

Uplatněná certifikovaná metodika 25/14

DETEKCE GENOVĚ SPECIFICKÝCH
MIKROSATELITNÍCH MARKERŮ
JETELE LUČNÍHO

**Jana Řepková, Jan Ištváněk, Jana Simandlová,
Jan Nedělník, Hana Jakešová**

Brno 2014



Uplatněná certifikovaná metodika

Metodika 25/14

DETEKCE GENOVĚ SPECIFICKÝCH MIKROSATELITNÍCH MARKERŮ JETELE LUČNÍHO

**Jana Řepková, Jan Ištváněk, Jana Simandlová,
Jan Nedělník, Hana Jakešová**



Dedikace:

Uplatněná certifikovaná metodika vznikla za finanční podpory projektu QI111A019 „Nové genomické postupy pro šlechtění cizosprašných plodin na zlepšení užitkových vlastností“.

Metodiku zpracovali:

Doc. RNDr. Jana Řepková, CSc.¹; Mgr. Jan Ištváněk¹; Mgr. Jana Simandlová¹; RNDr. Jan Nedělník, Ph.D.²; Ing. Hana Jakešová, CSc.³

¹ Ústav experimentální biologie, Masarykova univerzita, Brno

² Zemědělský výzkum, spol. s r.o., Troubsko

³ Ing. Hana Jakešová, šlechtění, Hladké Životice

Oponenti:

Prof. Ing. Jaroslava Ehrenbergerová, CSc.

Mendelova univerzita v Brně, Agronomická fakulta, Ústav pěstování, šlechtění rostlin a rostlinolékařství

Ing. Lydie Čechová

Ústřední kontrolní a zkušební ústav, Brno, pracoviště Hradec n. Svitavou

Obsah

1	Cíl metodiky.....	4
2	Teoretický úvod	5
3	Vlastní popis metodiky	8
3.1	Sekvenování jaderné DNA a vyhledání sekvencí mikrosatelitů	8
3.2	Postup pro validaci SSR markerů	11
3.2.1	Izolace DNA	12
3.2.2	Polymerázová řetězová reakce	12
3.2.3	Detekce fragmentů DNA	13
4	Srovnání novosti postupů	16
5	Popis uplatnění certifikované metodiky	16
6	Ekonomické aspekty	16
7	Seznam použité literatury	18
8	Seznam publikací předcházejících metodice	20
9	Přílohy	21



1 Cíl metodiky

Metodický postup a řešení se týká navržení sady mikrosatelitních markerů ve specifických genech *Trifolium pratense* (jetel luční) definovaných na základě vlastního sekvenování další generace. Zahrnuje druhově specifické SSR markery v kódujících sekvencích (exonech) genů pro desaturázy, polyfenoloxidázy a noduliny, které jsou součástí metabolických drah biosyntézy mastných kyselin, důležitých pro kvalitu bílkovin a pro fixaci dusíku, a dále zahrnuje SSR markery v genech pro rezistenci a pro výnosové charakteristiky. Metodika se také vztahuje k postupu amplifikace cílových sekvencí SSR markerů pomocí specifických sond. Testovací sada mikrosatelitních markerů je určena pro oblast šlechtění jetele lučního.



2 Teoretický úvod

Pojmem genetický marker obecně označujeme znak s jednoduchou genetickou determinací. Genetické markery můžeme rozdělit na markery morfologické, jejichž změnou dochází i ke změně fenotypu, markery biochemické, jejichž změnou vznikají rozdílné proteinové produkty, a DNA markery, které se liší na úrovni sekvence DNA. První dva typy nejsou příliš využívány pro svoje nedostatky, jako je malý počet a nízká úroveň polymorfismu, vliv prostředí na fenotypový projev (morfologické), metoda detekce (izoenzymy). DNA markery mají řadu výhod, jako je jejich velký počet v genomech, jednoduché metody detekce, projev není ovlivňován prostředím, vysoká úroveň polymorfismu, některé mají kodominantní charakter apod. Mohou to být sekvence velmi krátké, například jen několik párů bází, ale i poměrně dlouhé.

Mikrosatelity, nebo také SSR (simple sequence repeat) či STR (short tandem repeat) markery, jsou tandemové repetice sekvencí, které jsou dlouhé pouze 2 až 6 bází. Polymorfismus těchto markerů bývá způsoben různým počtem opakování této krátké sekvence, jež je možné analyzovat elektroforeticky po jejich amplifikaci metodou polymerázové řetězové reakce (PCR). Pro reakci se navrhuje primery ohraničující SSR marker. Vývoj primerů je klíčovým krokem, pro jejich návrh je nutná předchozí charakterizace sekvencí obsahujících SSR. SSR markery se vyznačují velkým počtem alel, bývají obvykle kodominantní, což znamená, že je možné identifikovat obě alely daného organismu. Výhodou těchto markerů je jejich rovnoměrné rozložení po celém genomu a také jejich vysoký počet. Vyskytují se nejčastěji v nekódujících oblastech genomu. Odlišením jednotlivých alel markeru lze prozkoumat odlišné populace s cílem popsat např. ekonomicky zajímavé vlastnosti rostlinného materiálu a jejich genetickou podstatu. Jejich pomocí lze identifikovat a geneticky zmapovat major- i minorgeny, které determinují danou vlastnost. Mikrosatelitní markery však lze použít i pro následné klonování genu pro danou vlastnost, využít je pro popis variability mezi odrůdami, či je využít v rámci šlechtění pro úvodní selekci výchozího rostlinné-



ho materiálu. Variabilita mezi jedinci je pak zjišťována pomocí PCR. Pro získávání SSR markerů v kódujících sekvencích se často využívají známé sekvence EST (sequence-tagged site). U modelových druhů čeledě Fabaceae bylo s jejich pomocí doposud získáno 32 126 SSR u *Medicago truncatula*, 12 298 SSR u *Lotus japonicus* a 37 308 SSR u kulturního druhu *Glycine max* (<http://www.intranet.icrisat.org/gt1/SSR/SSRdatabase.html>). Genomy jetelů (*Trifolium*) jsou kvůli cizosprašnosti a silné gametofytické inkompatibilitě vysoce variabilní a polymorfní, což má za následek vysokou míru heterozygotnosti. Využívání SSR markerů i z blízkce příbuzných druhů, ale i odlišných genotypů, je málo efektivní. SSR markery byly u jetele lučního využity především ke konstrukci vysycených genetických map na základě rekombinačního mapování (Sato et al. 2005; Herrmann et al. 2006; Isobe et al. 2009).

V databázi markerů spravované Výzkumným institutem Kazusa (Kazusa DNA Research Institute; <http://marker.kazusa.or.jp>) je v současné době dostupných 7 785 markerů pro jetel luční, z nichž 7 262 jsou markery SSR, 228 markery RFLP (restriction fragment length polymorphism) a 209 SSR markerů je odvozených z kolekce EST jetele plazivého. Pro jetel plazivý je zde dostupných 1 993 markerů. V databázi EST GenBank (<http://www.ncbi.nlm.nih.gov/dbEST>) je v současné době vloženo 53 731 EST jetele, přičemž 38 353 pochází z jetele lučního, 15 323 z jetele plazivého, 53 z *T. purpureum* a 2 z *T. subterraneum*. Při využití SSR markerů je problém s jejich nízkou přenositelností mezi druhy, ale i genotypy, a to především u cizosprašných druhů, jakým je i *T. pratense*.

Před rozvojem metod sekvenování další generace (NGS, next-generation sequencing) byl vývoj SSR markerů velmi pracný a nákladný. Zahrnoval konstrukci genomových knihoven, izolaci a sekvenování klonů obsahujících SSR. Metody NGS představují efektivnější způsob pro izolaci SSR, jsou rychlejší, výkonnější, levnější a mají větší výnosnost izolovaných SSR. Pro izolaci SSR u rostlin se používají metody 454 a Illumina. Výhodou shotgun sekvenování pomocí technologie 454 jsou relativně dlouhé fragmenty, delší než u metody Illumina. Metoda Illumina vychází levněji a produkuje větší množství dat, která zcela po-



kryjí genom. Hlavní výhodou NGS oproti klasickému Sangerovu sekvenování je nižší cena a větší rychlost a výkon dosažené díky paralelnímu zpracování obrovského množství vzorků. Proto se těmto metodám také říká metody masivně paralelního sekvenování („massively parallel sequencing“, Quail et al. 2012). Nedostatkem je produkce krátkých fragmentů, kterých je ohromné množství. Následné zpracování těchto fragmentů vyžaduje speciální software a vybavení pro analýzu a také uchovávání dat. Obě metody mají oproti klasickému Sangerovu sekvenování vyšší chybovost, která je redukována zvýšením pokrytí.

Pro vyhledání SSR bylo použito sekvenování genomu nebo transkriptomu. Hlavní výhodou markerů získaných z transkriptomu je možnost jejich spojování s užitečnými geny a fenotypy. Wang et al. (2010) použili metodu Illumina k vyhledání SSR v transkriptomu druhu *Ipomoea batatas* (batátovník jedlý) a metodu 454 využili např. Dutta et al. (2011) k vyhledání SSR v transkriptomu druhu *Cajanus cajan* (kajan indický). U *T. pratense* Yates et al. (2014) publikovali sadu SSR markerů v genech asociovaných s tolerancí k suchu. Nevýhodou známých SSR markerů *T. pratense*, které jsou lokalizovány v genetických mapách, je neznalost jejich asociace s konkrétními specifickými geny. Známé SSR markery jiných rostlinných druhů asociované s konkrétními geny zase nemají využití u *T. pratense*.



3 Vlastní popis metodiky

3.1 Sekvenování jaderné DNA a vyhledání sekvencí mikrosatelitů

V prvním kroku je nezbytné získat genomické sekvence cílových genotypů. Genomická sekvence dvou odrůd jetele lučního (*Trifolium pratense*; Start, Tatra) je získána pomocí sekvenování další generace na platformě HiSeq2000. Pro přípravu DNA jsou ze semen získány rostliny a z 30denních rostlin konkrétních genotypů vypěstovaných ve skleníku jsou odebrány listy. Z 10 g mladých listů je izolována DNA z jader (Zhang et al. 1995).

V případě odrůdy Start byly z DNA připraveny a sekvenovány dvě genomické knihovny s různou délkou fragmentů, kratšími a delšími, u odrůdy Tatra byla sekvenována jedna genomická knihovna (viz Tab. 1). V dalším postupu jsou sekvenační ready zkontrolovány pomocí FastQC a případné artefakty sekvenování (zbytky adaptérů) je nutné spolu s nízkou kvalitními bázemi odstranit. Na základě velikosti haploidního genomu jednotlivých druhů z rodu *Trifolium* bylo vypočítáno pokrytí uvedené v následující tabulce.

	Start		Tatra
Genomická knihovna	300–1200 bp	700–2000 bp	280–580 bp
Počet fragmentů	534 mil.	113,7 mil.	243,6 mil.
Pokrytí genomu	122,3×	25,2×	55,4×

Sekvenační ready jsou opraveny programem Echo v1.11 (Kao et al. 2011), který eliminuje sekvenační chyby na základě pravděpodobnosti jejich výskytu. Program Abyss v1.3.3 (Simpson et al. 2009) lze poté využít pro sestavení genomu osekvenovaných odrůd *T. pratense*.

Výsledné kontigy jsou filtrovány na duplicitu a program SSRLocator (da Maia et al. 2008) je použit pro vyhledání mikrosatelitních lokusů. Počet mikrosatelit-



ních markerů a zastoupení základních motivů bylo zhodnoceno pomocí vlastního PERL skriptu SSR_motif_length.pl (Příloha 1). Zastoupení základních motivů SSR markerů pro celý genom studovaných odrůd je následující:

T. pratense Start 64 174 SSR

mono:	16 758 (26,11%)
di:	7 843 (12,22%)
tri:	18 377 (28,64%)
tetra:	4 575 (7,13%)
penta:	9 069 (14,13%)
hexa:	3 307 (5,15%)
komplexní:	4 245 (6,62%)

T. pratense Tatra 79 579 SSR

mono:	22 401 (28,1%)
di:	9 818 (12,3%)
tri:	21 374 (26,9%)
tetra:	5 770 (7,3%)
penta:	10 534 (13,2%)
hexa:	3 799 (4,8%)
komplexní:	5 883 (7,4%)

Byly predikovány mikrosatelitní markery v kódujících oblastech odrůd Start a Tatra, přičemž tyto markery se nachází přímo v exonech jednotlivých genů. Protein kódující geny jsou předpovězeny programem Augustus (Stanke et al. 2004) a většina z nich je anotována pomocí softwaru Blast2GO v2.6.6 (Conesa et al. 2005) na základě sekvenční homologie, predikce proteinových domén a podobnostech s Gene Ontology databází (<http://www.geneontology.org>).

Počty mikrosatelitních markerů a rozdělení dle jejich zastoupení dle délky základního motivu je následující:



T. pratense Start kódující sekvence 5 540 SSR

mono:	6 (0,11%)
di:	49 (0,88%)
tri:	4 489 (81,03%)
tetra:	14 (0,25%)
penta:	67 (1,21%)
hexa:	440 (7,94%)
komplexní:	475 (8,58%)

T. pratense Tatra kódující sekvence 6 749 SSR

mono:	8 (0,12%)
di:	40 (0,59%)
tri:	5365 (79,49%)
tetra:	14 (0,21%)
penta:	83 (1,23%)
hexa:	530 (7,85%)
komplexní:	709 (10,51%)

Pro identifikaci SSR markerů ve specifických genech je využita nejčastěji používaná matrice pro vyhledávání rostlinných genů založená na modelu *Arabidopsis thaliana*. K predikci kompletních i parciálních genů jsou využity kontigy *T. pratense* delší než 200 bp. Anotace genů je provedena pomocí softwaru Blast2GO, je založena nejen na výsledcích BLASTP, ale i na vyhledávání proteinových domén pomocí InterProScan (Zdobnov a Apweiler 2001), který prochází proteinové databáze ProDom, PRINTS, Pfam, Gene3D, PANTHER, SuperFamily, SignalP, TMHMM, PIR, SMART, TIGR, PROFILE, a PROSITE. Geny následně na základě anotace získají EC number (enzymové klasifikační označení) a jsou dosazeny do KEGG map, které popisují jednotlivé biosyntetické dráhy primárních i sekundárních metabolitů.

Na základě sekvenční homologie jsou identifikovány cílové geny a v nich pomo-



cí softwaru SSRLocator (da Maia et al. 2008) SSR markery. Celkem jsou identifikovány následující geny s SSR markery (Tab. 2):

- 1) 59 nodulinových genů, které hrají esenciální roli při fixaci vzdušného dusíku; v kódujících oblastech těchto genů bylo identifikováno 5 SSR markerů,
- 2) 33 genů pro desaturázy participující na biosyntéze mastných kyselin se 3 SSR markery,
- 3) 6 genů pro polyfenoloxidázy participující na kvalitě bílkovin s 1 SSR markerem,
- 4) 445 genů rezistence s 27 SSR markery,
- 5) 25 genů pro výnosové charakteristiky s 2 SSR markery.

3.2 Postup pro validaci SSR markerů

V tabulce 2 je uveden přehled identifikovaných mikrosatelitních markerů s dalšími charakteristikami jako je základní motiv SSR markeru, počet opakování tohoto motivu, přímý a zpětný oligonukleotid, které ohraničují mikrosatelit, délka amplifikovaného produktu při teplotě nasedání oligonukleotidů 58 °C. Provedení je založeno na izolaci genomové DNA z jednotlivých rostlin běžnými metodami molekulární biologie. Izolovaná DNA je podrobena vyšetření metodou PCR pomocí párů sond (oligonukleotidů) komplementárních s DNA ohraničující SSR marker

3.2.1 Izolace DNA

Pro izolaci celkové genomové rostlinné DNA analyzovaných vzorků je možné využít komerční kity, metodu CTAB (hexadecyltrimethylammonium bromid; Rogers a Bendich 1988; Příloha 1) nebo modifikovanou metodu dle Dellaporta et al. (1983; Příloha 2). Měření koncentrace a čistoty DNA lze provést pomocí NanoDrop (M.G.P. s.r.o.).

3.2.2 Polymerázová řetězová reakce

Komponenty pro PCR:

GO-TAQ™ DNA polymeráza od firmy SIGMA®



5× GOtaq PCR Reaction Buffer do firmy SIGMA®

Deoxynucleotide Mix od firmy SIGMA®

Složení reakční směsi o finálním objemu 10 µl pro PCR SSR markerů:

Komponenta	Koncentrace	Objem (µl)
dNTP	200 µM	0,2
5× pufr		2,0
Primer P	5 pM	0,2
Primer L	5 pM	0,2
Go-Taq polymeráza	1,5 U	0,1
Sterilní H ₂ O		5,8
DNA	1–10 ng	1,5

Komponenty reakční směsi pro PCR napipetovat do mikrozkušavek, promíchat pipetou, krátce zcentrifugovat, umístit do termocykleru značky BIOMETRA.

Program SSR markery:

1.	94 °C	3 min	
	58 °C	1 min	
	72 °C	1 min	
2.	94 °C	30 sec] opakovat 30×
	58 °C	30 sec	
	72 °C	30 sec	
3.	72 °C	5 min	

3.3.3 Detekce fragmentů DNA

Detekce amplifikovaných produktů je prováděna elektrofoteticky v 3% agaróзовém gelu nebo polyakrylamidovém gelu.



Tab. 2: Testovací sada mikrosatelitních markerů v souboru znaků jetele lučního ve specifických genech se vyznačuje následující skladbou.

Lokus	Přímý oligonukleoid	Zpětný oligonukleoid	Velikost produktu (bp)	Motiv
Geny pro desaturázu				
g5504.t1.cds13	TGGGTCTTAACTTGAGGT	GATTGTCACGAAAGGTTG	189	(CTT)6
g12626.t1.cds8	CCGAGGAAGAAGAAGAGG	TTGAGCAGCCCAATAAAG	230	(TGT)4
g16356.t1.cds8	TAAAGCCACTTGCTCCAG	CCATGAATCTTTCACCCA	320	(AGAACA)3
Geny pro noduliny				
g19969.t1.cds1	CCAATACACCATCCTCCA	GGTGGTTTTTCATGGTGA	104	(ACC)4
g22526.t1.cds1	CTTGTCAGGTTTGTCTGG	GGAGCCTTACCCAACACT	348	(AGC)4
g27967.t1.cds1	AAGCAAGACATCAAAGCG	AAAAAGCTGCAAGCAATG	262	(TGA)5
g34222.t1.cds4	GCAATGAGTGCAATTTCC	CTTCCTGCAGGTCTGTTG	130	(ACC)4
g34578.t1.cds1	TTTAGTAGCCGGTGCTTG	GCTGATGCTAATGCTGCT	180	(AAT)4
Geny pro polyfenoloxidázu				
g49764.t1.cds1	TGGAGTTGGAAATACCGA	TGACACCATCCCCATATC	200	(AGA)4-(GAT)4
Geny rezistence				
g2655.t1.cds4	GCCTAAAAGATCCGAAGC	CGTCCTTGGTGAATGGTA	142	(TTC)4
g3211.t1.cds2	CCATGTACCGCAAAGAAG	ATTACCGATATCTCCCCG	195	(CAA)4
g5346.t1.cds1	TAAATGGATTGTCGACGG	GCTGTGGGAAATGGTACA	273	(CTC)4
g6033.t1.cds1	CCCTCAGTTGCTGTGAAG	CGAGGAACAATAGTCGGA	274	(TCA)4
g6171.t1.cds3	GGGTGTGTGTTGGAAAA	TGTC AACATCAGCAGCAG	152	(GAT)4
g9093.t1.cds1	GGTACCGGTGGCTTAAAT	TTCCTGGAACACCAAAAA	316	(CTG)4
g9093.t1.cds1	GGATGATCGTGCAAGC	CCGAGATTCTCTTGGTT	328	(TGC)4
g12019.t1.cds1	AGGGATTGTGGTGGAGTT	TGGAAAACCAAAGGAACA	329	(TGG)5
g13444.t1.cds1	TCCGAAATCGAAGAAATG	AGGAGAGTAATGTAAGTGCACA	270	(AGC)4
g13783.t1.cds1	TAAGTTTTCATTGCCGC	AACCCTTTTCAGTTGGG	257	(AAT)5
g13969.t1.cds3	TTTGAAAATGTTCCGCAT	CTGGAAGCATGGTATGT	223	(TCA)4
g14097.t1.cds4	AAAGGTCTTCGGAGGCTA	TTCCAGCCACAATAGGAA	271	(TCTCCA)3
g14625.t1.cds3	GAAGTCATGTTCCGCAC	CATTAGGATGCAAGTGGC	267	(AGA)5

g15189.t1.cds2	AATACCACAATGCGCCTA	AATGTTTGCCAAGTCCAA	122	(ATC)4
g17034.t1.cds5	CACCAATGGAACGTCTTC	GAGGAAAGGAAGGGAACA	234	(ATG)4
g18749.t1.cds4	AATACCACAATGCGCCTA	CACCGGAAGGATATTTT	215	(ATC)4
g21880.t1.cds2	TGAGCACCTCCAAATAA	GAAAGGGAGTTCGGAAGA	301	(TCT)4
g22400.t1.cds8	AGAATTGAGGTTGGGGAG	TCCATAGCATAGGAATCCA	308	(GAT)5
g25533.t1.cds1	GCCTAAAAGATCCGAAGC	CGTCCTTGGTGAATGGTA	142	(TTC)4
g26114.t1.cds6	CACCCAGGTTTGGTGATA	CCTGATAATCATCGGCTG	322	(AGC)5
g26468.t1.cds4	CTTCCTCTCAAAAAGGC	TTTCATCGTCCCAGAAAA	186	(TAC)4
g31403.t1.cds1	GCTTGTGTCCATGGAGTG	TAATTCCAAGCCAGTTGC	233	(CTT)5
g34996.t1.cds2	GCAAAAAGAGCCACAAGA	GCCTCTCCCATCTATTC	348	(AAG)5
g36642.t1.cds1	CCAATTTGATTTTGCCTC	GAGTGGTCTTACCCACCC	273	(CCT)5
g38112.t1.cds2	TGCACGAAGTAGCTAGGG	CTCAACCACAAAAGGTGC	119	(CTG)5
g42040.t1.cds1	GGCTTGCAAGTATCATCCA	GCACCATGTTGAAGAAGC	304	(TCT)13
g42780.t1.cds1	TGAAAATCATCCAAACAACA	CTTTGCGGATGTCTGAAC	194	(TCA)4
Geny pro výnosové charakteristiky				
g13114.t1.cds10	CGTGTGAAAATGGGAGAA	AGTGGCTGTTGAGGATCA	266	(ATG)4
g13114.t1.cds11	TCCAGAACCAGAAGGTGA	GCCACCATTTTGTGAAG	295	(TTC)14



4 Srovnání novosti postupů

Aktuálnost navrženého postupu vyplývá ze současných možností rozvoje genomiky, získání celogenomových sekvencí, identifikace genů a jejich funkcí. Sekvenování další generace je vhodnou a spolehlivou alternativou klasických a biotechnologických metod. Vývoj markerů k získaným sekvencím konkrétních genotypů na základě sekvencí získaných sekvenováním další generace, popř. sekvencí získaných z dostupných databází, zajistí spolehlivou detekci cílových sekvencí a studium jejich variability. Přenositelnost SSR markerů přejatých z literatury není tak spolehlivá.

5 Popis uplatnění certifikované metodiky

Metodika bude využita v oblasti diagnostiky genotypů jetele pro získání nových zdrojů diverzity využitelných při tvorbě nového šlechtitelského materiálu pro vyšlechtění odrůdy s požadovanými hospodářskými znaky jako je odolnost k patogenům, zlepšená fixace dusíku a vyšší obsah polyfenoloxidáz důležitých pro výživářské parametry. Diagnostické markery jsou funkční (kauzální) markery v cílových genech. Takové soubory SSR markerů jetele je možné využít v populačních studiích ve šlechtění jetele lučního při 1) charakterizaci rozsáhlých kolekcí genových bank pro vyloučení duplicit, 2) odlišení odrůd, 3) asociačních studiích a detekci markerů v genetické vazbě s hospodářsky významnými geny (MAS; marker assisted selection).

6 Ekonomické aspekty

Metodika vychází z potřeby maximálně optimalizovat šlechtitelský proces, protože tvorba nových odrůd jetele lučního je dlouhodobá záležitost. Hlavním zá-



měřem je uplatnění nových metod skríningu a diagnostiky ve šlechtění jetele lučního založených na molekulárně genetických metodách. Aplikace těchto metod bude znamenat zvýšení efektivity šlechtění, jeho výrazné urychlení a tím i ekonomický efekt.



6 Seznam použité literatury

- Conesa A., Götz S., Garcia-Gomez J.M., Terol J., Talon M., Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- da Maia L.C., Palmieri D.A., de Souza V.Q., Kopp M.M., de Carvalho F.I.F., de Oliveira A.C. 2008. SSR Locator: tool for Simple Sequence Repeat discovery integrated with primer design and PCR simulation. *Internat. J. Plant Genomics* 2008: 412696.
- Dellaporta S.L., Wood J., Hicks J.B. 1983. A plant DNA miniprep: Version II. *Plant Mol. Biol. Rep.* 1, 19–21.
- Dutta S., Kumawat G., Singh B.P., Gupta D.K., Singh S., Dogra V., Gaikwad K., Sharma T.R., Raje R.S., Bandhopadhyaya T.K. 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* 11: 17.
- Herrmann D., Boller B., Studer B., Widmer F., Kölliker R. 2006. QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). *Theor. Appl. Genet.* 112: 536–545.
- Isobe S., Kölliker R., Hisano H., Sasamoto S., Wada T., Klimenko I., Okumura K., Tabata S. 2009. Construction of a consensus linkage map for red clover *Trifolium pratense* (L.). *BMC Plant Biol.* 9: 57.
- Kao W.C., Chan A.H., Song Y.S. 2011. Echo: a reference-free short-read error correction algorithm. *Genome Res.* 21: 1181–1192.
- Mardis E.R. 2013. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* 6: 287–303.
- Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A. et al. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13: 341.
- Rogers P.P., Bendich A.J. 1988. *Plant Molecular Biology manual* A6: 1–10. Kluwer Academic Publishers, Dodrecht.
- Sato S., Isobe S., Asamizu E., Ohmido N., Kataoka R., Nakamura Y., Kaneko T. et al. 2005. Comprehensive structural analysis of the genome of red clover *Trifolium pratense* L. *DNA Res.* 12: 301–364.
- Simpson J., Wong K., Jackman S., Schein J., Jones S., Birol I. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19: 1117–1123.
- Stanke M., Steinkamp R., Waack S., Morgenstern B. 2004. Augustus: a web server for gene finding in eukaryotes. *Nucl. Acids Res.* 32: W309–W312.
- Wang J., Drayton M.C., George J., Cogan N.O., Baillie R.C., Hand M.L., Kearney G. A., Erb S., Wilkinson T., Bannan N.R., Forster J.W., Smith K.F. 2010. Identification of genetic factors influencing salt stress tolerance in white clover (*Trifolium repens* L.) by QTL analysis. *Theor. Appl. Genet.* 120: 607–619.
- Yates S.A., Swain M.T., Hegarty M.J., Chernukin I., Lowe M., Allison G.G., Rutink T., Abberton M.T., Jenkins G., Skøt L. 2014. De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genomics* 15: 453.
- Zdobnov E.M., Apweiler R. 2001. InterProScan – An integration platform for the signature recognition methods in Interpro. *Bioinformatics* 17: 847–848.
- Zhang H. B., Zhao Z., Ding X., Paterson A.H., Wing R.A. 1995. Preparation of megabase-size DNA from plant nuclei. *Plant J.* 7: 175–184.



7 Seznam publikací předcházejících metodice

- Ištvánek J., Řepková J., Nedělník J., Jakešová H.: **De novo assembly of *Trifolium pratense* genome. Biotechnology in Legume Breeding**, 23rd–25th October 2012, Šumperk, p. 35. ISBN 978-80-87360-12-5
- Ištvánek J., Jaroš M., Křenek A., Řepková J. **Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae)**. American Journal of Botany, 101(2), 327–337, 2014. DOI: 10.3732/ajb.1300340.
- Řepková J., Ištvánek J., Nedělník J., Jakešová H., Simandlová J. **Next-generation sequencing and genome characterization of red clover and its wild relative zig-zag clover**. Book of Abstracts: International Conference on Enhanced genepool utilization – Capturing wild relative and landrace diversity for crop improvement, Cambridge, United Kingdom, 16–20 June, 91–92, 2014. ISBN 978929043885-0
- Řepková J., Nedělník J., Krtková V., Schulzová V., Novotná H., Hajšlová J., Jakešová H. **Phytoestrogen content in clover (*Trifolium* spp.) and in grass stands depending on treatment and storage**. Proceedings of the European Grassland Federation, Aberystwyth, Wales, 7–11 September 2014. Grassland Science in Europe, Volume 19, 495–498, 2014. ISBN 978-0-9926940-1-2
- Dluhošová J., Řepková J. **Srovnávací analýza repetitivních sekvencí *Trifolium pratense* a *Trifolium medium* sekvenováním nové generace**. Konferenční sborník z Genetické konference GSGM 2014, 24.–26. 9., Průhonice, 42, 2014.
- Dluhošová J., Pátková L., Ištvánek J., Soldánová M., Řepková J. **Využití diverzity mikrosatelitních lokusů získaných sekvenováním další generace u jetele**. Konferenční sborník z XIV. setkání biochemiků a molekulárních biologů 2014, 11.–12. 11., Brno, 2014.



Příloha 1

PERL skript SSR_motif_length.pl

```
#!/usr/bin/perl
#AUTHOR: Jan Istvanek, Laboratory of Molecular Plant Genetics, Masaryk University

unless ($ARGV[0]) {print "This script counts length of SSR basic motif after prediction.\nUsage:
perl SSR_motif_length.pl [SSRfile.txt]\n";}
if ($ARGV[0] eq "-h" | $ARGV[0] eq "-help") {print "This script counts length of SSR basic motif
after prediction.\nUsage: perl SSR_motif_length.pl [SSRfile.txt]\n";}

$input = $ARGV[0];

open (IN, $input);
open (OUT, ">$input.motif");
open (STATS, ">$input.stats");

($one, $two, $three, $four, $five, $six, $complex, $all) == 0;

while ($line=<IN>) {
    chomp ($line);
    @data = split(/\t/, $line);
    if ($data[7] =~ /Motif/) {
        print OUT join ("t", @data) . "tSSRlength\n";
        next;
    }

    if ($data[7] =~ /\-/) {
        $SSR_length = "complex";
        $complex++;
    }
    else {
        $data[7] =~ s/[()d+]/g;
        if ($data[7] =~ /\w{6}/) {
            $SSR_length = 6;
            $six++;
        }
        elseif ($data[7] =~ /\w{5}/) {
            $SSR_length = 5;
            $five++;
        }
        elseif ($data[7] =~ /\w{4}/) {
            $SSR_length = 4;

```

```

        $four++;
    }
    elseif ($data[7] =~ /\w{3}/) {
        $SSR_length = 3;
        $three++;
    }
    elseif ($data[7] =~ /\w{2}/) {
        $SSR_length = 2;
        $two++;
    }
    elseif ($data[7] =~ /\w{1}/) {
        $SSR_length = 1;
        $one++;
    }
}

print OUT join ("t", @data) . "tSSR_length\n";

}

$all = ($one + $two + $three + $four + $five + $six + $complex);
if ($one > 0) {$rel_one = $one / $all;}
if ($two > 0) {$rel_two = $two / $all;}
if ($three > 0) {$rel_three = $three / $all;}
if ($four > 0) {$rel_four = $four / $all;}
if ($five > 0) {$rel_five = $five / $all;}
if ($six > 0) {$rel_six = $six / $all;}
if ($complex > 0) {$rel_complex = $complex / $all;}

print STATS "SSR statistics-basic motif statistics:\nMotif\tCount\tRelat.\n";
if ($one > 0) {print STATS "mono:\t$one\t$rel_one\n";} else {print STATS "mono:\t0\t0\n";}
if ($two > 0) {print STATS "di:\t$two\t$rel_two\n";} else {print STATS "mono:\t0\t0\n";}
if ($three > 0) {print STATS "tri:\t$three\t$rel_three\n";} else {print STATS "mono:\t0\t0\n";}
if ($four > 0) {print STATS "tetra:\t$four\t$rel_four\n";} else {print STATS "mono:\t0\t0\n";}
if ($five > 0) {print STATS "penta:\t$five\t$rel_five\n";} else {print STATS "mono:\t0\t0\n";}
if ($six > 0) {print STATS "hexa:\t$six\t$rel_six\n";} else {print STATS "mono:\t0\t0\n";}
if ($complex > 0) {print STATS "compl.:\t$complex\t$rel_complex\n";} else {print STATS "mono:\t0\t0\n";}

close IN;
close OUT;
close STATS;
exit;

```



Příloha 2

Izolace celkové rostlinné DNA metodou CTAB

Použité roztoky

- **2× CTAB buffer**

2 % CTAB (firma Sigma-Aldrich); 100 mM TRIS (pH 8,0); 1,4 M NaCl; 1 % PVP (Mr 40000)

- **10 % CTAB**

10 % CTAB; 0,7 M NaCl

- **CTAB precipitation buffer**

1 % CTAB; 50 mM TRIS (pH 8,0); 10 mM EDTA (pH 8,0)

- **High-salt TE buffer**

10 mM TRIS (pH 8,0); 1 mM EDTA (pH 8,0); 1 M NaCl

- **0,1× TE buffer**

1 mM TRIS (pH 8,0); 0,1 EDTA (pH 8,0)

1. 100 mg listů umístit do 1,5 ml eppendorfky. Zkumavku ponořit do 2/3 do tekutého dusíku. Zamražený materiál ihned homogenizovat.
2. Vyjmout zkumavku a nechat asi 3 min ve stojanu, aby se teplota homogenátu zvýšila těsně pod bod mrazu. Homogenát zalít 140 µl 2× CTAB zahřátého na 65 °C. Roztok promíchat pipetou.
3. Mikrozukumavku uzavřít a nechat 5 min ve vodní lázni 65 °C.
4. Přidat 210 µl chloroformu, důkladně protřepat v ruce a 30 s odstředovat při 12 000 rpm.
5. Odebrat 120 µl supernatantu, přidat 12 µl 10% CTAB solution (10:1) – předehřátý na 65 °C, jinak je moc viskózní. Protřepat v ruce, přidat 120 µl chloroformu, důkladně protřepat v ruce a 30 s odstředovat při 12 000 rpm.
6. Odebrat 100 µl supernatantu do čisté eppendorfky (odebírat pomalu, aby se neodebral chloroform), přidat 100 µl CTAB precipitation buffer (1:1) a jemně promíchat pipetou. Odstředovat 60 s při 14 000 rpm (zobáček mikrozukumavky směřuje ven, aby se vědělo, kde je pelet, nemusí být vidět). Odstranit supernatant (pozor na pelet. Je rozprostřen na stěně zkumavky).
7. Přidat 50 µl High-salt TE buffer, protřepat na vortexu, 10 min ponechat v lázni 65 °C, znovu vortexovat.
8. Přidat 100 µl 96% etanolu vychlazeného na -20 °C, promíchat pipetou, odstředit 10 min při 14 000 rpm, odstranit supernatant (pozor na pelet, je na dně a může se také vznášet v supernatantu).
9. Přidat 50 µl 80% etanolu vychlazeného na -20 °C, odstředit 5 min při 14 000 rpm. Odstranit supernatant, nechat asi 1 hod vysušit při laboratorní teplotě.
10. Rozpustit v 50 µl 0,1× TE buffer. Nechat 1 hod rozpouštět. Uchovávat při -20 °C.



Příloha 3

Izolace celkové genomické rostlinné DNA

Použité roztoky

- **Extrakční buffer**

100 mM Tris-HCl, pH 7.5, 500 mM NaCl, 50 mM EDTA, 10 mM β -merkaptoethanol, 1% SDS

- **TE buffer**

50 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0

1. den

1. Z rostliny odebrat 1g listů, listy dát do třecí misky a zalít tekutým dusíkem a pořádně nadrtit.
2. Nadrcenou hmotu přenést do nové 50 ml centrifugační zkumavky.
3. Přidat předehřátých 15 ml EB pufru, dobře protřepat.
4. Přidat předehřátý 1 ml 20% SDS, opatrně protřepat.
5. Inkubovat 20 min ve vodní lázni 65 °C.
6. Ihned přidat 5 ml octanu draselného, protřepat, nechat 20 min na ledu.
7. Centrifugovat 15 min 4tis otáček 4 °C.
8. Supernatant přelit na ledu přes sterilní gázu do připravené vychlazené centrifugační zkumavky s isopropanolem.
9. Nechat 30 min při -20 °C.
10. Centrifugovat 15min 4tis otáček 4 °C.
11. Supernatant vylít.
12. Přilít malé množství 80% ethanolu, jemně opláchnout sediment a ethanol důkladně vylít/odsát.
13. Sediment nechat oschnout v zapnuté digestoři 30 min.
14. Přidat 700 μ l TE pufru.
15. Nechat rozpouštět přes noc v lednici.

2. den

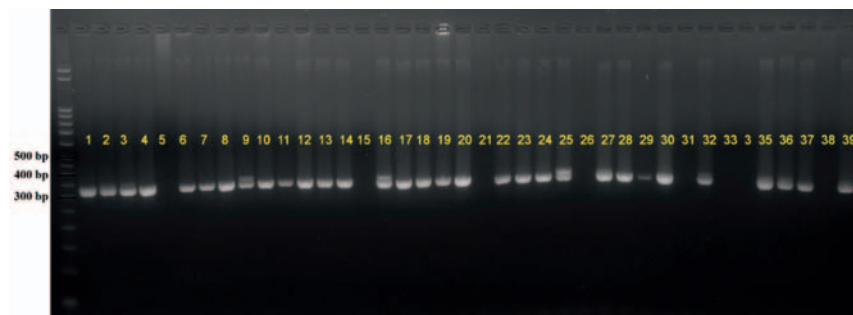
1. Zkontrolovat, jestli se sediment rozpustil, pokud ne, tak vortexovat/promíchat špičkou až do rozpouštění.
2. Celý objem přepipetovat do 2ml ependorfy.
3. Přidat 700 μ l fenolu.
4. 3 min v ruce obracet, centrifugovat 7 min 13tis. otáček 20 °C.
5. Odebrat horní fázi do nové 1,5ml ependorfy.
6. Přidat 700 μ l fenolu.
7. 3 min v ruce obracet, centrifugovat 7 min 13tis. otáček 20 °C.
8. Odebrat horní fázi do nové 1,5ml ependorfy.
9. Přidat 350 μ l fenolu a 350 μ l chloroformu.
10. 3 min v ruce obracet, centrifugovat 7 min 13tis. otáček 20 °C.



11. Odebrat horní fázi do nové 1,5 ml ependorfy.
12. Přidat 700 μ l chloroformu.
13. 3 min v ruce obracet, centrifugovat 7 min 13tis. otáček 20 °C.
14. Odebrat horní fázi do připravené vychlazené ependorfy s octanem sodným.
15. Přidat připravených vychlazených 500 μ l isopropanolu.
16. Důkladně promíchat, centrifugovat 15 min 13tis. otáček 4 °C.
17. Supernatant vylít.
18. Přilít malé množství 80% ethanolu, jemně opláchnout pelet a ethanol pořádně vylít/odsát.
19. Nechat oschnout v zapnuté digestoři 20 min.
20. Pelet rozpustit ve 100 μ l TE pufru.
21. Přidat 12 μ l RNázy, inkubovat 1 hodinu při 37 °C.
22. Přidat 10 μ l octanu amonného a připravených vychlazených 220 μ l ethanolu.
23. Dobře promíchat, nechat srážet 1 hodinu při -20 °C.
24. Centrifugovat 15 min 13tis. otáček 4 °C.
25. Supernatant vylít.
26. Přilít malé množství 80% ethanolu, jemně opláchnout pelet a ethanol důkladně odsát.
27. Nechat oschnout v zapnuté digestoři 20 min.
28. Rozpustit ve 200 μ l vody.

Příloha 4

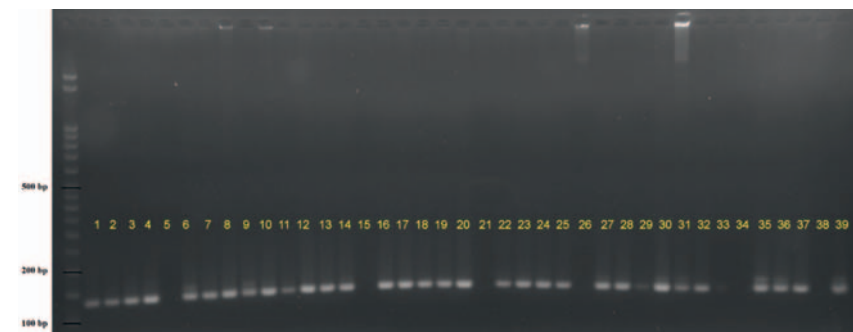
Obr. 1: Detekce fragmentů SSR markeru v lokusu g49764.t1.cds1 genu pro enzym polyfenoloxidázu u odrůd jetele lučního.



1 Agil, 2 Bonus, 3 Brisk, 4 Garant, 5 Chlumecký, 6 Nemaro, 7 Pavo, 8 Radan, 9 Respect, 10 Slavín, 11 Slavon, 12 Spurt, 13 Start, 14 Suez, 15 Tábor, 16 Trubadur, 17 Van, 18 Vendelín, 19 Vltavín, 20 Astur, 21 Atlantis, 22 Beskyd, 23 Bivoj, 24 Blizard, 25 Cyklon, 26 Dolina, 27 Dolly, 28 Fresko, 29 Kvarta, 30 Kvarta, 31 Radegast, 32 Rezista, 33 Sigord, 34 Sprint, 34 Tempus, 36 Titus, 37 Vesna, 38 Vulkán, 39 Amos

Fragmenty jsou detekovány na agarózovém gelu, kde je zřejmý polymorfismus. U odrůd Chlumecký, Tábor, Atlantis, Dolina, Radegast, Sigord, Sprint a Vulkán nebylo v rámci sekvence genu cílové místo pro amplifikaci.

Obr. 2: Detekce fragmentů SSR markeru v lokusu g34222.t1.cds4 genu pro noduliny u odrůd jetele lučního.



1 Agil, 2 Bonus, 3 Brisk, 4 Garant, 5 Chlumecký, 6 Nemaro, 7 Pavo, 8 Radan, 9 Respect, 10 Slavín, 11 Slavon, 12 Spurt, 13 Start, 14 Suez, 15 Tábor, 16 Trubadur, 17 Van, 18 Vendelín, 19 Vltavín, 20 Astur, 21 Atlantis, 22 Beskyd, 23 Bivoj, 24 Blizard, 25 Cyklon, 26 Dolina, 27 Dolly, 28 Fresko, 29 Kvarta, 30 Kvarta, 31 Radegast, 32 Rezista, 33 Sigord, 34 Sprint, 34 Tempus, 36 Titus, 37 Vesna, 38 Vulkán, 39 Amos

Fragmenty jsou detekovány na agarózovém gelu, kde je zřejmý polymorfismus. U odrůd Chlumecký, Tábor, Atlantis, Dolina, Radegast, Sigord, Sprint a Vulkán nebylo v rámci sekvence genu cílové místo pro amplifikaci.

Vydal: Zemědělský výzkum, spol. s r.o., Troubsko,

Zahradní 1, 664 41 Troubsko

Fotografie: Mgr. Jana Simandlová

Vydání: první

Náklad: 500 výtisků

Grafická úprava: Jana Adamová

Tisk: Agriprint, s.r.o., Wellnerova 7, 779 00 Olomouc

Tato publikace neprošla jazykovou úpravou

ISBN 978-80-88000-00-6