tc1 transpozon u háďátka obecného a mariner transpozon u octomilky. Elementy této třídy se vyskytují u všech živočichů (Robertson, 1995), zvláště jsou rozšířeny u obratlovců (včetně člověka). Distribuce mezi druhy je ale nerovnoměrná a tyto transpozony se mohou vyskytovat u blízce příbuzných druhů nebo zcela chybět (Robertson a Lampe, 1995). Obvykle mají 1.5 kb, kódují transponázu, která je lemována TIR o 30 párech bází (Craig, 2015).

Krypton

Kryptony představují jedinečnou třídu DNA transpozonů využívajících tyrosin rekombinázu k rozštěpení a opětovnému připojení rekombinujících molekul DNA. Kryptony byly původně identifikovány v houbách a později v mořských sasankách, mořských jezcích a hmyzu (Kojima a Jurka, 2011). Dosud je o Kryptonech poměrně málo známo. Nebyly provedeny žádné studie jejich translace, transkripce, transpozice ani regulace. Sekvenční podobnost mezi Kryptony je velmi nízká, zejména v oblastech nekódujících protein (Kojima a Jurka, 2011).

Maverick/ Polinton

Polintony (také nazývané Maverick) jsou největší a nejsložitější známé DNA transpozony. Název byl odvozen od dvou klíčových proteinů, DNA polymerázy a retrovirové integrázy (Kapitonov a Jurka, 2006). Kódují až 10 jednotlivých proteinů. Občas jsou označovány jako „samosyntetizující“ transpozony, protože kódují proteiny nezbytné ke své replikaci (Krupovic a Koonin, 2016). Obsahují geny, které jsou homologní s virovými proteiny a často se vyskytují v eukaryotických genomech (kromě Archaeplastida – červené a zelené řasy, glaukofyty a vyšší rostliny). Typický Polinton má 15-20 kb (Haapa-Paananen et al., 2014). Polintonové sekvence obsahují TIR dlouhé 100-1000 párů bází (Krupovic a Koonin, 2015).

2.2. Regulace aktivity TE

2.2.1. Epigenetika

Termín epigenetika označuje změny v genové expresi (a případně fenotypu), které nelze přímo připsat změnám v sekvenci DNA (Jaenisch a Bird, 2003). Tyto změny jsou řízeny několika typy epigenetických mechanismů. Týkají se nejen genů, ale i transponovatelných elementů a dalších sekvencí, a odehrávají se na úrovni kontroly transkripce nebo ve formě posttranskripčního umlčení.

2.2.1.1. Transkripční umlčování genů

Transkripční umlčování genů (z anglického Transcriptional gene silencing, TGS) označuje stabilní potlačení transkripce a postihuje hlavně transpozony, tandemové repetyce a transgenní inzerty. Neméně důležitou roli zaujímá v řízení tkáňově nebo časově specifické exprese genů kódujících proteiny (Nishimura et al., 2012).

U mnohobuněčných eukaryot je obvyklé, že TGS přetrvává i po mitotickém dělení a v některých případech bývá děděna i transgeneračně - meioticky. Takové stability TGS je dosaženo společným působením několika epigenetických mechanismů, které vytvářejí a udržují konkrétní vzorce modifikací cílové DNA a histonových proteinů během replikace DNA (Law a Jacobsen, 2010). Dvěma hlavními mechanismy TGS jsou tedy metylace DNA a modifikace histonů.

Metylace DNA

Metylace DNA je nejlépe charakterizovaná forma epigenetické modifikace. Probíhá v poloze 5 cytosinu a u živočichů se obvykle vyskytuje v kontextu CG. U rostlin najdeme metylované cytosiny ve všech kontextech sekvencí, CG, CHG a CHH, kde H označuje bázi C, A nebo T (Bird, 2002; Law a Jacobsen, 2010). To je důvod proč mají rostliny složitější a redundantnější epigenetické umlčovací mechanismy než živočichové (Law a Jacobsen, 2010; Simon a Meyers, 2011; Zaratiegui et al., 2007). Metylací DNA nejčastěji nacházíme v regulačních oblastech genů, promotorech, a v některých případech i v sekvenci kódující proteiny. Promotor obsahující metylované cytosiny je obecně spojován s transkripčně inaktivním, umlčeným stavem genu, a to buď prevencí vazby transkripčních faktorů nebo přitahováním proteinů vázajících se na metylové skupiny cytosinů, které rekrutují korepresory transkripce (Bird, 1986; Klose a Bird, 2006). Metylace nacházející se v kódující oblasti genu naopak koreluje se stálou mírnou expresí genu, ale mechanismus účinku dosud není objasněn. Metylace DNA hraje důležitou roli v mnoha fyziologických jevech, jako je například genomický imprinting nebo inaktivace jednoho ze dvou samičích chromozomů X u savců (Jaenisch a Bird, 2003). Metylace DNA také hraje klíčovou roli v dalších buněčných procesech, jako je buněčná diferenciace a tkáňově specifická exprese genů. Metylace je tedy běžně detekována ve tkáňově specifických a zárodečně specifických genech, X-vázaných genech a imprintovaných genech, ale zdaleka nejtypičtějším cílem metylace DNA jsou transpozony (Jaenisch a Bird, 2003). Metylace DNA je relativně stabilní ve srovnání s histonovými modifikacemi (Tuesta a Zhang, 2014).

Histonové modifikace

Kromě metylace DNA hrají při regulaci genové exprese důležitou roli také kovalentní modifikace histonů. U eukaryot jsou DNA a histony spojeny do nukleosomů, základní jednotky chromatinu. Nukleosomy jsou složeny ze 147 párů bází DNA ovinutých kolem oktameru čtyř jaderných histonových proteinů: H2A, H2B, H3 a H4 (Kornberg a Lorch,

1999). Mezi hlavní modifikace histonů patří acetylace, metylace, ubikvitinace a fosforylace (Bannister a Kouzarides, 2011). Tyto modifikace mohou buď aktivovat nebo potlačit transkripci v závislosti na modifikaci a specifických zbytcích substrátu (Margueron et al., 2005; Martin a Zhang, 2005; Zhang, 2003; Kouzarides, 2007) a mohou být uloženy nebo odstraněny velkou rodinou enzymů modifikujících histonové proteiny. Acetylace histonů, ke které dochází u určitých zbytků lysinu histonů H3 a H4, je jednou z běžných forem modifikace histonu spojenou s transkripční aktivací. Po acetylaci je chromatin dekonzenzován kvůli neutralizaci pozitivně nabitých zbytků lysinu v histonových zbytcích (Tuesta a Zhang, 2014). Naopak dimethylace lysinu 9 na histonu H3 (H3K9me2) je dominantním signálem pro transkripční inaktivaci a je spojována s umlčením transpozonů.

2.2.1.2. Posttranskripční umlčování genů

Posttranskripční umlčování genů (z anglického Post-Transcriptional Gene Silencing, PTGS) je jedním z mechanismů, které degradují specifické mRNA, a tím snižují expresi genu. PTGS pouze snižuje akumulaci mRNA v rostlinné cytoplazmě, ale neovlivňuje její transkripci (de Carvalho et al., 1992; Van Blokland et al., 1994). PTGS ovlivňuje expresi všech transkribovaných elementů (geny, nekódující RNA, transpozony) ale i transgenů. Mechanismus PTGS začíná dvouvláknovou RNA (dsRNA), která může vznikat buď z vláskových struktur RNA nebo hybridizací opačně orientovaných molekul RNA. Dvouřetězcové molekuly RNA jsou následně štěpeny na malé interferující RNA (short interfering RNA, siRNA), které se vážou na komplementární mRNA a označují ji tak pro degradaci (Fire, 1999; Vaucheret et al., 2001). Tohoto faktu se využívá například pro výzkumné účely, kdy se zavedením transgenů nebo dvouřetězcové RNA (dsRNA) do hostitele může spustit posttranskripční umlčení všech homologních hostitelských genů anebo transgenů. Transgenem indukované umlčení genu bylo poprvé detekováno u rostlin a stálo za objevem PTGS (Van Der Krol et al., 1990). U rostlin může být PTGS také indukováno viry exprimujícími hostitelské geny v procesu zvaném virově indukované umlčení genů (Baulcombe, 1996). Samotné viry mohou být cílem PTGS mechanismů (Baulcombe, 1996), a proto si mnoho virů vyvinulo strategie, jak působit proti PTGS a úspěšně infikovat rostliny (Vaucheret et al., 2001).

2.2.2. Epigenetická regulace transpozonů u rostlin

Epigenetické mechanismy a přírodní výběr společně definují rozsah, v němž mohou transponovatelné elementy ovlivňovat evoluci genomu (Underwood et al., 2017). TE jsou u rostlin, stejně jako u většiny eukaryot, zdaleka nejvariabilnějšími částmi genomu. Tvoří dominantní část rostlinných genomů a prostřednictvím vln expanzí různých rodin TE v čase mohou být příčinou dramatických rozdílů ve velikosti nebo i celkové architektuře genomů u blízce příbuzných druhů rostlin (Bennetzen et al., 2005). Kromě hrubých účinků na celkovou architekturu genomů může aktivita TE způsobovat širokou škálu změn v genové expresi a funkci, od jemných kvantitativních účinků na expresi genů, po tvorbu nových regulačních genových sítí a vývoj zcela nových genů (Yao et al., 2001). Změny navozené aktivitou TE však nelze považovat za inherentně pozitivní, protože TE působí v první řadě jako inzerční mutageny, narušují kódující genové sekvence nebo mohou negativně ovlivnit expresi přilehlých genů. Cílem buněčných organismů je proto prostřednictvím epigenetických mechanismů dosáhnout stavu, aby většina přežívajících změn způsobených TE byla selektivně neutrální nebo jen mírně škodlivá (Tollis a Boissinot, 2012). Nicméně, rozsah potenciálně užitečné variability generované TE naznačuje, že tyto endogenní mutageny podstatně přispěly k evoluci svých hostitelů (Lisch, 2013).

TE v genomech eukaryot mohou existovat ve více epigenetických stavech a jsou přepínány z transkripčně aktivního stavu, který může vést k transpozici, do transkripčně inaktivního - umlčeného stavu. Epigenetický stav TE u rostlin je závislý na několika faktorech. Prvním z nich je umístění dané kopie elementu (Sigman a Slotkin, 2016). TE nacházející se v repetitivních oblastech chudých na geny, tzv. ostrovech TE, bývají konstitutivně heterochromatinizovány. V tomto trvale umlčeném stavu zůstávají díky působení udržovací metylace DNA, která během replikace DNA kopíruje metylaci cytosinu na nově vznikajícím vlákně podle vzoru na templátovém vlákně DNA. Udržovací metylace je závislá na přítomnosti DNA remodelačních proteinů, jako je DDM1 (Decrease in DNA Methylation 1) (Zemach et al., 2013) a je charakterizována metylací v kontextu CG. Heterochromatický charakter ostrovů TE přetrvává mitotická dělení a ze značné části se přenáší i do další pohlavní generace. U TE nacházejících se v blízkosti genů je situace odlišná. Rostlinné geny mohou podléhat metylaci pomocí silných a ne vždy zcela specifických metylačních aparátů nebo je třeba geny epigeneticky aktivovat během diferenciací buněk, a proto jsou genové sekvence aktivně

demetylovány. Tato aktivní demetylace genů často přechází do okolních oblastí a vede k transkripční aktivaci blízkých TE. Reaktivované TE jsou potenciálně nebezpečné, a proto jsou opětovně metylovány pomocí *de-novo* metylace DNA, pro niž je typický metylační kontext CHG a CHH. *De-novo* metylace je zajišťována pomocí mechanismu RNA řízené methylyace DNA (RNA-Directed DNA-Methylation, RdDM) (Gent et al., 2014, 2013; Li et al., 2015; Vergara a Gutierrez, 2017; West et al., 2014). Klíčovým prvkem RdDM jsou krátké interferující molekuly RNA (siRNA), které vykazují homologii k danému TE a vznikají buď z mikro RNA genů, z transkriptů specializovaných RNA polymeráz (například RNA-pol IV) nebo degradací dsRNA. Pomocí RdDM jsou umlčovány také nově začleněné kopie TE nebo TE, které se objevily v genomu nově například následkem mutací, křížení nebo horizontálním genovým přenosem.

Druhým základním faktorem ovlivňujícím epigenetický stav TE je typ buněčného pletiva a fáze životního cyklu rostliny. V somatických buňkách je většina TE udržována v umlčeném stavu charakterizovaném metylací v kontextu CG. Během diferenciac buněčných pletiv dochází ke změnám epigenetického stavu genů a s tím souvisí i vyšší aktivita *de-novo* metylace cílené na TE v blízkosti genů. Proto mají kopie TE nacházející se v blízkosti genů v meristematically aktivních pletivech vyšší míru metylace v kontextu CHG a CHH. Největšími změnami epigenetických stavů TE procházejí během diferenciac pohlavních buněk z květních meristémů. V mateřských buňkách samčích i samičích rostlinných gamet dochází k masivní demetylaci DNA, která má za cíl obnovit totipotentní stav buňky. Této masivní demetylaci podléhají také TE, jež jsou tímto aktivovány. Rostliny si ale vyvinuly obranné mechanismy, aby minimalizovaly riziko inserční mutagenese TE v gametách. Tyto mechanismy jsou založeny na molekulách siRNA, které jsou generovány ve vegetativních buňkách rostlinných gametofytů, centrální buňce zárodečného vaku v samičím gametofytu a ve vegetativním jádře pylu. Tyto siRNA migrují z vegetativních buněk do pohlavních buněk, vajíčka respektive spermatických buněk pylu, kde řídí transkripční a posttranskripční umlčení TE. Tato ochrana funguje v principu podobně u samčích i samičích gamet rostlin, ale existují indicie, že se v některých detailech liší natolik, že se většina TE transgeneračně množí buď převážně v samčí nebo v samičí linii (Kubát et al., 2014; Hobza et al., 2017). Předmětem této práce je nalézt důkazy pro toto pohlavně specifické chování jednotlivých TE a epigenetických regulačních mechanismů.

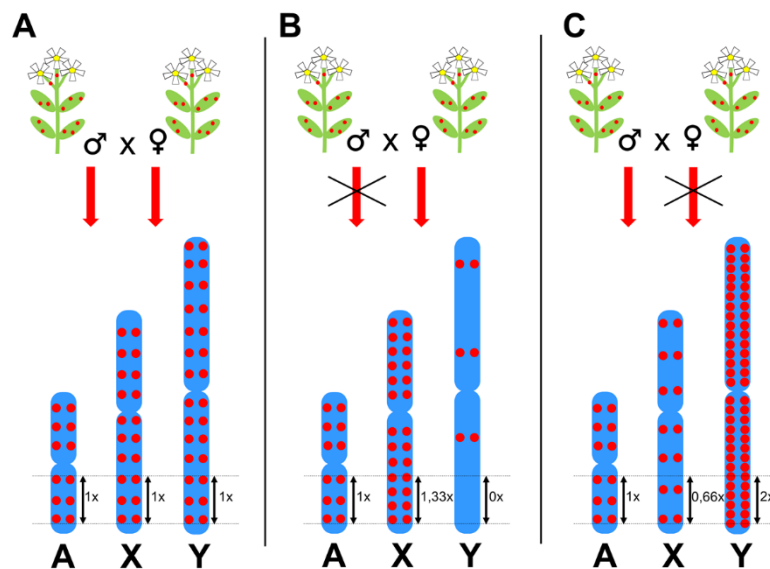
2.3. TE u rostlin s pohlavními chromozomy

Pohlavní chromozomy se vyvinuly mnohokrát nezávisle u rostlin i živočichů z dvojice běžných autozomů. Na rozdíl od živočichů jsou oddělená pohlaví a pohlavní chromozomy u rostlin velmi vzácné (Puterova et al., 2018). Pohlavní chromozomy byly identifikovány u 40 druhů. Pouze 19 druhů má heteromorfní pohlavní chromozomy (Ming et al., 2011; Renner, 2014) a většina z nich nese velký chromozom Y, který svědčí o raném stadiu vývoje pohlavních chromozomů (Hobza et al., 2015). Chromozom Y je jedinečnou součástí genomu, protože obsahuje oblasti, v nichž nedochází k rekombinaci (Bachtrog, 2013). Hlavní fáze vývoje chromozomů Y začíná založením oblasti určující pohlaví, kterému následuje lokální potlačení rekombinace, hromadění repetitivních sekvencí, degenerace genů, a nakonec smršťování delecí již nepotřebných geneticky degenerovaných lokusů. Expanze a smršťování jsou paralelní procesy ve formování struktury chromozomu Y, které mění svůj vliv na proměny chromozomu Y v jednotlivých fázích jeho evoluce (Vyskot a Hobza, 2015). Expanze převážně nerekombinujících oblastí pohlavních chromozomů je často doprovázena akumulací repetitivních sekvencí. To má za následek nejen zvětšování pohlavních chromozomů, ale může stát i za významnou variabilitou velikosti genomu mezi blízce příbuznými dvoudomými a jednodomými druhy, jak bylo ukázáno u silenky (Cegan et al., 2012) a chřestu (Harkess et al., 2016).

I přesto, že byl vývoj pohlavních chromozomů předmětem četných studií, dopad repetitivních sekvencí na architekturu pohlavních chromozomů není zcela objasněn. Nedávné studie však ukázaly, že jak expanze, tak smršťování chromozomu Y mohou být ovlivněny pohlavně specifickým šířením mobilní DNA, zejména transpozonů (Puterova et al., 2018). Prvotní náznaky pohlavně specifického šíření TE byly objeveny pomocí experimentů s fluorescenční *in situ* hybridizací (FISH), které byly prováděny na dvoudomých rostlinách s heteromorfními pohlavními chromozomy. Zejména některé retrotranspozony vykazují slabší nebo silnější hybridizační signál na pohlavních chromozomech, ve srovnání s autozomy dvoudomých druhů *Rumex* a *Silene* (Cermak et al., 2008; Kejnovsky et al., 2009; Kralova et al., 2014; Steflova et al., 2013). Konkrétně TE rodiny, které na chromozomech Y chybí, jsou poměrně časté na X (Cermak et al., 2008; Filatov et al., 2009; Steflova et al., 2013). Podrobným rozbořením velkého množství výsledků FISH v těchto publikacích bylo zjištěno, že TE, které jsou nedostatečně zastoupeny na chromozomech Y, jsou současně akumulovány na chromozomech X.

A naopak, TE akumulované na chromozomech Y mají méně kopií na X. Tato pozorování silně podporují hypotézu, že se TE šíří přednostně buď v samčí nebo samičí linii.

Na tomto základě byl vytvořen model, který popisuje, jak proliferace TE závislá na pohlaví ovlivňuje jejich distribuci na X, Y a autozomech (Obr. 4) (Hobza et al., 2017). Model je založen na skutečnosti, že v typické populaci je pravděpodobnost přítomnosti autozomů u samců nebo samic 1/2. V případě chromozomů X je pravděpodobnost u samců 1/3 a u samic 2/3. Pro chromozomy Y je to 1 u samce a 0 u samice. Tento model byl částečně doložen u *S. latifolia* pomocí bioinformatické analýzy složení samčích a samičích genomů, kdy bylo na chromozomu X zjištěno přibližně 1,5 krát více kopií LTR retrotranspozonu OGRE CL5 než na autozomech (Puterova et al., 2018). Předchozí výsledky naznačují, že OGRE CL5 je odlišně regulován epigenetickými mechanismy během formování zárodečných linií samců a samic (Kubat et al., 2014).



Obr. 4 Model proliferace transponovatelných elementů (TE) závislé na pohlaví. Tři scénáře transgeneračního šíření TE v rostlinách a jejich dopad na chromozomální distribuci TE. Červené tečky označují inserce TE - v modře zbarvených meiotických chromozomech, A-autozomech, X-chromozomech a Y-chromozomech. Čísla vedle dvojitých šipek označují očekávanou hustotu insercí TE na jednu jednotku délky příslušného chromozomu. (A) Pokud je TE přenášen na potomstvo stejně u samců i samic, je hustota insercí TE na všech chromozomech identická. (B) Pokud je proliferace TE narušena u samců, je hustota TE 1.33krát vyšší na chromozomu X než na autozomech, ale téměř nulová na chromozomu Y. (C) Pokud je proliferace TE přerušena u samic, hustota insercí TE je nižší na chromozomu X ve srovnání s autozomy a dvakrát vyšší na chromozomu Y než na autozomech. Scénáře B a C představují extrémní případy pohlavně specifické aktivity TE. Reálné TE se pohybují na plynulé škále mezi těmito extrémami. Převzato z (Hobza et al., 2017).

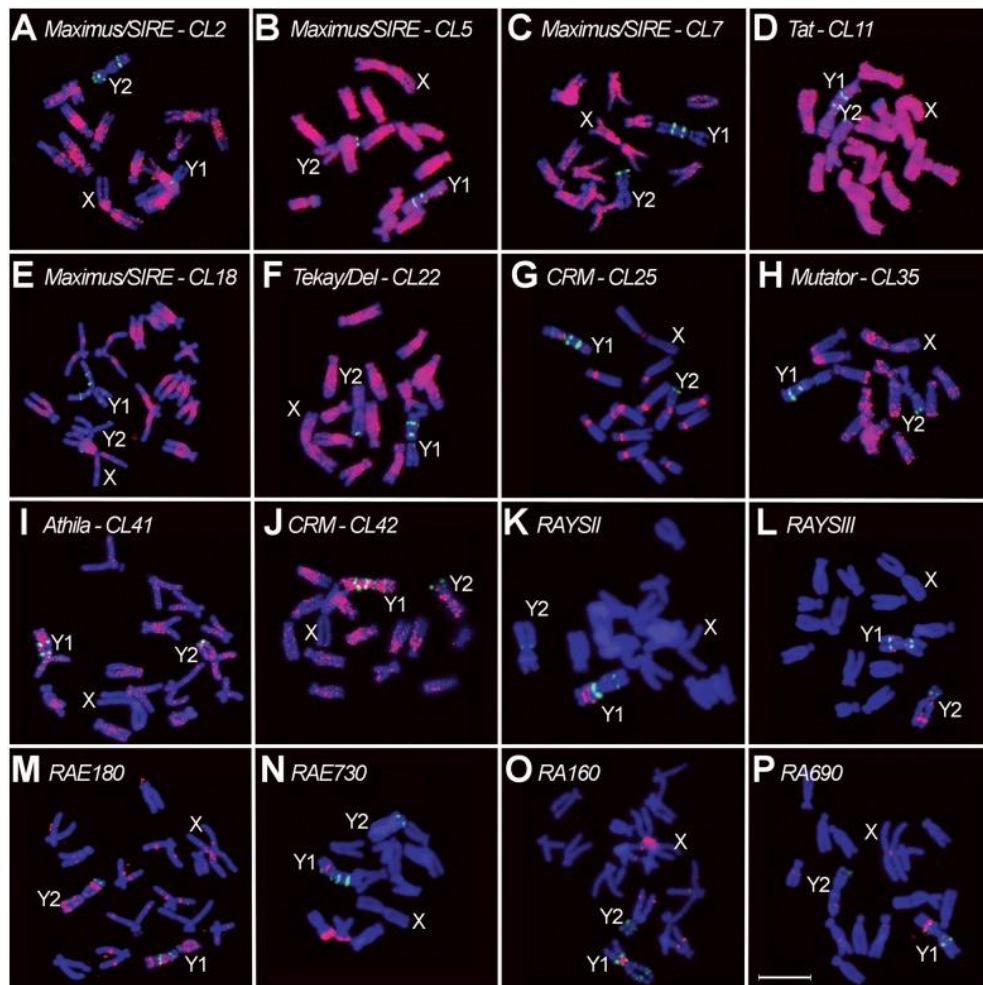
Rumex acetosa

V této práci jsme si dali za cíl potvrdit výše zmíněný model u *Rumex acetosa* (šťovík kyselý), dvoudomé rostliny s vysoce diferencovaným pohlavním chromozomovým systémem, ve kterém jsou samice $2n = 12 + XX$, zatímco samci jsou $2n = 12 + XY_1Y_2$. Na rozdíl od mnoha jiných živočichů a rostlin s pohlavními chromozomy, pohlaví *R. acetosa* není kontrolováno aktivním chromosomem Y, ale je řízeno rovnováhou mezi počtem chromozomů X a počtem autozomů (poměr X/A) (Manzano et al., 2017). Pohlavní chromozomy jsou z celého karyotypu největší, a zároveň je X o trochu větší než oba chromozomy Y. Dohromady, jsou ale oba Y větší než X. Chromozomy Y *R. acetosa* mají podobnou cytogenetickou strukturu. Oba jsou heterochromatické, degenerované a vysoce diferencované ve srovnání s euchromatickým X (Clark et al., 1993; Lengerova a Vyskot, 2001; Mosiołek et al., 2005; Rejčín et al., 1994).

Složení samčího a samičího genomu *R. acetosa* bylo již dříve studováno pomocí masivního paralelního sekvenování celogenomové DNA (Tab. 1). Identifikace a charakterizace hlavních skupin repetitivní DNA *R. acetosa* ukázaly, že největší zastoupení mají retrotranspozony typu copia a gypsy, následované DNA transpozony a non-LTR retrotranspozony. V gypsy superrodině převládaly CRM a Tat / Ogre retrotranspozony, zatímco u copia retrotranspozonů byly nejhojnější Maximus / Sireviry. Na chromozomech Y došlo k akumulaci gypsy (CRM), zatímco copia retrotranspozony, které jsou v genomu velmi hojné, chyběly nebo nebyly na chromozomu Y přítomny (Steflova et al., 2013).

Tab. 1 Zastoupení repeticí v genomu *Rumex acetosa* odhadnuté ze sekvenování genomové DNA pomocí technologie Illumina (Steflova et al., 2013).

Klasifikace			Poměr genomu	
Typ repeticie	Superrodina	Rodina	Samec (%)	Samice (%)
Retrotranspozony	Gypsy	Chromovirus – CRM	5.73	5.38
		Chromovirus – Tekay/Del	0.34	0.36
		Athila	0.48	0.41
		Tat/Ogre	5.36	5.69
	Copia	Maximus/Sire	34.92	35.58
		Bianka	0.19	0.20
		TAR	0.10	0.10
	LINE		0.01	0.01
DNA transpozony	Mutator		1.30	1.33
	CACTA		0.26	0.25
TE celkem			48.69	43.91



Obr. 5 Lokalizace transponovatelných elementů a satelitů na metafázových chromozomech *Rumex acetosa* pomocí FISH. Název každé rodiny transpozonu je spolu s odpovídajícím číslem klastru uvnitř každého obrázku. Měřítko označuje 10 μm .

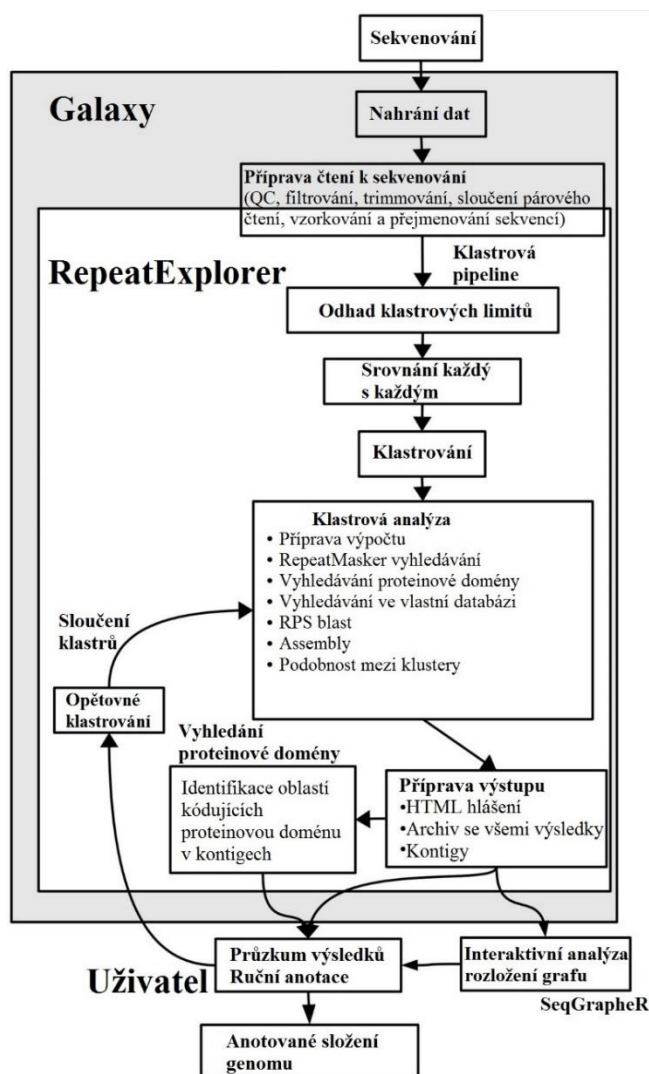
Obrázek získaný FISH dokumentuje rozdílné chromozomální distribuce různých TE. Některé TE mají méně kopií na chromozomech Y (Obr. 5A, B, C, D, E, F) a jiné se na Y naopak kumulují (Obr. 5I, J). V případě satelitů dochází k akumulaci především na chromozomu Y (Obr. 5K, L, N, P) a méně na X (Obr. 5M, O) (Steflova et al., 2013).

V této práci jsme se rozhodli zopakovat analýzu složení genomu *R. acetosa* s využitím masivního paralelního sekvenování chromozomů X, Y a autozomů roztříděných pomocí průtokové cytometrie. Analýza tříděných chromozomů nám umožnila popsat složení těchto různých složek genomu a testovat model pohlavně specifického šíření TE (Obr. 4) navržený v práci (Hobza et al., 2017).

3. EXPERIMENTÁLNÍ ČÁST

3.1. RepeatExplorer

RepeatExplorer (RE) je výpočetní systém, který obsahuje soubor bezplatných nástrojů pro objevování a charakterizaci opakujících se sekvencí v eukaryotických genomech. Systém běží na platformě Galaxy (Goecks et al., 2010) a je přístupný na adrese <http://repeatexplorer.org/>. RE využívá shlukování sekvencí na základě grafů (Novák et al., 2010), které usnadňují *de novo* opakovanou identifikaci bez potřeby referenčních databází známých repetecí, zkoumá fylogenetické vztahy retroelementů a provádí srovnávací analýzu složení repetecí mezi více druhy (Obr. 6). Protože algoritmus na vstupu používá krátké náhodně vzorkované sekvence z genomu, je ideální pro analýzu NGS (Next-Generation Sequencing) dat (Novák et al., 2013).



Obr. 6 Součásti a pracovní postup analýzy RepeatExploreru (Novák et al., 2013).

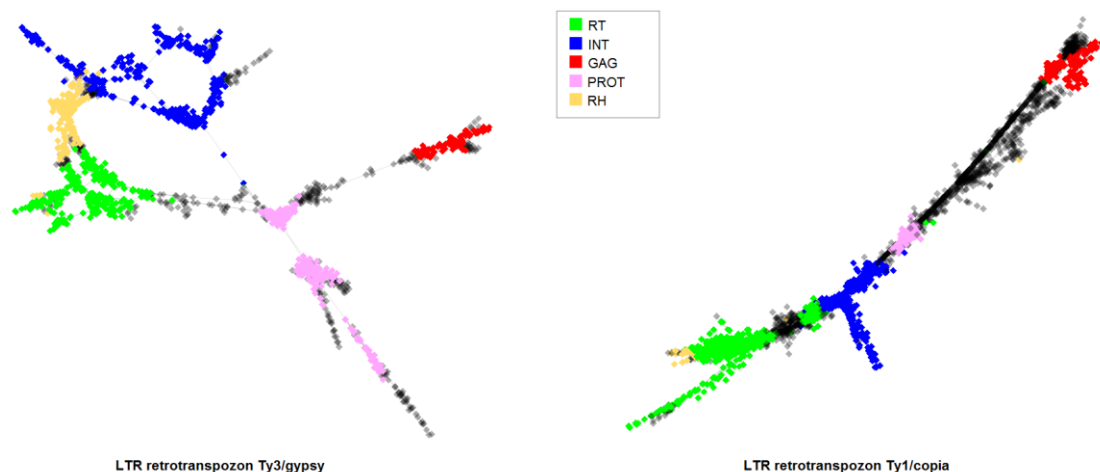
RE také obsahuje různé nástroje pro práci s FASTQ a FASTA sekvencemi, jako je například přejmenování sekvencí, přidání předpony (prefix) a přípony (suffix), sloučení párových čtení. Umí také filtrovat kvalitu dat a ořezat je. Jednotlivé nástroje lze kombinovat a vytvářet vlastní analytické pracovní postupy (pipeline) zaměřené na konkrétní úkoly.

3.2. Klastrová (shluková) analýza

Tato analýza je založena na nalezení a kvantifikaci podobností mezi jednotlivými sekvencemi čtení, které se obvykle získají masivním paralelním sekvenováním náhodně nastříhané genomové DNA. Tyto podobnosti se používají ke konstrukci grafů, ve kterých vrcholy odpovídají sekvenčním čtením. Překrývající se čtení jsou spojeny s hranami a jejich skóre podobnosti jsou vyjádřeny jako hmotnosti hran. Topologie grafu je poté analyzována, aby se identifikovaly a oddělily klastry často propojených čtení (Novák et al., 2010). Vytvořené klastry pak představují rodiny různých opakujících se elementů. Klastrová analýza tedy slouží ke třídění čtení tak, aby si byly čtení z jednoho klastru podobnější, než čtení z jiných klastrů.

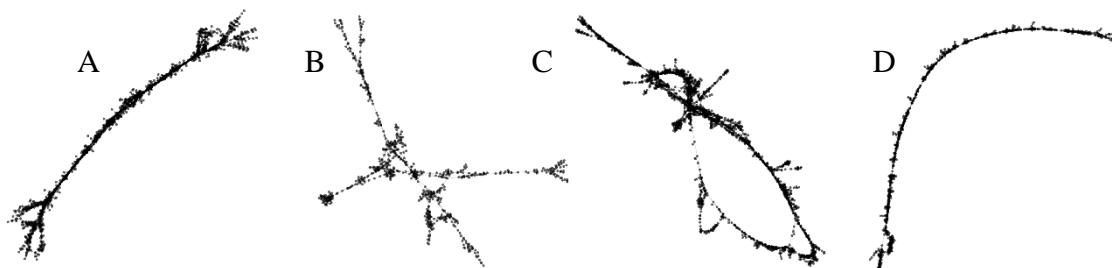
Klastrová analýza je výpočetně náročná. V závislosti na počtu čtení a složení repeticí analyzovaného vzorku může analýza trvat několik hodin, dní a v extrémních případech až dva týdny (pro genomy bohaté na určité typy repeticí, např. satelity) (Novák et al., 2010).

RE poskytuje automatickou anotaci pomocí programu RepeatMasker a Databáze konzervovaných domén (CCD). Na základě pořadí a procentuálního výskytu nalezených domén, dokáže dané klastry určit (Obr. 7).



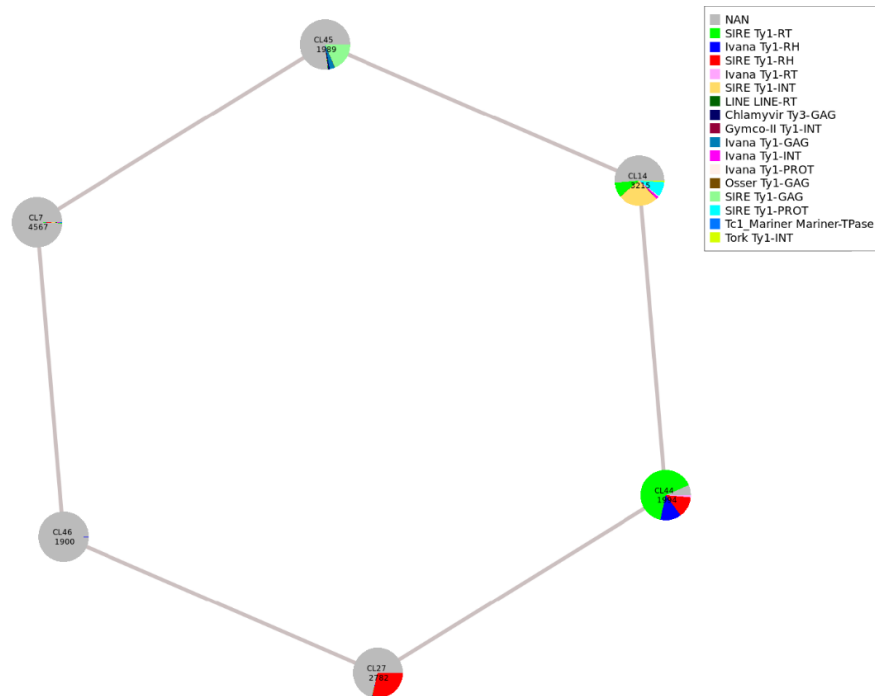
Obr. 7 Klastry RE s identifikovanými doménami. LTR retrotranspozony Ty3/gypsy se od LTR retrotranspozonů Ty1/copia liší rozdílným uspořádáním domén.

Nicméně tato anotace není vždy spolehlivá. Domény nemusí být vždy identifikované nebo v dostatečném zastoupení. Transpozony mohou být roztrhané na menší části (Obr. 8A), které se mohou spojit na základě podobnosti s jinými roztrhanými transpozony (Obr. 8B) a vytvořit tak nový klastř (tvořený čteními z rozdílných rodin TE). Další možností jak anotovat TE je pomocí Blastn nebo ručně. Nicméně určit ručně jen podle grafu, o který transponovatelný element se jedná, je téměř nemožné. LTR transpozony, které mají na svých koncích repetici, by se v grafu měly spojit do „tenkého kruhu“ se zdvojenou LTR částí (Obr. 8C). Bohužel se velmi často stává, že RepeatExplorer dlouhé transpozony roztrhá na více menších klastřů, které se zobrazují jako „čára“ (Obr. 8A). Ostatní transpozony, které koncovou repetici nemají, by se také měly zobrazovat jako „čára“ (Obr. 8D). Často jsou ale také roztrhané do více klastřů. Je tedy téměř nemožné určit, o jaký TE se jedná.

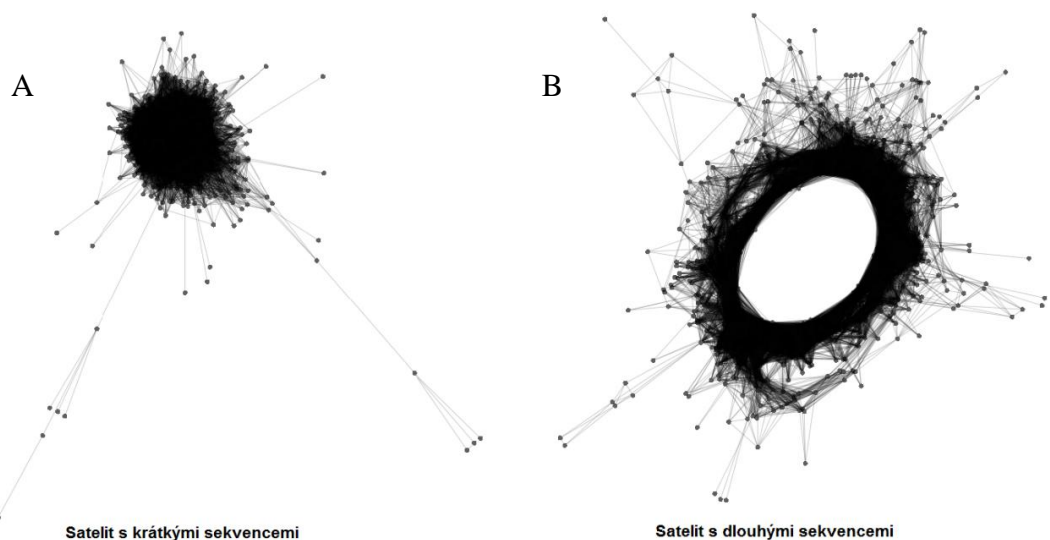


Obr. 8 Různé možnosti klastřů RE. (A) Část transpozonu, který byl při analýze roztržen. (B) Dvě části různých transpozonů, které byly na základě podobnosti sloučeny do jednoho klastř. (C) LTR retrotranspozón. (D) DNA transpozón.

Další způsob jak určit, o jaký transpozon se jedná, jsou superklastry. RE tak řeší to, že některé transpozony trhá do více klastrů. Superklastr je skupina klastrů, které mají hodně spojení (překryvů) přes párová čtení. Párová čtení jsou fyzicky blízko u sebe. Pokud jsou blízko sebe párová čtení z různých klastrů, může to znamenat, že pochází ze stejného elementu. Jednotlivé domény jsou pak rozprostřeny po superklastru (Obr. 9).



Obr. 9 Superklastr LTR retrotranspozonu Ty1/copia, který zahrnuje šest podobných klastrů. U některých klastrů byly identifikovány domény, na jejichž základě jsme poznali, že se jedná o retrotranspozon rodiny Ty1/copia.



Obr. 10 Graf satelitních sekvencí. (A) Satelit s monomerem kratším než je délka čtení. (B) Satelit s monomerem delším než je délka čtení.

To však neplatí pro satelity. Satelity jsou sekvence, které se těsně za sebou neustále opakují. Když je sekvence satelitu kratší než délka čtení, budou se jednotlivá čtení stejného satelitu překrývat. Vznikne nám tak „hustý vyplněný kruh“ (Obr. 10A). Když je sekvence satelitu delší než délka čtení, čtení se v grafu uspořádají do „nevyplněného hustého kruhu“ (Obr. 10B).

3.3. Programovací jazyk R

R je programovací jazyk vhodný pro statistické výpočty, analýzu velkého množství dat a jejich grafického znázornění (<https://www.r-project.org>). K jazyku existuje velké množství knihoven, které rozšiřují možnosti analýzy i grafického zobrazení. R je volně přístupný programovací jazyk. Díky tomu je komunita vývojářů obrovská a vývoj probíhá rychle. Kód lze spustit bez kompilátoru, mnoho výpočtů probíhá pomocí vektorů. Proto je R rychlejší než některé ostatní jazyky. Pomocí jazyka R jsme počítali například Pearsonův korelační koeficient nebo lineární regresi. Také jsme zde tvořili grafy pro potvrzení našeho modelu.

3.4. Metody

3.4.1. Bioinformatická analýza

Surová data dvoudomé *Rumex acetosa* z Illumina MiSeq jsme filtrovali a ořezali podle kvality pomocí nástrojů FastQC (Andrews, 2010) a Trimmomatic (Bolger et al., 2014) (s parametry: illuminaclip 2:30:10, leading a trailing 20, minlen 100, crop 215, slidingwindow 4:15). Získali jsme párová čtení o minimální délce 100 bází a maximální délce 251 bází. Vhodně jsme párová čtení přejmenovali a náhodně vzorkovali za účelem snížení množství dat podle procentuálního zastoupení v genomu: 72.24 % A, 12.04 % X a 15.72 % Y₁Y₂. Výsledný soubor připravený pro klastrovou analýzu byl po úpravách a vzorkování zmenšen na 500 000 párových čtení.

Pro *de novo* identifikaci repetitivních sekvencí jsme použili pipeline RE (Novák et al., 2010). Nastavili jsme hranici pro podrobnou analýzu klastru na > 0.01 % (minimální velikost klastrů, které mají být vybrány je definována jako podíl počtu všech analyzovaných čtení). Zadali jsme také provedení srovnávací analýzy. Zůstalo nám 50.55 % párových čtení, tedy 330 klastrů. Výsledné klastry jsme identifikovali pomocí

automatické anotace RepeatExploreru včetně ruční kontroly a na základě podobností s veřejnými databázemi GenBank. Podařilo se nám identifikovat 205 klastrů z 330.

Vypočítali jsme procentuální zastoupení jednotlivých klastrů na chromozomech X, Y a autozomech a zjistili jsme procentuální zastoupení rodin v knihovnách NGS dat. Pro potřeby této práce předpokládáme, že každý klaster zastupuje buď celou unikátní repetici nebo její část. Dále jsme zjišťovali, jestli jsou dané repeticity rozmístěny na X a Y rovnoměrně, náhodně, anebo se na některém chromozomu kumulují. Následně jsme data očistili o satelitní DNA a vše znovu přepočítali. Výsledné informace jsme pro lepší přehlednost zpracovali pomocí sloupcových nebo bodových grafů.

3.4.2. Statistická analýza

Předpokládá se, že pohlavní chromozomy X a Y se v průběhu evoluce vyvinuly z běžných autozomů (Muller 1914; Ohno 1967). Oproti autozomům jsou však X a Y obohaceny nebo naopak ochuzeny o geny různých funkčních skupin. Příkladem je obohacení na chromozomech X a Y o geny určující pohlaví a reprodukci (Saifi & Chandra, 1999). V této diplomové práci nás zajímá především akumulace transponovatelných elementů a satelitů na chromozomech X a Y, která je způsobena omezením rekombinace na pohlavních chromozomech.

Obohacení pohlavních chromozomů

Z výsledků procentuálního zastoupení jednotlivých repeticí na pohlavních chromozomech jsme zjistili, že se každý klaster odpovídající určité linii repetitivních elementů nenáhodně akumuluje více na jednom nebo druhém pohlavním chromozomu. Tato nenáhodná akumulace repeticí není zřejmá z celkových součtů repeticí každého typu, například LTR retrotranspozonů typu Ty1/copia. Rozhodli jsme se tedy zjistit, kolik procent pohlavních chromozomů X a Y je tvořeno stejnými repeticemi a kolik procent představují chromozomově unikátní repeticity, tedy takové, jež jsou na daném pohlavním chromozomu navíc ve srovnání s druhým pohlavním chromozomem. Pro srovnání jsme výpočty provedli pro data, která obsahují satelity (Tab. 2), a poté pro očištěná data (Tab. 3) (viz Kapitola 4.4.).

Postup pro výpočet hodnot v (Tab. 2) a (Tab. 3) je následující. Součet X a Y označuje procentuální (%) zastoupení daného typu repeticity v příslušném pohlavním chromozomu. Rodiny transpozonů, které jsou na X, ale chybí na Y (nebo jsou na Y a chybí na X) jsme vypočítali jako součet rozdílů % zastoupení repeticity v genomu X a Y pro každý klaster.

Součet kladných hodnot nám udává přebytky na X a součet záporných hodnot zase přebytky na Y. Dále nás zajímalo, jaké procento počtu kopií transpozonů každá repetice zaujímá. Sdílené TE byly spočítány jako rozdíl součtů a přebytků. Procentuální zastoupení TE na X sdílené s Y (nebo % zastoupení TE na Y sdílené s X) jsme spočítali jako podíl sdílených transpozonů a % zastoupení daného typu repetice v příslušném pohlavním chromozomu (součet X nebo součet Y).

Relativní zisk repeticí

Při výpočtu relativního zisku repeticí současných pohlavních chromozomů ve srovnání s ancestrálními autozomy vycházíme ze dvou předpokladů. Prvním předpokladem je, že složení původního (ancestrálního) autozomu, ze kterého se X a Y vyvinuly, je podobné současnému složení autozomů. Druhým předpokladem je, že se absolutní množství nerepetitivní části v evoluci výrazně nezměnilo.

Při výpočtu vycházíme z rovnice (3.1), která nám dává informace o procentuálním množství pohlavních chromozomů (X, Y) a autozomů (A) v samčím genomu. Velikost samčího genomu je přibližně 7335 Mbp. Repetitivní částí genomu rozumíme superrodiny Ty1/copia, Ty3/gypsy, DNA transpozony, nedefinované repetice a satelity. Protože nám předpoklady neříkají nic o tom, jestli je absolutní množství nerepetitivní části ancestrálního autozomu podobné chromozomu X nebo spíše Y, všechny výpočty děláme pro obě varianty, které ve výsledku zprůměrujeme.

$$\frac{10.8 * A + 1.8 * X + 2.35 * Y}{14.95} \quad (3.1)$$

Spočítali jsme množství nerepetitivní a repetitivní části procentuálního zastoupení A, X a Y v genomu, a na základě prvního předpokladu, i jejich absolutního množství v genomu. Absolutní množství nerepetitivní části jsme počítali jako součin velikosti samčího genomu, % zastoupení X (Y) v genomu a množství nerepetitivní části X (Y) v genomu. Absolutní množství repetitivní části jsme vypočítali podobně, avšak místo množství nerepetitivní části X, případně Y, jsme použili množství pro příslušné repetice. Dále jsme počítali velikost (délku) ancestrálního autozomu jako podíl nerepetitivního absolutního množství X (Y) a % zastoupení nerepetitivní části autozomu v genomu. Následně jsme spočítali absolutní množství ancestrálního autozomu pro každou repetici, jako součin velikosti původního autozomu a % zastoupení repeticí na autozomech. Díky

výpočtu velikosti původního autozomu a velikosti současného chromozomu X (Y) jsme dokázali zjistit procentuální nárůst současného X (Y). Obě čísla jsme pak zprůměrovali a určili odchylku.

Pearsonův vážený korelační koeficient

Výpočet Pearsonova korelačního koeficientu je statistická metoda, která měří lineární korelaci (lineární závislost) mezi dvěma náhodnými proměnnými X a Y. Lineární korelace může nabývat hodnot od -1 do 1. Negativní lineární korelace říká, že všechny body leží na přímce, pro kterou Y s rostoucím X klesá. Když je lineární korelace blízko nule, říkáme, že korelace není lineární, tedy neexistuje žádná lineární závislost mezi proměnnými. Pro pozitivní lineární korelaci zase platí, že všechny body leží na přímce, pro kterou se Y zvyšuje, když se zvyšuje X. Pro náš model ale obyčejná Pearsonova korelace nestačí. Každý klastr je různě velký a je v genomu zastoupen procentuálně více či méně. S přihlédnutím na tuto skutečnost jsme počítali Pearsonův vážený korelační koeficient, kdy jsme každému klastru přiřadili vlastní váhu. Váhu jsme zavedli pomocí rovnice (3.1). Pro výpočet váženého korelačního koeficientu jsme použili tři rovnice:

$$m(x; w) = \frac{\sum_i w_i x_i}{\sum_i w_i} \quad (3.2)$$

$$cov(x, y; w) = \frac{\sum_i w_i (x_i - m(x; w))(y_i - m(y; w))}{\sum_i w_i} \quad (3.3)$$

$$corr(x, y; w) = \frac{cov(x, y; w)}{\sqrt{cov(x, x; w) cov(y, y; w)}} \quad (3.4)$$

kde x, y jsou vektory, w je váhový vektor, $m(x; w)$ je vážený průměr, $cov(x, y; w)$ je vážená kovariance (tendence lineárního vztahu mezi dvěma náhodnými proměnnými X a Y) a $corr(x, y; w)$ je vážená korelace.

Grafické znázornění závislosti obohacení Y na obohacení X

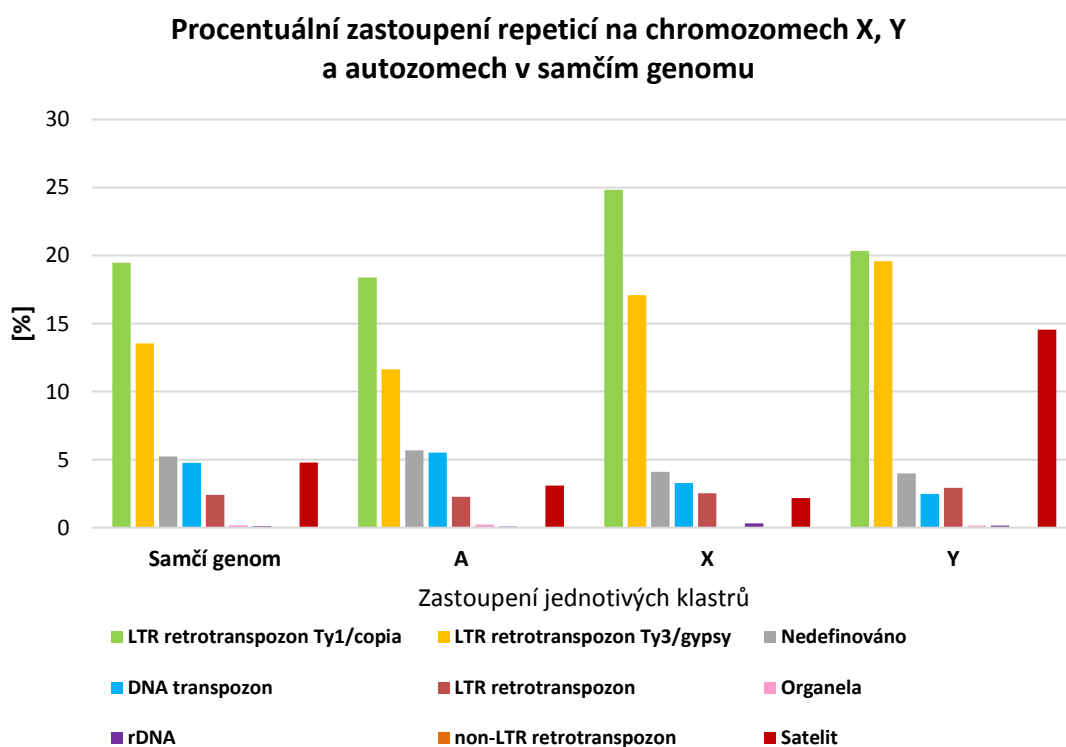
Dále jsme data analyzovali a konfrontovali s modelem pohlavně specifického šíření TE (Hobza et al., 2017) pomocí grafického zpracování v R. Použitá data obsahovala pouze LTR retrotranspozony, pro které by měl model fungovat nejlépe. LTR retrotranspozony dělíme do tří kategorií – Ty1/copia, Ty3/gypsy a LTR retrotranspozony, které neumíme z nedostatku informací do předchozích dvou kategorií zařadit. Pro lepší přehlednost má

každá kategorie svoji barvu. Každému klastru jsme opět přiřadili stupeň důležitosti na základě velikosti klastru. V grafu se to projevuje různou velikostí bodů. Do grafu dosazujeme obohacení na Y, což je podíl procentuálního zastoupení klastru na Y a procentuálního zastoupení na autozomech, a obohacení na X, který je roven podílu % zastoupení na X a % zastoupení na A. Úsečku vedenou grafem představují vektory $x = (1.333, 1, 0.666)$ a $y = (0, 1, 2)$, které odrážejí náš model. Předpokládáme, že se LTR retrotranspozony budou zobrazovat okolo této úsečky.

4. VÝSLEDKY

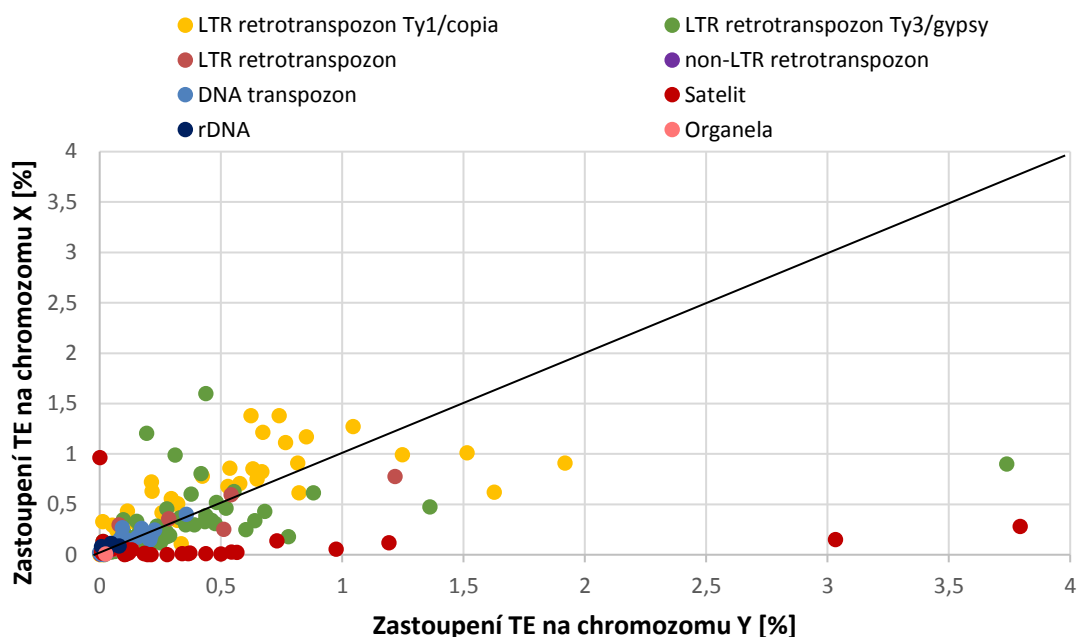
4.1. Složení samčího genomu a jeho komponent

Bioinformatickou analýzou se nám podařilo anotovat 205 klastrů. Největší procento zastoupení v samčím genomu, chromozomech X, Y a autozomech, zaujímají LTR retrotranspozony (Obr. 11). LTR retrotranspozony jsme rozdělili na tři kategorie - dvě superrodiny Ty1/copia a Ty3/gypsy, a LTR retrotranspozony, u kterých jsme nedokázali určit, do které ze dvou superrodin patří. Ve všech určených případech je repetice Ty1/copia zastoupena více než Ty3/gypsy, kromě chromozomu Y, kde zaujímají podobné procento čtení. Na druhém místě jsou DNA transpozony, poté satelity, organely (mitochondrie a plastidy) a rDNA. Přibližně 5 % klastrů se nám nepodařilo zařadit do žádné z podtříd. Z grafu si můžeme všimnout, že satelitní DNA tvoří podstatnou část chromozomu Y (přibližně 14 %), ale na ostatních chromozomech je její podíl nízký (menší než 5 %).



Obr. 11 Procentuální zastoupení repeticí na chromozomech X, Y a autozomech v samčím genomu.

Podíl zastoupení TE na chromozomech Y a X



Obr. 12 Podíl zastoupení transpozonů na chromozomech Y a X. Přímka vedená grafem znázorňuje rovnoměrné zastoupení transpozonů na X a Y.

Dále jsme zkoumali chování repeticí na chromozomu X a Y (Obr. 12). Přímka vedená grafem znázorňuje situaci, kdy je procentuální podíl klastru shodný na obou pohlavních chromozomech X a Y. Klastry nad přímkou tvoří větší podíl chromozomu X a klastry pod přímkou tvoří naopak větší poměrnou část chromozomu Y. Pro lepší přehlednost jsme vynechali nedefinované klastry. Z grafu je patrné, že DNA transpozony jsou rovnoměrně zastoupeny na X i Y (kumulují se na přímkce), zatímco jednotlivé Ty1/copia a Ty3/gypsy LTR retrotranspozony se vyskytují převážně náhodně buď na chromozomu X nebo na Y. To znamená, že každá linie LTR retrotranspozonů se akumuluje s větší pravděpodobností buď na chromozomu X nebo na Y. Zvláštním rysem satelitů je vysoký podíl na chromozomu Y a téměř nulový na X. Na grafu se také nachází odlehlé hodnoty, kdy se v případě satelitů může jednat o velké bloky satelitní DNA jen na několika místech genomu. Odlehlá hodnota u LTR retrotranspozonu Ty3/gypsy může být zapříčiněna vznikem satelitu z TE, kdy se transpozon začne chovat jako satelit.

4.2. Rozdíly ve složení pohlavních chromozomů

Z (Obr. 12) je zřejmé, že většina jednotlivých repeticí má výrazně nevyrovnaný podíl na pohlavních chromozomech. Přitom součty podílů všech repeticí každé z kategorií se mezi pohlavními chromozomy liší méně (Obr. 11).

Zajímalo nás tedy, kolik procent DNA pohlavních chromozomů je tvořeno sdílenými, stejnými, repetitivními elementy a kolik procent DNA pohlavních chromozomů je tvořeno unikátními kopii repetitivních elementů, tedy elementy, které jsou na chromozomu X navíc proti chromozomu Y a naopak. Podrobné srovnání složení pohlavních chromozomů X a Y je zobrazeno v (Tab. 2) a (Tab. 3).

Tab. 2 Podrobné srovnání složení pohlavních chromozomů X a Y před očištěním o satelity.

Typ repetice	Chromozomální DNA [%]					Počet kopií TE [%]	
	součet X	přebytky na X	součet Y	přebytky na Y	sdílené TE	TE na X sdílené s Y	TE na Y sdílené s X
Ty1/copia	24,84	7,81	20,35	3,32	17,02	68,54	83,66
Ty3/gypsy	17,10	4,90	19,59	7,39	12,20	71,35	62,29
DNA transpozony	3,28	0,94	2,47	0,13	2,34	71,33	94,74
Satelity	2,17	1,11	14,56	13,49	1,07	49,13	7,34
Součet	47,39	14,76	56,96	24,33	32,63	68,86	57,29

Tab. 3 Podrobné srovnání složení pohlavních chromozomů X a Y očištěné o satelity.

Typ repetice	Chromozomální DNA [%]					Počet kopií TE [%]	
	součet X	přebytky na X	součet Y	přebytky na Y	sdílené TE	TE na X sdílené s Y	TE na Y sdílené s X
Ty1/copia	25,48	4,08	29,50	8,10	21,40	83,98	72,54
Ty3/gypsy	17,24	3,46	28,28	14,50	13,77	79,91	48,71
DNA transpozony	3,45	0,49	3,58	0,63	2,96	85,66	82,43
Součet	46,17	8,04	61,36	23,24	38,13	82,58	62,13

(Tab. 3) říká, že pohlavní chromozomy X a Y jsou ze 38.13 % tvořeny stejnými repeticemi. Další 8.04 % chromozomu X a celých 23.24 % chromozomu Y je tvořeno unikátními chromozomově specifickými repeticemi. Když se zaměříme na počty monomerů satelitů a jednotlivých kopií transpozonů v (Tab. 2), pak vidíme, že chromozom X jich sdílí 68.86 % s chromozomem Y. Chromozom Y ale sdílí s X jen 57.29 % satelitových monomerů a kopií transpozonů, zbylých téměř 43 % kopií repeticí se nachází unikátně jen na Y. Lze tedy shrnout, že chromozomy X a Y obsahují přibližně 1/3 u X a téměř 1/2 u Y chromozomově specifických repeticí.

4.3. Podíl jednotlivých typů repeticí na zvětšení pohlavních chromozomů

Pohlavní chromozomy X a Y šťovíku kyselého jsou největšími chromozomy u tohoto druhu a předpokládá se, že se nacházejí v expanzivní fázi evoluce. Zajímalo nás, jakým dílem se jednotlivé typy repeticí podílejí na jejich růstu (Tab. 4).

Počítali jsme relativní zisk repeticí současných pohlavních chromozomů ve srovnání s ancestrálními autozomy. Při výpočtech jsme vycházeli z předpokladů, že procentuální složení původního autozomu, ze kterého se pohlavní chromozomy vyvinuly, odpovídá skladbě současných autozomů, a že se v průběhu evoluce příliš nezměnilo absolutní množství nerepetitivní části tohoto chromozomu. (Tab. 4) nám říká, o kolik procent dnešní X a Y obsahují více či méně repeticí v porovnání s ancestrálními autozomy. Rozsah (\pm) zachycuje odlišnost výsledku podle toho, jestli byla pro odhad velikosti nerepetitivní části ancestrálního chromozomu použita velikost nerepetitivní části současných chromozomů X nebo Y.

Tab. 4 Relativní nárůst podílu různých repeticí na současných pohlavních chromozomech X a Y ve srovnání s předpokládanými ancestrálními autozomy.

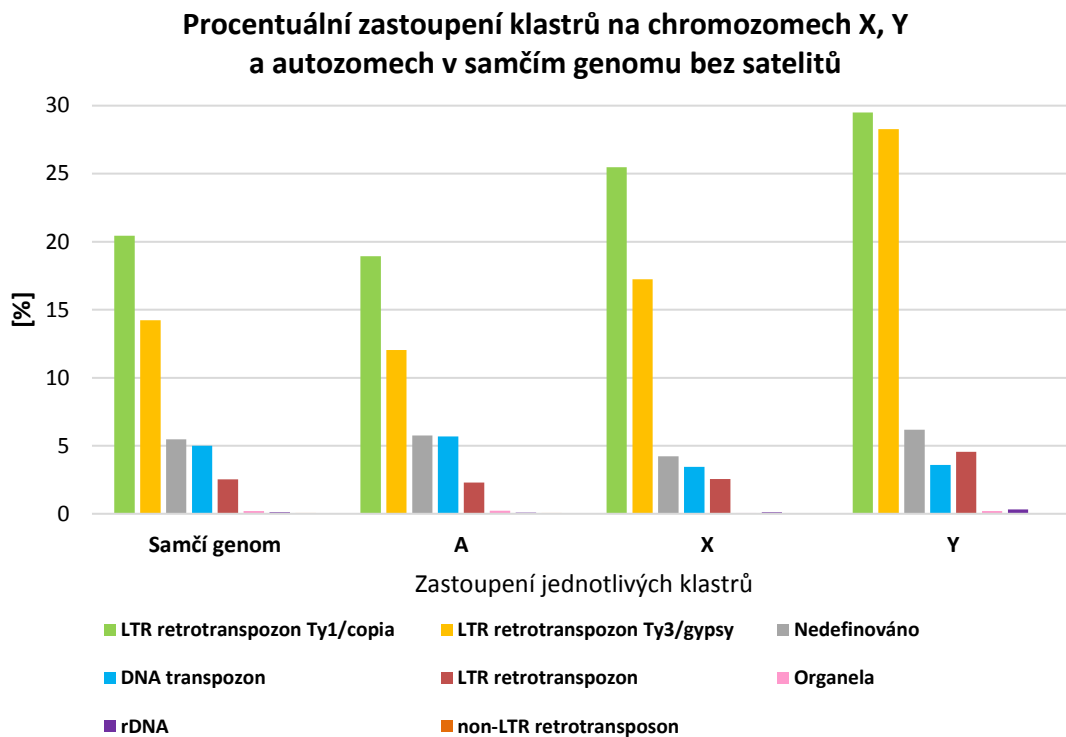
Typ repeticce	Relativní zisk X	Relativní zisk Y
Ty1/copia	+56% \pm 2%	+66% \pm 2%
Ty3/gypsy	+69% \pm 2%	+153% \pm 3%
DNA transpozon	-32% \pm 1%	-33% \pm 1%
Nedefinováno	-17% \pm 1%	+6% \pm 1%
Satelit	-19% \pm 1%	+608% \pm 9%

Největšího nárůstu si můžeme všimnout na chromozomu Y. Pozorujeme zde nárůst satelitní DNA přibližně o 608 % a LTR retrotranspozonů Ty3/gypsy o cca 153 %. Největší pokles pozorujeme u DNA transpozonů na X i Y, což je pravděpodobně způsobeno nižší transpoziční aktivitou DNA transpozonů ve srovnání s LTR retrotranspozony a amplifikací satelitů. Z toho důvodu byly DNA transpozony na rostoucích chromozomech X a Y přečísleny ostatními repeticemi, takže se procentuální zastoupení DNA transpozonů na pohlavních chromozomech snížilo.

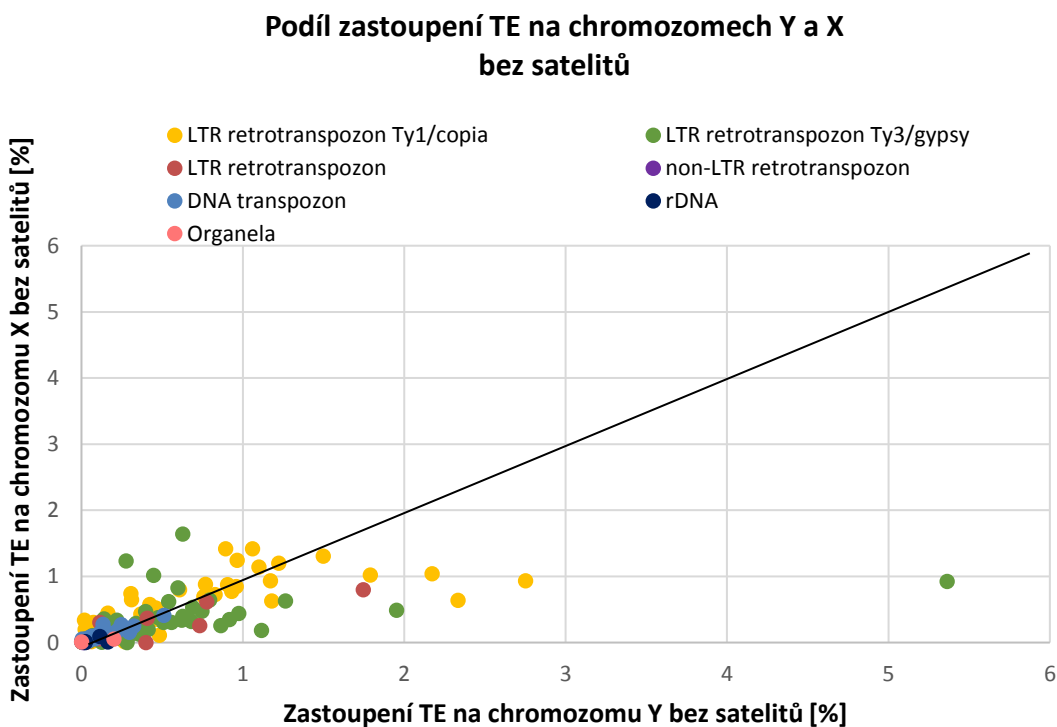
4.4. Ověřování modelu distribuce transpozonů

Pro ověření modelu proliferace transpozonů závislé na pohlaví jsme očistili výsledky od sekvencí, které se kumulují pouze v některých částech genomu. V našem případě se jedná o satelity, které se výrazně amplifikují na chromozomu Y díky nízké míře rekombinace ve srovnání s autozomy a chromozomem X. Satelity tak zkreslují rychlost fixace ostatních repeticí na chromozomu Y ve srovnání s ostatními chromozomy. Očištěná data jsme přepočítali.

Kromě LTR retrotranspozonů Ty1/copia a Ty3/gypsy se procentuální zastoupení repeticí po přepočítání téměř nezměnilo (Obr. 13). Největší procento zastoupení v samčím genomu bez satelitů opět zauímají LTR retrotranspozony, kde na chromozomu Y pozorujeme nárůst obou repeticí přibližně o 10 %. Je zde největší nárůst, protože chromozom Y obsahoval nejvíce satelitní DNA. Ve srovnání LTR retrotranspozonů Ty1/copia a LTR retrotranspozonů Ty3/gypsy jsou v genomu, na chromozomu X a autozomech více zastoupeny Ty1/copia než Ty3/gypsy. U chromozomů Y je zastoupení přibližně stejné.



Obr. 13 Procentuální zastoupení klastrů na chromozomech X, Y a autozomech v samčím genomu bez satelitů. Očištěný data set je přepočítán.

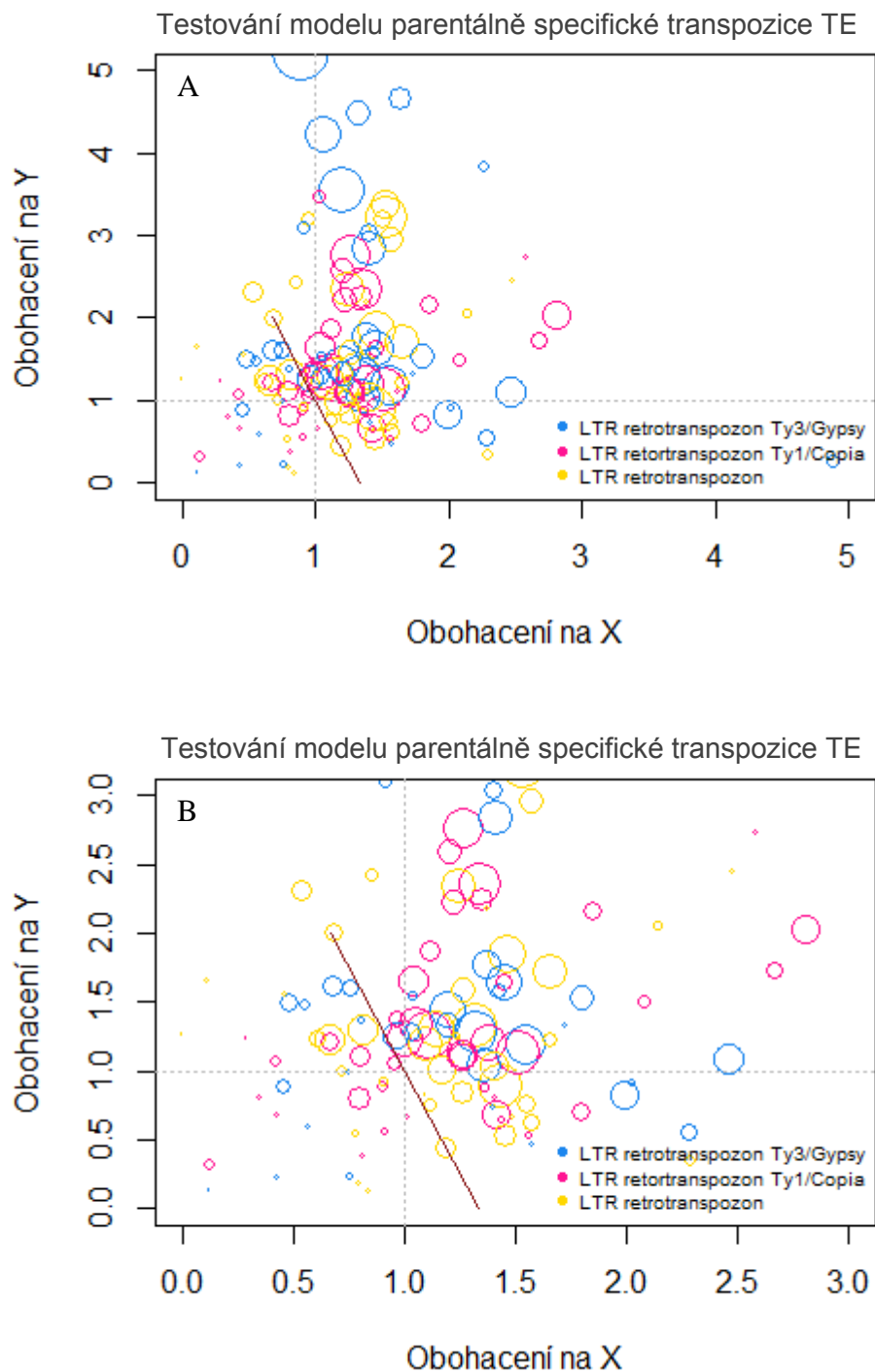


Obr. 14 Podíl zastoupení transpozónů na chromozomech Y a X bez satelitní DNA. Přímka vedená grafem znázorňuje rovnoměrné zastoupení transpozónů na X a Y.

Po očištění genomu o satelitní DNA došlo k mírné korekci distribuce klastrů v okolí přímky reprezentující shodný podíl klastrů na obou pohlavních chromozomech (Obr. 14). Část klastrů (zejména DNA transpozony) se přesunula blíže k přímce, ale náhodná akumulace jednotlivých klastrů odpovídajících LTR retrotranspozonům buď na chromozomu X nebo Y zůstává patrná.

V rámci testování modelu jsme počítali Pearsonovu lineární korelaci, která měří lineární závislost mezi dvěma proměnnými v rámci celého data setu. V případě, že by model pohlavně specifického množení byl hlavním vlivem určujícím distribuci jednotlivých rodin transpozonů na pohlavních chromosomech, očekávali bychom negativní lineární korelaci. Po prvním měření jsme dostali kladné číslo blízké nule, tedy jsme významnou korelaci nenašli. Vyzkoušeli jsme také vypočítat vážený Pearsonův korelační koeficient. Při tomto výpočtu jsme každému klastru přiřadili stupeň důležitosti v podobě váhy, která vyplývá z procentuálního zastoupení jednotlivých klastrů v samčím genomu. Váha byla získána pomocí rovnice (3.1), kde A, X a Y značí procentuální zastoupení daného clusteru v odpovídajících genomických frakcích. Do výpočtu byly zahrnuty pouze pro LTR retrotranspozony. Vážený Pearsonův korelační koeficient vyšel jako kladné číslo blízké nule, významnou korelaci jsme tedy nenašli.

Porovnání dat s modelem jsme zpracovali v podobě grafu (Obr. 15). Pro tuto analýzu byly použity pouze takové klastry, u kterých můžeme s jistotou říct, že patří mezi LTR retrotranspozony (obsahují anotované domény), pro které předpokládáme, že se mohou chovat podle našeho modelu. U DNA transpozonů nebo satelitů toto chování neočekáváme.



Obr. 15 Obohacení LTR klastrů na chromozomu Y ve vztahu k obohacení na chromozomu X. Úsečka vedená grafem znázorňuje model parentálně specifického šíření transpozonů. Její krajní body odpovídají extrémním případům čistě paternálního nebo maternálního šíření transpozonu. Plocha bodů v grafu odpovídá přiřazeným váhám (procentuálnímu zastoupení v samčím genomu). (A) a (B) jsou stejné. Graf (B) je výřezem z grafu (A) pro lepší viditelnost bodů okolo modelové úsečky.

Graf (Obr. 15) znázorňuje chování LTR retrotranspozonů. Úsečka v grafu znázorňuje model. Pro konstrukci grafu byly použity pouze LTR retrotranspozony Ty3/gypsy, Ty1/copia a ostatní LTR retrotranspozony, které jsme nedokázali přesněji určit. Každá kategorie LTR retrotranspozonů je barevně rozlišena. Každý LTR retrotranspozon má přiřazenou váhu, která vyplývá z jeho procentuálního zastoupení v genomu. Graf nezobrazuje všechny LTR retrotranspozony, protože je zde mnoho odlehlých hodnot a úsečka modelu by nebyla dobře viditelná.

Pokud graf o trochu přiblížíme (Obr. 15B) a zaměříme se na úsečku vedenou grafem, můžeme si povšimnout, že některé LTR retrotranspozony by se podle našeho modelu chovat mohly. Nemůžeme to však říct s jistotou.

5. DISKUSE

Při zkoumání složení samčího genomu a jeho komponent jsme si všimli silné kumulace satelitů na chromozomu Y. Chromozom Y je z velké části tvořen DNA s omezenou rekombinací, tzv. nerekombinující oblastí. Předpokládá se, že nerekombinující oblasti na pohlavních chromozomech skýtají podmínky pro rychlou akumulaci satelitních sekvencí a naše výsledky tuto hypotézu potvrzují. Kromě toho jsme objevili, že jeden z LTR retrotranspozonů Ty3/gypsy se v grafech zobrazuje obdobně jako satelity (Obr. 12), tedy v podobě odlehklých hodnot ukazujících na silnou akumulaci na chromozomu Y. Taková akumulace LTR retrotranspozonu může být zapříčiněna vznikem satelitu z TE, kdy se část nebo celý transpozon začne chovat jako satelit. Již dříve bylo ukázáno, že z 3' netranslatovaných oblastí retrotranspozonů mohou vznikat tandemové repetice (Vondrak et al., 2020).

Srovnávali jsme také, jaký je rozdíl repetitivní části mezi ancestrálními autozomy a dnešními pohlavní chromozomy, kterého se z těchto autozomů vyvinuly. Největší procentuální nárůst jsme pozorovali u satelitů na chromozomu Y, protože obsahuje oblasti, ve kterých nedochází k rekombinaci. To vede k akumulaci tandemových repeticí. Nejvíce akumulovanými repeticemi na pohlavních chromozomech obecně byly satelity následované LTR retrotranspozony Ty3/gypsy a Ty1/copia. Naopak procentuální zastoupení DNA transpozonů se na pohlavních chromozomech snížilo. To s největší pravděpodobností souvisí s mechanismem transpozice DNA transpozonů. Tyto „vystřihni a vlož“ mobilní elementy se v genomu šíří pomalejším tempem než „zkopíruj a vlož“ retrotranspozony. Přestože počet kopií DNA transpozonů roste nebo stagnuje, jejich podíl na zvětšujících se pohlavních chromozomech se snižuje vlivem rychlé expanze retrotranspozonů na těchto chromozomech.

Při zkoumání rozdílu mezi pohlavními chromozomy jsme zjistili, že ač je složení chromozomů X a Y z hlediska hlavních typů repeticí (LTR retrotranspozony, DNA transpozony, tandemové repetice) podobné, ve skutečnosti je velká část pohlavních chromozomů (přibližně 30-50 %) tvořena chromozomálně specifickými liniemi transpozonů a specifickými satelity. Z těchto zjištění plyne, že se jednotlivé linie LTR retrotranspozonů s větší pravděpodobností kumulují buď na jednom nebo na druhém pohlavním chromozomu. Mechanismus, který by za to mohl být zodpovědný, není znám,

ale předpokládá se, že zde svoji roli sehrávají epigenetické regulační mechanismy a rozdíly v regulaci jednotlivých TE (Kubat et al., 2014; Hobza et al., 2017).

Při výpočtu vážené lineární korelace obohacení pohlavních chromozomů jsme predikovali, že lineární korelace bude negativní. Výsledkem ale byla kladná lineární nezávislost, kterou následně potvrdil (Obr. 15). To, že se výsledek od predikce liší, může být způsobeno různými vlivy, které na transponovatelné elementy působí. Jedním z nich je rozdílná rekombinace na X a Y. Na X probíhá rekombinace jen v polovině případů, tedy jen u samice. Na Y rekombinace neprobíhá, protože není párový. Rekombinace umí odstraňovat TE z chromozomů vlivem nerovnoměrné rekombinace anebo nehomologní rekombinace (Vitte and Panaud, 2003). Nerovnoměrná rekombinace mezi LTR může vést ke tvorbě sólo-LTR (Shirasu et al., 2000), kdy dojde k odstranění vnitřní domény LTR retrotranspozonů (Bennetzen a Kellog, 1997). Tím se v místě vyříznutí vytvoří jediný samostatný sólo-LTR (Shirasu et al., 2000). Nehomologní (nelegitimní) rekombinace pomocí opravného mechanismu tvoří delece, čímž se odstraní mnohem více DNA než nerovnoměrnou rekombinací (Devos et al., 2002). Rekombinace může být ovlivněna i velikostí TE. Čím delší je LTR, tím je pravděpodobnější, že element podstoupí rekombinaci (Rabinowicz, 2000). Všechny tyto faktory v souhrnu mohou ovlivnit rychlost fixace nových inzercí TE na jednotlivých chromozomech. Nejrychlejší fixace nových inzercí TE se předpokládá na chromozomu Y, poté na X a nejpomaleji se TE fixují na autozomech.

Další vliv může mít i stáří retrotranspozonů, kdy amplifikace retroelementu je časově omezená a mnoho z rodin TE, které jsme popsali, již pravděpodobně není aktivní (Rabinowicz, 2000). Staré rodiny mohou být poničeny inzerční aktivitou novějších rodin. To se projevuje ve velké míře na rostoucím chromozomu X, ale nejvíce na Y. Vlivem poničení se tedy může zdát, že je na autozomech více LTR retrotranspozonů, než je tomu ve skutečnosti. Další možností může být sekvenční nebo strukturní preference TE při vkládání do chromozomu.

Model pohlavně specifického šíření retrotranspozonů, který jsme testovali, může pravděpodobně vycházet jen pro mladé (transpozičně aktivní) rodiny LTR retrotranspozonů. Pro potvrzení modelu je tak potřeba provést další selekci klastrů, aby analýza zahrnovala jen transpozičně aktivní rodiny TE. Dále je třeba kvantifikovat každou linii TE s co největší přesností, tedy zabránit sdružení více podobných linií TE, čemuž klastrová analýza nedokáže plně zabránit. Toho se dá dosáhnout například

využitím naší klastrové analýzy k anotaci dlouhých genomických čtení, jako jsou například sekvence BAC (Bacterial Artificial Chromosome) nebo čtení ze sekvenační metody Oxford Nanopore. Anotované genomické kopie TE pak mohou posloužit k přesné kvantifikaci jednotlivých linií TE pomocí mapování čtení z chromozomů tříděných průtokovým cytometrem, jež jsme použili v této práci.

6. ZÁVĚR

V této práci jsme analyzovali složení genomu *R. acetosa* pomocí RepeatExploreru. Využili jsme sortované chromozomy, což nám umožnilo u tohoto druhu poprvé detailně popsat složení autozomů, chromozomů X a Y.

Zjistili jsme, že se složení pohlavních chromozomů liší od složení autozomů. Zejména chromozom Y se odlišuje akumulací satelitů, která je patrně způsobena omezením rekombinace na Y. Dále jsme zjistili, že chromozomy X i Y jsou z mnoha desítek procent tvořeny unikátními rodinami mobilních elementů a repeticí. To znamená, že se některé repetice kumulují více na Y než na X a jiné se chovají naopak.

Tento fenomén může souviset s odlišnou regulací TE v samčí a samičí linii. Pokusili jsme se testovat model, který popisuje chování TE, pokud platí tento předpoklad. Bohužel se nám nepodařilo model statisticky průkazně potvrdit, nemůžeme ale vyloučit jeho platnost. Za tím účelem je potřeba provést další hlubší analýzy TE, selekci aktivních rodin TE a pokusit se odhadnout efekt dalších chromozomálně specifických fenoménů, které ovlivňují fixaci nových inzercí TE na chromozomech.

7. LITERATURA

- Andrews, S., (2010): FastQC - A quality control tool for high throughput sequence data. *Babraham Bioinformatics*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Arensburger, P., Hice, R.H., Zhou, L., Smith, R.C., Tom, A.C., Wright, J.A., Knapp, J., O'Brochta, D.A., Craig, N.L., Atkinson, P.W., (2011): Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* **188**, 45-57.
- Arkhipova, I.R., Pyatkov, K.I., Meselson, M., Evgen'ev, M.B., (2003): Retroelements containing introns in diverse invertebrate taxa. *Nature Genetics* **33**, 123–124.
- Atkinson, P.W., Warren, W.D., O'Brochta, D.A., (1993): The hobo transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the Ac element of maize. *Proceedings of the National Academy of Sciences* **90**, 9693-9697.
- Bachtrog, D., (2013): Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* **14**, 113–124.
- Bannister, A.J., Kouzarides, T., (2011): Regulation of chromatin by histone modifications. *Cell Research* **21**, 381–395.
- Baulcombe, D.C., (1996): RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. In: *Post-Transcriptional Control of Gene Expression in Plants*, (Filipowicz W., Hohn T., eds), Springer, Dordrecht, 79-88.
- Becker, H.A., Saedler, H., Lönnig, W.E., (2001): Transposable Elements in Plants. In: *Encyclopedia of Genetics*, (Brenner S., Jefferey H., Miller S., eds), Academic Press, USA, 2020–2033.
- Bennetzen, J.L., (1996): The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology* **4**, 347-353.
- Bennetzen, J.L., Kellogg, E.A., (1997): Do plants have a one-way ticket to genomic obesity? *The Plant Cell* **9**, 1509-1514.
- Bennetzen, J.L., Ma, J., Devos, K.M., (2005): Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95**, 127–132.
- Bird, A.P., (1986): CpG-Rich islands and the function of DNA methylation. *Nature* **321**, 209-213.
- Bird, A.P., (2002): DNA methylation patterns and epigenetic memory. *Genes & Development* **16**, 6-21.
- Boeke, J.D., Garfinkel, D.J., Styles, C.A., Fink, G.R., (1985): Ty elements transpose through an RNA intermediate. *Cell* **40**, 491-500.
- Boeke, J.D., Corces, V.G., (1989): Transcription and Reverse Transcription of Retrotransposons. *Annual Review of Microbiology* **43**, 403-434.
- Boeke, J.D., (1997): Lines and *Alus* — the polyA connection. *Nature Genetics* **16**, 6–7.
- Bolger, A.M., Lohse, M., Usadel, B., (2014): Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
- Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L., Feschotte, C., (2018): Ten things you should know about transposable elements. *BMC Genome Biology*, **19**:199. <https://doi.org/10.1186/s13059-018-1577-z>.
- Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M., (1987): Correct integration of retroviral DNA in vitro. *Cell* **49**, 347-356.

- Cegan, R., Vyskot, B., Kejnovsky, E., Kubat, Z., Blavet, H., Šafář, J., Doležel, J., Blavet, N., Hobza, R., (2012): Genomic diversity in two related plant species with and without sex chromosomes - *Silene latifolia* and *S. vulgaris*. *PLoS One*, **7**:2. <https://doi.org/10.1371/journal.pone.0031898>.
- Cermak, T., Kubat, Z., Hobza, R., Koblizkova, A., Widmer, A., Macas, J., Vyskot, B., Kejnovsky, E., (2008.): Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. *Chromosome Research* **16**, 961–976.
- Chalopin, D., Naville, M., Plard, F., Galiana, D., Volff, J.N., (2015): Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution* **7**, 567–580.
- Chuong, E.B., Elde, N.C., Feschotte, C., (2017): Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetic* **18**, 71–86.
- Clark, M.S., Ainsworth, C.C., Parker, J.S., (1993): Repeated dna and heterochromatin structure in *Rumex acetosa*. *Heredity* **70**, 527–536.
- Coen, E.S., Robbins, T.P., Almeida, J., Hudson, A., Carpenter, R., (1989): Consequences and mechanisms of transposition in *Antirrhinum majus*. In: *Mobile DNA*, American Society for Microbiology, Washington DC, USA, 1771 stran.
- Cost, G.J., Boeke, J.D., (1998): Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* **37**, 18081-18093.
- Craig, N.L., (2015): *Mobile DNA III*. 3rd ed., ASM Press, Washington DC, USA, 1350 stran.
- de Carvalho, F., Gheysen, G., Kushnir, S., Van Montagu, M., Inzé, D., Castresana, C., (1992): Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *The EMBO Journal* **11**, 2595-2602.
- Devos, K., Brown, J., Bennetzen, J., (2002): Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* **12**, 1075-1079.
- Drouin, G., Dover, G.A., (1987): A plant processed pseudogene. *Nature* **328**, 557-558.
- Du, C., Fefelova, N., Caronna, J., He, L., Dooner, H.K., (2009): The polychromatic Helitron landscape of the maize genome. *Proceedings of the National Academy of Science* **106**, 19916-19921.
- Eickbush, T.H., Jamburuthugoda, V.K., (2008): The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research* **134**, 221-234.
- Evgen'ev, M.B., Arkhipova, I.R., (2005): Penelope-like elements - A new class of retroelements: Distribution, function and possible evolutionary significance. *Cytogenetic and Genome Research* **110**, 510–521.
- Feschotte, C., Wessler, S.R., (2001): Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes. *Proceedings of the National Academy of Science* **98**, 8923-8924.
- Filatov, D.A., Howell, E.C., Groutides, C., Armstrong, S.J., (2009): Recent spread of a retrotransposon in the *Silene latifolia* genome, apart from the Y chromosome. *Genetics* **181**, 811-817.
- Flavell, A.J., Dunbar, E., Anderson, R., Pearce, S.R., Hartley, R., Kumar, A., (1992): Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Research* **20**, 3639–3644.
- Gent, J.I., Ellis, N.A., Guo, L., Harkess, A.E., Yao, Y., Zhang, X., Dawe, R.K., (2013): CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. *Genome Research* **23**, 628-637.

- Gent, J.I., Madzima, T.F., Bader, R., Kent, M.R., Zhang, X., Stam, M., McGinnis, K.M., Dawea, R.K., (2014): Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-Directed DNA methylation. *The Plant Cell* **26**, 4903-4917.
- Georgiev, G.P., (1984): Mobile genetic elements in animal cells and their biological significance. *European Journal of Biochemistry* **145**, 203-220.
- Geyer, P.K., Spana, C., Corces, V.G., (1986): On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. *The EMBO Journal* **5**, 2657-2662.
- Gifford, R.J., Blomberg, J., Coffin, J.M., Fan, H., Heidmann, T., Mayer, J., Stoye, J., Tristem, M., Johnson, W.E., (2018): Nomenclature for endogenous retrovirus (ERV) loci. *BMC Retrovirology*, **15**:59. <https://doi.org/10.1186/s12977-018-0442-1>.
- Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Ananda, G., Baker, D., Blankenberg, D., Chakrabarty, R., Coraor, N., Goecks, J., Von Kuster, G., Lazarus, R., Li, K., Taylor, J., Vincent, K., (2010): Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *BMC Genome Biology*, **11**:R86. <https://doi.org/10.1186/gb-2010-11-8-r86>.
- Goodwin, T.J.D., Poulter, R.T.M., (2004): A New Group of Tyrosine Recombinase-Encoding Retrotransposons. *Molecular Biology and Evolution* **21**, 746–759.
- Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Doring, A., Kapitonov, V., Diem, T., Dalda, A., Jurka, J., Pritham, E.J., Dyda, F., Izsvak, Z., Ivics, Z., (2016): A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nature Communication*, **7**:10716 <https://doi.org/10.1038/ncomms10716>.
- Greenblatt, I.M., Brink, R.A., (1963): Transpositions of modulator in maize into divided and undivided chromosome segments. *Nature* **197**, 412–413.
- Haapa-Paananen, S., Wahlberg, N., Savilahti, H., (2014): Phylogenetic analysis of Maverick/Polinton giant transposons across organisms. *Molecular Phylogenetic and Evolution* **78**, 271-274.
- Harkess, A., Mercati, F., Abbate, L., McKain, M., Pires, J.C., Sala, T., Sunseri, F., Falavigna, A., Leebens-Mack, J.J., (2016): Retrotransposon proliferation coincident with the evolution of dioecy in asparagus. *G3: Genes, Genomes, Genetics* **6**, 2679-2685.
- Havecker, E.R., Gao, X., Voytas, D.F., (2005): The sireviruses, a plant-specific lineage of the Ty1/copia retrotransposons, interact with a family of proteins related to dynein light chain 8. *Plant Physiology* **139**, 857-868.
- Hobza, R., Kubat, Z., Cegan, R., Jesionek, W., Vyskot, B., Kejnovsky, E., (2015): Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Research* **23**, 561-570.
- Hobza, R., Cegan, R., Jesionek, W., Kejnovsky, E., Vyskot, B., Kubat, Z., (2017): Impact of repetitive elements on the Y chromosome formation in plants. *Genes (Basel)*, **8**:11. <https://doi.org/10.3390/genes8110302>.
- Jaenisch, R., Bird, A., (2003): Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics* **33**, 245-254.
- Kapitonov, V. V., Jurka, J., (2001): Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Science* **98**, 8714-8719.
- Kapitonov, V. V., Jurka, J., (2006): Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Science* **103**, 4540-4545.
- Kaul, S., Koo, H.L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L.J., et al. (2000): Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Kazazian, H.H., Moran, J. V., (1998): The impact of L1 retrotransposons on the human genome.

- Nature Genetics* **19**, 19–24.
- Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., Vyskot, B., (2009): The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* **102**, 533–541.
- Kidwell, M.G., Lisch, D.R., (2001): Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**, 1-24.
- Klose, R.J., Bird, A.P., (2006): Genomic DNA methylation: The mark and its mediators. *Trends in Biochemical Sciences* **31**, 89-97.
- Kojima, K.K., Jurka, J., (2011): Crypton transposons: Identification of new diverse families and ancient domestication events. *Mobile DNA*, **2**:12. <https://doi.org/10.1186/1759-8753-2-12>.
- Kornberg, R.D., Lorch, Y., (1999): Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285-294.
- Kralova, T., Cegan, R., Kubat, Z., Vrana, J., Vyskot, B., Vogel, I., Kejnovsky, E., Hobza, R., (2014): Identification of a novel retrotransposon with sex chromosome-specific distribution in *Silene latifolia*. *Cytogenetic and Genome Research* **143**, 87-95.
- Krupovic, M., Koonin, E. V., (2015): Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Reviews Microbiology* **13**, 105–115.
- Krupovic, M., Koonin, E. V., (2016): Self-synthesizing transposons: Unexpected key players in the evolution of viruses and defense systems. *Current Opinion in Microbiology* **31**, 25-33.
- Kubat, Z., Zluvova, J., Vogel, I., Kovacova, V., Cermak, T., Cegan, R., Hobza, R., Vyskot, B., Kejnovsky, E., (2014): Possible mechanisms responsible for absence of a retrotransposon family on a plant Y chromosome. *New Phytologist* **202**, 662-678.
- Kumar, A., (1996): The adventures of the Ty1-copia group of retrotransposons in plants. *Trends in Genetics* **12**, 41-43.
- Kumar, A., (1998): The evolution of plant retroviruses: Moving to green pastures. *Trends in Plant Science* **3**, 371-374.
- Kumar, A., Bennetzen, J.L., (1999): Plant Retrotransposons. *Annual Reviews of Genetics* **33**, 479-532.
- Law, J.A., Jacobsen, S.E., (2010): Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**, 204–220.
- Lengerova, M., Vyskot, B., (2001): Sex chromatin and nucleolar analyses in *Rumex acetosa* L. *Protoplasma* **217**, 147–153.
- Li, Q., Gent, J.I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C.D., Hirsch, C.N., Dawe, R.K., Madzima, T.F., McGinnis, K.M., Lisch, D., Schmitz, R.J., Vaughn, M.W., Springer, N.M., (2015): RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proceedings of the National Academy of Science* **112**, 14728-14733.
- Lisch, D., (2013): How important are transposons for plant evolution? *Nature Reviews Genetics* **14**, 49-61.
- Loguercio, L.L., Wilkins, T.A., (1998): Structural analysis of a hmg-coA-reductase pseudogene: Insights into evolutionary processes affecting the hmgr gene family in allotetraploid cotton (*Gossypium hirsutum* L.). *Current Genetics* **34**, 241–249.
- 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**, 595-605.
- Lynch, M., (2007): The origins of genome architecture. *Journal of Heredity* **98**, 633–634.

- Makałowski, W., Gotea, V., Pande, A., Makałowska, I., (2019): Transposable elements: Classification, identification, and their use as a tool for comparative genomics. In: *Methods in Molecular Biology* vol 1910, (Anisimova M., ed.), Humana, New York, 177-207.
- Manzano, S., Megías, Z., Martínez, C., García, A., Aguado, E., Chileh, T., López-Alonso, D., García-Maroto, F., Kejnovský, E., Šíroky, J., Kubát, Z., Králová, T., Vyskot, B., Jamilena, M., (2017): Overexpression of a flower-specific aerolysin-like protein from the dioecious plant *Rumex acetosa* alters flower development and induces male sterility in transgenic tobacco. *The Plant Journal* **89**, 58-72.
- Margueron, R., Trojer, P., Reinberg, D., (2005): The key to development: Interpreting the histone code? *Current Opinion in Genetics & Development* **15**, 163-176.
- Martin, C., Zhang, Y., (2005): The diverse functions of histone lysine methylation. *Natural Reviews Molecular Cell Biology* **6**, 838–849.
- McClintock, B., (1950): The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Science* **36**, 344-355.
- McClintock, B., (1956): Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology* **21**, 197-216.
- Ming, R., Bendahmane, A., Renner, S.S., (2011): Sex Chromosomes in Land Plants. *Annual Reviews of Plant Biology* **62**, 485-514.
- Mita, P., Boeke, J.D., (2016): How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development* **37**, 90-100.
- Moore, J.K., Haber, J.E., (1996): Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383**, 644–646.
- Mosiółek, M., Pasierbek, P., Malarz, J., Moś, M., Joachimiak, A.J., (2005): *Rumex acetosa* Y chromosomes: Constitutive or facultative heterochromatin? *Folia Histochemica et Cytobiologica* **43**, 161-167.
- Muller, H.J., (1914): A gene for the fourth chromosome of *Drosophila*. *Journal of Experimental Zoology* **17**, 325-336.
- Nishimura, T., Molinard, G., Petty, T.J., Broger, L., Gabus, C., Halazonetis, T.D., Thore, S., Paszkowski, J., (2012): Structural basis of transcriptional gene silencing mediated by *Arabidopsis* MOM1. *PLoS Genetics*, **8**:2. <https://doi.org/10.1371/journal.pgen.1002484>.
- Novák, P., Neumann, P., Macas, J., (2010): Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, **11**:378. <https://doi.org/10.1186/1471-2105-11-378>.
- Novák, P., Neumann, P., Pech, J., Steinhaisl, J., MacAs, J., (2013): RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792–793.
- Ohno, S., (1967): Sex Chromosomes and Sex-Linked Genes, *Springer*, Berlin, 192 stran.
- Orgel, L.E., Crick, F.H.C., (1980): Selfish DNA: The ultimate parasite. *Nature* **284**, 604–607.
- Pardue, M.L., Danilevskaya, O.N., Traverse, K.L., Lowenhaupt, K., (1997): Evolutionary links between telomeres and transposable elements. *Genetica* **100**, 73-84.
- Peifer, M., Bender, W., (1988): Sequences of the gypsy transposon of *Drosophila* necessary for its effects on adjacent genes. *Proceedings of the National Academy of Science* **85**, 9650-9654.
- Piednoël, M., Gonçalves, I.R., Higuët, D., Bonnivard, E., (2011): Eukaryote DIRS1-like retrotransposons: An overview. *BMC Genomics*, **12**:621. <https://doi.org/10.1186/1471-2164-12-621>.
- Pray, L., (2008): Transposons: The jumping genes. *Nature Education* **1**, 204.

- Puterova, J., Kubat, Z., Kejnovsky, E., Jesionek, W., Cizkova, J., Vyskot, B., Hobza, R., (2018): The slowdown of Y chromosome expansion in dioecious *Silene latifolia* due to DNA loss and male-specific silencing of retrotransposons. *BMC Genomics*, **19**:153. <https://doi.org/10.1186/s12864-018-4547-7>.
- Pyatkov, K.I., Arkhipova, I.R., Malkova, N. V., Finnegan, D.J., Evgen'ev, M.B., (2004): Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. *Proceedings of the National Academy of Science* **101**, 14719-14724.
- Rabinowicz, P. D., (2000): Are obese plant genomes on a diet? *Genome Research* **10**, 893-894.
- Rejón, C.R., Jamilena, M., Ramos, M.G., Parker, J.S., Rejoin, M.R., (1994): Cytogenetic and molecular analysis of the multiple sex chromosome system of *Rumex acetosa*. *Heredity* **72**, 209–215.
- Renner, S.S., (2014): The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* **101**, 1588-1596.
- Robertson, H.M., (1995): The Tc1-mariner superfamily of transposons in animals. *Journal of Insect Physiology* **41**, 99–105.
- Robertson, H.M., Lampe, D.J., (1995): Distribution of Transposable Elements in Arthropods. *Annual Review of Entomology* **40**, 333-357.
- Rubin, E., Lithwick, G., Levy, A.A., (2001): Structure and evolution of the hAT transposon superfamily. *Genetics* **158**, 949-957.
- Rubin, G.M., Kidwell, M.G., Bingham, P.M., (1982): The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. *Cell* **29**, 987-994.
- Saifi, G.M., Chandra, H.S., (1999): An apparent excess of sex- and reproduction-related genes on the human X chromosome. *Proceedings of the Royal Society B: Biological Sciences* **266**, 203-209.
- Sanmiguel, P., Bennetzen, J.L., (1998): Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* **82**, 37–44.
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., Bennetzen, J.L., (1996): Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765-768.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L., (1998): The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**, 43–45.
- Schmid, C.W., (1998): Does SINE evolution preclude Alu function? *Nucleic Acids Research* **26**, 4541–4550.
- Sigman, M.J., Slotkin, R.K., (2016): The first rule of plant transposable element silencing: Location, location, location. *The Plant Cell* **28**, 304-313.
- Shirasu, K., Schulman, A.H., Lahaye, T., Schulze-Lefert, P., (2000): A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Research* **10**, 908-915.
- Simon, S.A., Meyers, B.C., (2011): Small RNA-mediated epigenetic modifications in plants. *Current Opinion in Plant Biology* **14**, 148-155.
- Slotkin, R.K., Martienssen, R., (2007): Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**, 272–285.
- Sormacheva, I.D., Blinov, A.G., (2011): LTR retrotransposons in plants. *Russian Journal of Genetics: Applied Research* **1**, 540–564.

- Steflova, P., Tokan, V., Vogel, I., Lexa, M., Macas, J., Novak, P., Hobza, R., Vyskot, B., Kejnovsky, E., (2013): Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*. *Genome Biology and Evolution* **5**, 769–782.
- Tollis, M., Boissinot, S., (2012): The evolutionary dynamics of transposable elements in eukaryote genomes. In: *Repetitive DNA* vol 7, (Garrido-Ramos, M.A, ed.), Karger, Basilej, 68–91.
- Tuesta, L.M., Zhang, Y., (2014): Mechanisms of epigenetic memory and addiction. *The EMBO Journal* **33**, 1091-1103.
- Underwood, C.J., Henderson, I.R., Martienssen, R.A., (2017): Genetic and epigenetic variation of transposable elements in Arabidopsis. *Current Opinion in Plant Biology* **36**, 135–141.
- Van Blokland, R., Van der Geest, N., Mol, J.N.M., Kooter, J.M., (1994): Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. *The Plant Journal* **6**, 861-877.
- Van Der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M., Stuitje, A.R., (1990): Flavonoid genes in petunia: Addition of a limited number of gene copies may lead to a suppression of gene expression. In: *Plant Cell* vol. 2, 4rd ed., American Society of Plant Biologists, 291–299.
- Vaucheret, H., Béclin, C., Fagard, M., (2001): Post-transcriptional gene silencing in plants. *Journal of Cell Science* **114**, 3083-3091.
- Vergara, Z., Gutierrez, C., (2017): Emerging roles of chromatin in the maintenance of genome organization and function in plants. *BMC Genome Biology*, **18**:96. <https://doi.org/10.1186/s13059-017-1236-9>.
- Vitte, C., Panaud, O., (2003): Formation of Solo-LTRs Through Unequal Homologous Recombination Counterbalances Amplifications of LTR Retrotransposons in Rice *Oryza sativa* L. *Molecular Biology and Evolution* **20**, 528–540.
- Vondrak, T., Robledillo, L.A., Novák, P., Koblížková, A., Neumann, P., Macas, J., (2019): Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *The Plant Journal* **101**, 484-500.
- Vogt, P., (1997): Historical Introduction to the General Properties of Retroviruses. In: *Retroviruses*, (Coffin J., Hughes S., Varmus H., eds.), Cold Spring Harbor, New York, U.S.A., 1-26.
- Vyskot, B., Hobza, R., (2015): The genomics of plant sex chromosomes. *Plant Science* **236**, 126-135.
- West, P.T., Li, Q., Ji, L., Eichten, S.R., Song, J., Vaughn, M.W., Schmitz, R.J., Springer, N.M., (2014): Genomic distribution of H3K9me2 and DNA methylation in a maize genome. *PLoS One*, **9**:8. <https://doi.org/10.1371/journal.pone.0105267>.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., (2007): A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**, 973–982.
- Xiong, Y., Eickbush, T.H., (1990): Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO Journal* **9**, 3353-3362.
- Yang, L., Bennetzen, J.L., (2009): Distribution, diversity, evolution, and survival of Helitrons in the maize genome. *Proceedings of the National Academy of Sciences* **106**, 19922-19927.
- Yao, J.L., Dong, Y.H., Morris, B.A.M., (2001): Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proceedings of the National Academy of Sciences* **98**, 1306-1311.

- Zaratiegui, M., Irvine, D. V., Martienssen, R.A., (2007): Noncoding RNAs and Gene Silencing. *Cell* **128**, 763-776.
- Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., (2013): The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193-205.
- Zhang, Y., (2003): Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Development* **17**, 2733-2740.
- Zuker, C., Lodish, H.F., (1981): Repetitive DNA sequences cotranscribed with developmentally regulated *Dictyostelium discoideum* mRNAs. *Proceedings of the National Academy of Sciences* **78**, 5386-5390.

8. SEZNAM POUŽITÝCH ZKRATEK

A	autozom
int	integráza
kbp	kilo párů bází
LINE	dlouhé rozptýlené jaderné elementy
LTR	dlouhá koncová repetice
Mbp	milion párů bází
non-LTR	bez dlouhé koncové repetice
ORF	otevřený čtecí rámeček
PR	proteáza
PTGS	posttranskripční umlčování genů
RE	RepeatExplorer
RH	ribonukleáza H
RT	reverzní transkriptáza
SINE	krátké rozptýlené jaderné elementy
TE	transponovatelné elementy, transpozony
TGS	transkripční umlčování genů
Ty1/copia	LTR retrotranspozon Ty1/copia
Ty3/gypsy	LTR retrotranspozon Ty3/gypsy
X	chromozom X
Y	chromozom Y
YR	tyrozin rekombináza