

UŽITNÝ VZOR

(11) Číslo dokumentu:

34 500

(13) Druh dokumentu: **U1**

(51) Int. Cl.:

C12Q 1/68 (2018.01)
C12Q 1/6895 (2018.01)
C12N 15/11 (2006.01)
C12N 15/29 (2006.01)

(19)
ČESKÁ
REPUBLIKA



ÚŘAD
PRŮMYSLOVÉHO
VLASTNICTVÍ

(21) Číslo přihlášky: **2020-37913**
(22) Přihlášeno: **25.08.2020**
(47) Zapsáno: **03.11.2020**

(73) Majitel:
Zemědělský výzkum, spol. s r.o., Troubsko, CZ
Masarykova univerzita, Brno, Brno-město, CZ
Ing. Hana Jakešová, CSc., Hladké Životice, CZ
ESSENCE LINE, s.r.o., Praha 5, Motol, CZ

(72) Původce:
Ing. Hana Jakešová, CSc., Hladké Životice, CZ
Ing. Oldřich Trněný, Rousínov, Slavíkovice, CZ
Ing. Petr Novotný, Praha 5, Stodůlky, CZ
Mgr. Ján Boroň, Prešov, SK
Mgr. David Vlk, Brno, Královo Pole, CZ
RNDr. Jan Nedělník, Ph.D., Brno, Trnitá, CZ
prof. RNDr. Jana Řepková, CSc., Brno, Černá Pole,
CZ

(74) Zástupce:
Ing. Libor Markes, patentový zástupce, Grohova
145/54, 602 00 Brno, Veverčí

(54) Název užitného vzoru:
**Sada markerů v genech efektivních pro
vyšší fixaci dusíku u jetele lučního**

Sada markerů v genech efektivních pro vyšší fixaci dusíku u jetele lučního

Oblast techniky

5

Technické řešení zahrnuje sadu genetických markerů v genech jetele lučního (*Trifolium pratense* L.), které ovlivňují efektivitu fixace dusíku. Sada markerů je založena na jednonukleotidových polymorfismech (SNP) v genech podílejících se na procesu nodulace a fixaci vzdušného dusíku. Sada markerů nalezne uplatnění v oblasti šlechtění jetele lučního.

10

Dosavadní stav techniky

Fixace vzdušného dusíku jakožto mutualistický vztah mezi rostlinami a hlízkovitými bakteriemi představuje pro rostliny výraznou kompetitivní výhodu, protože jim umožňuje růst i na půdách chudých na dusík. Z hlediska využití ve šlechtění rostlin je nejvíce prozkoumán symbiotický vztah mezi bakteriemi rodu *Rhizobium* a rostlinami z čeledi *Fabaceae*. Rostliny této čeledi mají široké hospodářské využití v potravinářském a farmaceutickém průmyslu a mnoho z nich jsou i důležité zemědělské plodiny. Důležitou pícninou této čeledi je i jetel luční (*T. pratense* L.), který je využíván jako krmivo pro dobytek a používá se při osevních sledech k zúrodnování půdy. Důležitým faktorem při jeho použití jako zelené hnojivo je jeho schopnost obohacovat půdu o dusík, který, je-li ho v půdě nedostatek, zvládá vytvořit s pomocí hlízkovitých bakterií. Efektivita obohacení půdy dusíkem ale závisí na samotné schopnosti rostliny fixovat vzdušný dusík. Ta se však velmi liší nejen mezi druhy, ale i mezi samotnými jedinci uvnitř druhu. Tato variabilita se vysvětluje existencí polymorfismů v genech, které zodpovídají jak za interakci mezi bakterií a rostlinou, tak za proces nodulace a samotnou fixaci dusíku. Některé polymorfismy se vyznačují jednoduchou genetickou determinací, a proto jsou vhodné k využití jako genetické markery.

V raných fázích se jako genetické markery používaly znaky založené na změně fenotypu, anebo znaky biochemického charakteru (isozymy), postupem času se s rozšířenými možnostmi analýzy DNA začaly zavádět genetické markery založené přímo na variabilitě v sekvenci DNA. Tyto markery mají výhodu, že jsou v hojné míře rozprostřeny po celém genomu, nejsou ovlivňovány prostředím, jejich variabilita je vysoká a jsou také snadno detekovatelné. V době před zavedením metody polymerázové řetězové reakce (PCR) v 80. letech 20. st., byly hojně používány polymorfismy v délce restrikčních fragmentů (RFLP). Z důvodu vyšší citlivosti a snadné automatizaci se po vývoji PCR začaly využívat markery více či méně založené na amplifikaci fragmentů, mezi nimiž k nejpoužívanějším patří jednonukleotidové polymorfismy (SNP).

Jednonukleotidový polymorfismus jakožto variabilita v rámci jediného nukleotidu je nejmenší jednotka genetické variability. K praktickým účelům se ve větší míře začala využívat až později, když se usnadnilo jejich vyhledávání v genomech organismů a jejich následná detekce. V rámci genomu je distribuce SNP nehomogenní, častěji se vyskytují v nekódujících oblastech, nebo obecně tam, kde přírodní výběr nenapomáhá fixaci určité alely v daném prostředí. Záměna jednoho nukleotidu v protein kódujících oblastech sice nemusí nutně vést ke změně fenotypu, a dokonce ani ke změně aminokyseliny (tzv. synonymní substituce), často ale tato záměna vede ke změně aminokyseliny (nesynonymní substituce) a také k odlišnému proteinovému produktu, příp. ke zkrácenému proteinovému řetězci. Výhodou SNP je jak jejich všudypřítomnost a kvantita uvnitř genomů, tak možnost masivní a automatické detekce. Vzhledem k jejich abundanci je možné sestavit podrobné genetické mapy, které se používají k mapování jak cílových regionů, tak celého genomu, rychlé asociaci genetického markeru s fenotypovým znakem anebo ke klonování genů. Vzhledem k tomu, že tyto markery jsou méně polymorfní než jiné genetické markery (např. mikrosatelity), je potřeba tento nedostatek kompenzovat využitím co nejvyššího počtu SNP. Vyhledávání jednotlivých SNP je poměrně snadné u modelových organismů a organismů s malým genomem. Ovšem u rostlin byla tato identifikace kvůli velikosti a struktuře jejich genomů, kdy

velkou část zabírají repetitivní elementy, a častému polyploidnímu charakteru, náročná a nákladná, a to až do vývoje technologií sekvenování nové generace (NGS).

Před vývojem NGS se k získávání markerů SNP využívalo resekvenování ampliconů získaných Sangerovou metodou a *in silico* vyhledávání uvnitř databází EST, následované validací pomocí techniky PCR. Vzhledem ke své povaze ale tyto metody neumožnily identifikaci SNP v nekódujících regionech a v oblastech mezi geny, ale hledaly nové SNP pouze uvnitř genů, kde je frekvence SNP poměrně nízká, obzvláště v konzervovaných regionech. Vyhledávání u rostlin navíc komplikoval fakt, že spousta genů se objevuje ve více kopiích, což při identifikaci SNP *in silico* vedlo k nalezení velkého počtu nealelických SNP, které byly spíše reliktem paralogních genů než znak genetické variability. Tyto SNP tudíž nebyly vhodné pro další využití ve šlechtění rostlin. Metody NGS se začaly objevovat na začátku 21. st. (454/Roche, Illumina) a prakticky od svého počátku představovaly průlom ve studiu genomů. Tyto techniky umožňují masivní a rychlé zjišťování sekvencí DNA bez nutnosti klonování. Díky relativně nízké ceně se sekvenování genomů stalo dostupnější pro laboratoře po celém světě a získaná data našla uplatnění při evolučních analýzách, studiu regulace genů a při výzkumu genové exprese, umožnila rovněž masivní a rychlé vyhledávání genetických markerů. Pomocí resekvenování transkriptomu bylo najednou možné identifikovat SNP i u rostlin s obrovským a komplexním genomem, jako je kukuřice, pšenice anebo cukrová třtina. Firma Roche obohatila možnosti masivní identifikace SNP pomocí technologie NimbleGen sequence capture, která je založena na vytažení cílových sekvencí z genomu pomocí navržených sond, jejich amplifikaci a následném cíleném osekvenování, což vede k vyššímu pokrytí genů zájmu.

Pro oblast šlechtění rostlin je největším přínosem sekvenování a resekvenování hospodářsky důležitých plodin v milionech nalezených SNP po celém genomu, klíčovým předpokladem je ale schopnost využít obrovské množství získaných dat k systematické charakterizaci fenotypů. V době, kdy ještě nebylo dostupných tolik genetických markerů, se často využívala selekce s podporou markerů (MAS). Během této techniky je kauzální mutace neznámá, ale je známý marker, který je s danou mutací ve vazebné nerovnováze (tzn. existuje asociace výskytu daného genetického markeru a efektu mutované alely), takže pro mapování lze místo samotného genu využít daný genetický marker. Ve šlechtění jsou ale často objektem zájmu znaky kvantitativní, jejichž genetická determinace je obtížnější než u znaků kvalitativních, kde se na daném znaku podílí jeden nebo málo genů. Genetická podstata kvantitativního znaku je totiž založena na existenci tzv. lokusů kvantitativního znaku (tzv. QTL), z nichž každý sám o sobě má na výsledné fenotypové hodnotě znaku často malý podíl, ale jejich kvantita po celém genomu může být vysoká a jejich součet nakonec dává výsledný fenotyp. Hlavním nedostatkem využití technik MAS pro šlechtění kvantitativních znaků tedy je to, že zjištěný QTL vysvětluje pouze zlomek z genetické variability u šlechtitelského cíle a zbytek zůstává neznámý.

Tento problém se v současné době překonává využitím SNP nalezených pomocí metod NGS. Obrovské množství těchto markerů umožňuje provádět celogenomové asociční studie (GWAS), jejichž podstatou je, že se shromáždí jedinci s kontrastním fenotypem, pomocí DNA čipu s jednotlivými SNP, které jsou ve vazbě s QTL, se zjistí alelové varianty na daných pozicích, a sleduje se asociace obrovského množství markerů najednou s žádaným znakem, přičemž je předpoklad, že kontrastní fenotypy se budou statisticky lišit právě v těch oblastech, které se na realizaci daného znaku podílejí. Informace ze studií GWAS jsou dále využívány i při testování pomocí GWAS asistované genomické selekce. Cílem je přiřadit každému genetickému markeru tzv. odhad účinku, a pro samotnou selekci vybrat pouze markery QTL, které se na výsledném znaku podílejí v největší míře. Šlechtitelská hodnota každého genotypu je pak vypočtena zahrnutím těch genetických markerů, které jsou statisticky významné.

Technické řešení si klade za cíl navrhnout sondy pro SNP markery kandidátních genů asociované s efektivitou fixace vzdušného dusíku. Tyto markery v genech *T. pratense* zapojených do procesu regulace a realizace nodulace a fixace dusíku budou využity při konstrukci populačního čipu, který umožní masivní a rychlé genotypování jedinců, a tak jejich odhad šlechtitelské hodnoty.

Podstata technického řešení

Tento úkol řeší testovací sada 7 nových SNP markerů uvedená v následující tabulce:

5

Číslo	Lokus	Pozice	Referenční alela	Alternativní alela	Sekvence sondy
1	Tp_9375	133	T	C	AAAATTTGGTTAAT(T/C)TTTTTATTATA AGTT
2	Tp_13036	2625	C	T	ATTTTAAATTAAG(C/T)TCTTTGAACAAT
3	Tp_7124	4424	G	A	GAGATGTTGAG(G/A)ATTTTAAGAGA
4	Tp_86	3276	C	G	GTAATTTGAT(C/G)TCATCGGTGT
5	Tp_2269	7537	A	G	TTTCCTTGAAT(A/G)CTTGTCTTAAC
6	Tp_31403	2100	C	T	AGATTAGCGC(C/T)GAGTTTG
7	Tp_8398	805	T	A	AACACTAAAGC(T/A)AAGTATCAC

Příklady uskutečnění technického řešení

10 Předmětem užitého vzoru je 7 nových SNP markerů získaných na základě vlastního cíleného sekvenování panelu 17 genů u celkem 48 různých genotypů *T. pratense*. Tyto SNP jsou lokalizovány ve specifických genech, které se podílejí na biologické fixaci dusíku a procesu nodulace.

15 Součástí užitého vzoru je soubor 14 oligonukleotidů o celkové délce 326 bází se SNP polymorfismem uprostřed sekvence. Tyto oligonukleotidy slouží jako fluorescenčně značené sondy imobilizované na SNP čipu, kde se hybridizují k DNA komplementární se sekvencí sond. Každá alelová varianta je označena jinou fluorescenční barvou (červená, zelená), což umožňuje jednoduchou detekci signálu a rychlé genotypování v daném lokusu. Sekvence oligonukleotidových sond jsou odvozeny z referenční sekvence *T. pratense* odrůdy Tatra.

20 Tato sada SNP markerů a s nimi souvisejících sond umožňuje masivní a rychlé genotypování jedinců v populaci a identifikaci rostlin s alelovými variantami, které jsou asociovány s vyšší efektivitou biologické fixace dusíku. Vlastní analýza začíná izolací a purifikací DNA rostlinného vzorku, následuje příprava na hybridizaci a fluorescenční značení. Navržené sondy se SNP markery budou součástí populačního čipu, který umožňuje rychlou identifikaci alelových variant několika SNP pro obrovské množství jedinců najednou, což zpřesňuje predikci žádaného fenotypového znaku u analyzovaných jedinců.

30 Sekvence obsahující vybrané SNP markery byly získány vlastním cíleným sekvenováním panelu 17 genů pomocí technologie NimbleGen sequence capture firmy Roche. Vybrány byly ty geny, u kterých byl experimentálně prokázán jejich vliv na proces fixace dusíku, anebo které jsou specificky exprimované v hlízkách. Jednalo se jak o transkripční faktory, leghemoglobiny, receptorové kinázy anebo cytokininové receptory. Panel genů obsahoval celkem 95 kb genomické sekvence vybraných genů, knihovny pro sekvenování byly vytvořeny pomocí metody SeqCap EZ HyperCap, kdy jednotlivé fragmenty byly sekvenovány z obou konců (paired-end) s délkou čtení 150 bp na platformě Illumina NextSeq 500. Cílené sekvenování proběhlo u celkem 48 genotypů *T. pratense*, u jedinců kontrastních v efektivitě fixace dusíku. Počet nalezených polymorfismů se nacházel v rozmezí 18 až 887 pro každý gen.

40

Základní analýza hrubých sekvenačních dat proběhla pomocí programu FastQC. Pomocí nástroje Trimmomatic pak byly odstraněny kontaminující sekvence, zbytky adaptérů a báze s nízkou kvalitou. Následné mapování opravených sekvenačních readů bylo provedeno na základě referenční sekvence *T. pratense* odrůdy Tatra s využitím algoritmu BWA-MEM z nástrojů BWA. Sekvenační data byla upravena na pokrytí 150x. Program GATK byl dále použit pro recalibraci

45

kvality namapovaných readů a nakonec pro vyvolání SNP lokusů mezi vzorky. Tyto varianty byly dále filtrovány s využitím standardních filtrovacích parametrů podle doporučení v manuálu. Na základě ploidie sekvenovaných jedinců byly nalezené varianty vyvolány v diploidních nebo tetraploidních stavech a následně byly anotovány pomocí programu Variant Effect Predictor.

5

Efektivita fixace vzdušného dusíku u jednotlivých genotypů byla hodnocena acetylen redukční metodou ARA (acetylene reduction assay), kdy výsledky z tohoto měření byly uvedeny jako hodnoty molární koncentrace etylenu (C_E). Hodnoty C_E byly standardizovány do Z-score, aby se mohly porovnávat výsledky z rozdílných měření. Asociace získaných SNP markerů s efektivitou fixace dusíku byla hodnocena pomocí lineární regrese mezi Z-score C_E a numericky vyjádřeným genotypem abundance referenčních alel analyzovaných genotypů. Vybrány byly ty SNP markery, které vykazovaly nejvyšší závislost C_E na genotypu v daném lokusu a jejichž efekt byl ověřen ve dvou po sobě jdoucích generacích rostlin.

10

15 Vybrané SNP markery jsou v těchto genech:

Označení genu <i>T. pratense</i> var. Tatra	Název genu	Funkce genu
Tp_9375	<i>ERF required for nodulation 1</i>	transkripční faktor, kontrola nodulace, regulace růstu infekčních vláken
Tp_13036	<i>PEN3-like</i>	rezistence vůči chorobám, specificky exprimovaný v hlízkách
Tp_7124	<i>SUNN</i>	kináza, kontrola počtu hlízek
Tp_86	<i>PHO2-like</i>	regulace akumulace fosforu v hlízkách
Tp_2269	<i>NFP</i>	receptor Nod faktorů
Tp_31403	<i>Ca/CaM-dependent kinase</i>	přenos signálu Nod faktoru, nodulace
Tp_8398	<i>RDNI</i>	regulace počtu hlízek

Tento soubor SNP markerů bude možné využít ve šlechtění *T. pratense* pro efektivnější fixaci dusíku, při selekcích rostlin prostřednictvím geneticky vázaných markerů a při konstrukci populačního čipu, který umožní masivní a rychlé genotypování mnoha jedinců v několika lokusech najednou.

20

NÁROK NA OCHRANU

1. Testovací sada SNP markerů v genech efektivních pro vyšší fixaci dusíku u jetele lučního,
5 **vyznačující se tím**, že je tvořena následujícími prvky:

Číslo	Lokus	Pozice	Referenční alela	Alternativní alela	Sekvence sondy
1	Tp_9375	133	T	C	AAAATTTGGTTAAT(T/C)TTTTATTATAAGTT
2	Tp_13036	2625	C	T	ATTTTAATTAAG(C/T)TCTTTGAACAAT
3	Tp_7124	4424	G	A	GAGATGTTGAG(G/A)ATTTAAGAGA
4	Tp_86	3276	C	G	GTAATTTGAT(C/G)TCATCGGTGT
5	Tp_2269	7537	A	G	TTTCCTTTGAAT(A/G)CTTGTCTTAAC
6	Tp_31403	2100	C	T	AGATTAGCGC(C/T)GAGTTG
7	Tp_8398	805	T	A	AACACTAAAGC(T/A)AAGTATCAC