

MAPOWANIE GENETYCZNE MARKERÓW STS W GENOMIE ŁUBINU WĄSKOLISTNEGO (*LUPINUS ANGUSTIFOLIUS* L.) I ICH ANALIZA FUNKCJONALNA

KATARZYNA KAMEL, MAGDALENA KROC, WOJCIECH ŚWIĘCICKI

Instytut Genetyki Roślin Polskiej Akademii Nauk w Poznaniu

kkam@igr.poznan.pl

Synopsis. Łubin wąskolistny jest jednym z trzech gatunków uprawnych łubinów. W ostatnich latach obserwuje się wzrost zainteresowania uprawą łubinu, cennego źródła białka pochodzenia krajowego. Niestety wiedza z zakresu genetyki tego gatunku, a w szczególności struktury i organizacji genomu, jest wciąż ograniczona. Celem prowadzonych badań było uzupełnienie najnowszej wersji mapy genetycznej łubinu wąskolistnego o nowe markery zdefiniowane sekwencyjnie, projektowane dla sekwencji kodujących. Opracowano 85 markerów polimorficznych, spośród których 81 wykazało sprzężenie na mapie genetycznej. Wzbogacenie mapy o nowe markery STS zwiększyło jej długość do 2611,1 cM, a średnia odległość pomiędzy markerami wynosi obecnie 3,76 cM. Kolejnym etapem była analiza funkcjonalna *in silico* 68 sekwencji nukleotydowych opracowanych markerów łubinu wąskolistnego. Terminy GO zostały przypisane dla 32 sekwencji. Produkty genów zdefiniowano pod względem funkcji molekularnej, komponentu komórkowego oraz procesów biologicznych.

Słowa kluczowe – *key words*: łubin wąskolistny – *narrow-leafed lupin*, mapowanie genetyczne – *genetic mapping*, markery zdefiniowane sekwencyjnie – *sequence-specific markers*, adnotacja funkcjonalna – *functional annotation*

WSTĘP

Łubin wąskolistny jest jednym z trzech gatunków użytkowych należących do rodzaju *Lupinus* uprawianych w Polsce. Swoją popularność zawdzięcza wysokiej zawartości białka w nasionach, stosunkowo niewielkim wymaganiom glebowym oraz asymilacji azotu atmosferycznego dzięki symbiozie z bakteriami brodawkowymi z rodzaju *Bradyrhizobium*. W ostatnich latach wzrasta zainteresowanie uprawą łubinu, cennego źródła białka pochodzenia krajowego, w związku z zamiarem ograniczenia importu śruty sojowej. Uprawa i wykorzystanie łubinu napotyka jednak na trudności związane z niestabilnością plonów oraz podatnością na choroby, w tym antraknozę i fuzariozę. Lepsze poznanie struktury i organizacji genomu łubinu wąskolistnego może ułatwić hodowlane ulepszenie gatunku i pełne wykorzystanie jego walorów uprawnych i użytkowych.

Celem przeprowadzonych prac było wzbogacenie najnowszej wersji mapy genetycznej łubinu wąskolistnego [Nelson i in. 2010] o nowe markery zdefiniowane sekwencyjne (STS – ang. Sequence Tagged Sites). Niezwykle ważnym aspektem badań struktury genomów jest możliwość powiązania poszczególnych sekwencji z informacjami na temat ich biologicznej funkcji [Conesa i in. 2005], dlatego w dalszej kolejności przeprowadzono analizę funkcjonalną *in silico* sekwencji nukleotydowych opracowanych markerów łubinu wąskolistnego. Pozwoliło to na przyporządkowanie im terminów GO (ang. Gene Ontology) oraz ich zdefiniowanie pod wzglę-

dem funkcji molekularnej, lokalizacji komórkowej oraz procesów biologicznych, w których mogą uczestniczyć.

MATERIAŁ I METODY

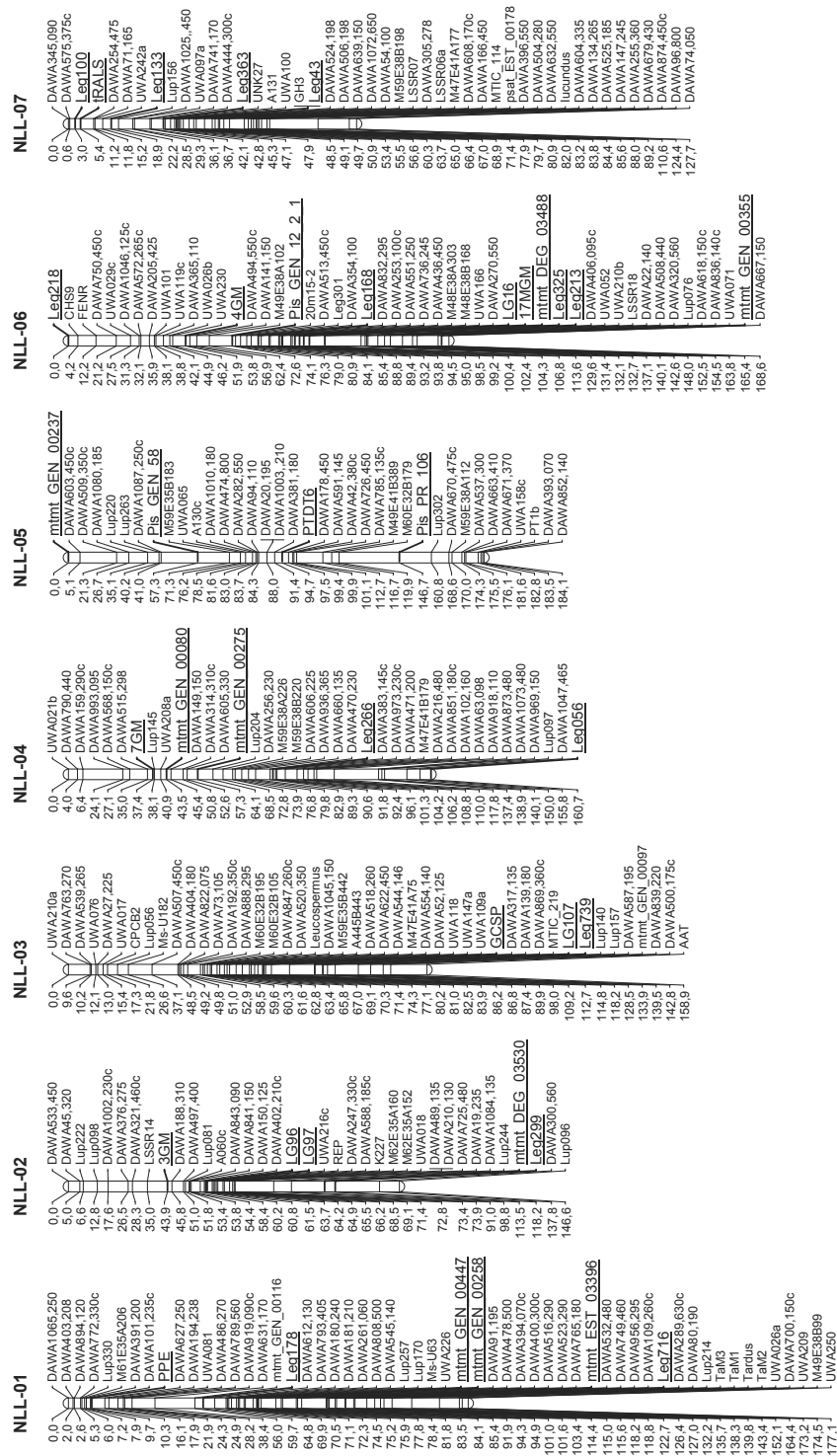
Mapowanie genetyczne markerów STS w genomie łubinu wąskolistnego przeprowadzono w latach 2005–2012 w Pracowni Analizy Genomu Instytutu Genetyki Roślin PAN. Populację mapującą stanowiło 89 linii wsobnych rekombinantów kombinacji krzyżówkowej linii hodowlanej (83A:476) i typu dzikiego (P27255) w pokoleniu F8 (współpraca z Department of Agriculture and Food, Perth, Western Australia). Wykorzystywane grupy starterów (GLIP, Leg, MLG i cross-legume), różniące się nieznacznie sposobem projektowania, zostały opracowane dla sekwencji różnych gatunków roślin motylkowatych. Strategia ich projektowania polegała na wykorzystaniu ewolucyjnie konserwatywnych sekwencji kodujących dla motylkowatych gatunków modelowych (*Lotus japonicus*, *Medicago truncatula*) oraz kilku gatunków uprawnych (*Glycine max*, *Phaseolus vulgaris*) i referencyjnego genomu *Arabidopsis thaliana*. Sekwencje wykorzystanych starterów otrzymano w ramach współpracy z kilkoma ośrodkami naukowymi w ramach 6 PR UE „Grain Legumes”.

W mapowaniu genetycznym wykorzystano dane źródłowe najnowszej wersji mapy łubinu wąskolistnego [Nelson i in. 2010], przy czym do konstrukcji mapy wykorzystano jedynie markery szkieletowe. Analizę sprzężeń nowych markerów STS wykonano za pomocą oprogramowania Map Manager v.QTXb20 [Manly i in. 2001], a wizualizację mapy z wykorzystaniem programu komputerowego MapChart [Voorrips 2002]. Do analizy funkcjonalnej zastosowano aplikację Blast2GO [Götz i in. 2011].

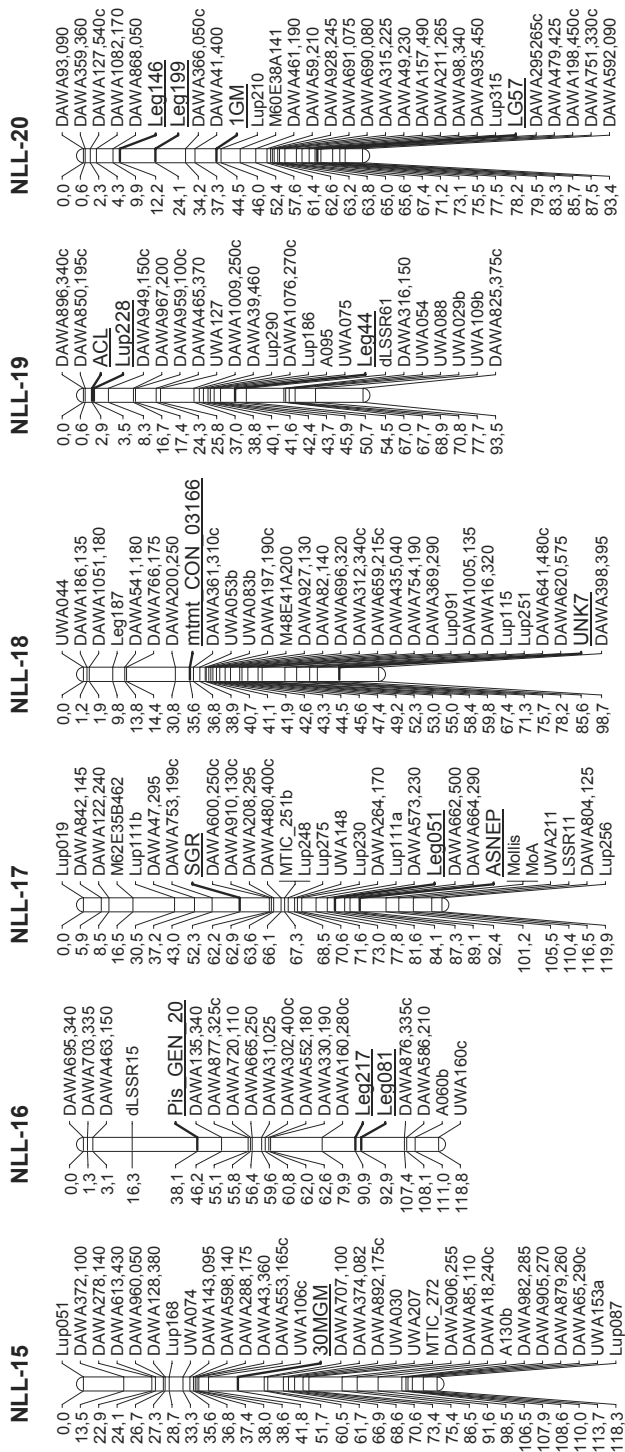
WYNIKI I DYSKUSJA

Analizie poddano łącznie 375 par starterów. Produkty reakcji PCR uzyskane z linii rodzicielskich analizowano w pierwszej kolejności pod względem polimorfizmu obserwowanego bezpośrednio w żelu agarozowym, tj. wielkości produktu amplifikowanego daną parą starterów oraz amplifikacji produktu tylko u jednego z rodziców. W przypadku braku polimorfizmu obserwowanego bezpośrednio w żelu agarozowym produkty amplifikacji obu linii rodzicielskich sekwencjonowano w poszukiwaniu polimorfizmu na poziomie pojedynczego nukleotydu – SNP. Gdy wykryta mutacja SNP zmieniała bezpośrednio miejsce rozpoznawane przez enzym restrykcyjny, opracowywano markery typu CAPS (ang. Cleaved Amplified Polymorphic Sequence) natomiast, gdy wykryty polimorfizm nukleotydu nie zmieniał bezpośrednio miejsca restrykcyjnego dla enzymu, opracowywano marker typu dCAPS (ang. derived-CAPS).

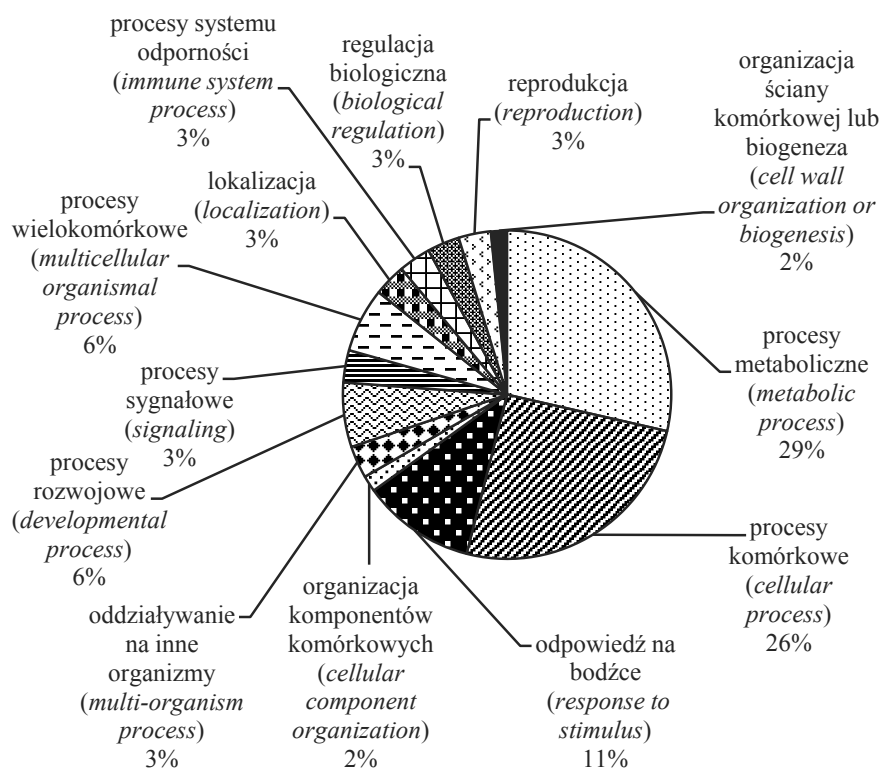
Łącznie wytypowano 85 markerów polimorficznych, spośród których 81 wykazało sprzężenia na mapie genetycznej gatunku (28 uwzględniono wcześniej w publikacji Nelson i in. 2010) (rys. 1). Polimorfizm markerów opracowanych z zastosowaniem badanej puli starterów wyniósł 22,7%. Polimorfizmem typu CAPS/dCAPS charakteryzowało się 68 markerów. Dla 13 markerów polimorfizm obserwowano bezpośrednio w żelu agarozowym (polimorfizm wielkości amplifikowanego produktu lub obecność produktu amplifikacji tylko u jednego z rodziców). Nowe markery zlokalizowano we wszystkich z dwudziestu opisanych grup sprzężeń łubinu wąskolistnego (od NLL-01 do NLL-20), a ich rozmieszczenie w poszczególnych grupach sprzężeń było zróżnicowane. Najwięcej – 10 markerów – zmapowano w grupie NLL-06. Obecnie całkowita długość mapy wynosi 2611,1 cM, a średnia odległość między markerami to 3,76 cM.



Rys. 1a. Mapa genetyczna tubinu wąskolistnego z naniesionymi nowymi markerami STS (podkreślone)
 Fig. 1a. Linkage map of the narrow-leafed lupin genome with the new STS markers (underlined)



Rys. 1c. Mapa genetyczna łubinu wąskolistnego z naniesionymi nowymi markerami STS (podkreślone)
 Fig. 1c. Linkage map of the narrow-leaved lupin genome with the new STS markers (underlined)



Rys. 2. Terminy GO zidentyfikowane dla 32 sekwencji markerów STS, sklasyfikowane według procesów biologicznych (terminy poziomu 2)

Fig. 2. Gene Ontology (GO) terms for the 32 STS markers, defined according to the biological processes (level 2 terms)

Wzbogacanie map genetycznych w markery reprezentujące sekwencje unikatowe w genomie, które dodatkowo ulegają ekspresji jest bardzo użyteczne, ze względu na możliwość identyfikacji i izolacji ważnych genów, wykrywania loci cech ilościowych oraz selekcję materiału hodowlanego z wykorzystaniem markerów – MAS (ang. Marker Assisted Selection) [Sato i in. 2005]. Ponadto, niezwykle cenna jest możliwość ich wykorzystania w mapowaniu porównawczym genomów gatunków pokrewnych, umożliwiającą wykorzystanie wiedzy dotyczącej struktury i funkcji genów gatunków modelowych w analizie i identyfikacji genów odpowiedzialnych za ważne cechy użytkowe gatunków o słabiej poznanych genomach [Zhu i in. 2005, Ellwood i n. 2008].

Analizę funkcjonalną *in silico* przeprowadzono dla 68 sekwencji nukleotydowych opracowanych markerów łubinu wąskolistnego (markery o polimorfizmie CAPS/dCAPS) z wykorzystaniem oprogramowania Blast2GO. Gatunkami, dla których najczęściej odnajdywano homologię sekwencji w analizie BLAST (Top-Hit species distribution) były: soja (*Glycine max*) – 11 sekwencji i *Medicago truncatula* – 6 sekwencji. Na tej podstawie można przypuszczać, że

najbardziej reprezentatywnym genomem referencyjnym dla łubinu wąskolistnego jest genom soi. Podobne wyniki odnotowali Pazos-Navarro i inni (2011) w badaniach dotyczących mało poznanego gatunku motylkowatego – *Bituminaria bituminosa*. W badaniach własnych terminy GO zostały zidentyfikowane dla 32 sekwencji (47% analizowanych sekwencji) i zostały sklasyfikowane pod względem lokalizacji komórkowej (terminy poziomu 3), procesów biologicznych (terminy poziomu 2) oraz funkcji molekularnej (terminy poziomu 2).

Największa pula produktów genów (52%) zidentyfikowanych pod kątem funkcji molekularnej charakteryzowała się aktywnością katalityczną, 41% uczestniczyło w wiązaniu, 4% w transdukcji sygnału, a 3% wykazywało aktywność transportową. W przypadku segregacji według lokalizacji komórkowej 54% terminów GO nie zostało precyzyjnie sklasyfikowanych (zostały określone jako niezdefiniowana część komórki lub jako niezdefiniowane organellum komórkowe), 36% przydzielono do organelli związanych z membranami, 6% do kompleksów białkowych, a po 2% do pęcherzyków lub organelli niezwiązanych z membranami.

Przeanalizowane terminy GO charakteryzowały się bardzo dużą różnorodnością pod względem procesów biologicznych, w których uczestniczą. Najwięcej produktów genów zidentyfikowano jako elementy uczestniczące w procesach metabolicznych (29%) oraz w procesach komórkowych (26%). Na rysunku 2 przedstawiono dystrybucję analizowanych sekwencji pod kątem procesów biologicznych.

Mapowanie genetyczne połączone z analizą funkcjonalną *in silico* daje możliwość rozwiązania jednego z najważniejszych problemów badań struktury genomów, a mianowicie powiązania sekwencji nukleotydowych opracowanych markerów z informacjami dotyczącymi ich ekspresji oraz z pełnią przez nie funkcją biologiczną. Adnotacja funkcjonalna pozwala na pogrupowanie genów w klasy funkcjonalne, które mogą okazać się bardzo użyteczne w poznaniu i zrozumieniu funkcji poszczególnych genów [Conesa i in. 2005, Götz i in. 2008]. Dlatego też tworzenie map genetycznych bogatych w markery STS połączone z ich równoczesną analizą funkcjonalną jest niezwykle ważnym narzędziem w poznaniu struktury, organizacji i funkcji genomów uprawnych gatunków niemodelowych, w tym łubinu wąskolistnego.

WNIOSKI

1. Wykorzystanie starterów zaprojektowanych dla ewolucyjnie konserwatywnych sekwencji kodujących pozwoliło wzbogacić mapę genetyczną łubinu wąskolistnego w 81 markerów zdefiniowanych sekwencyjnie.
2. Adnotacja funkcjonalna *in silico* nowych markerów STS umożliwiła ich podział pod względem klas funkcjonalnych.
3. Na podstawie przeprowadzonej analizy funkcjonalnej można przypuszczać, że genom *Glycine max* jest najbardziej reprezentatywnym genomem referencyjnym dla łubinu wąskolistnego.
4. Uzyskane wyniki stanowią podstawę do dalszej identyfikacji markerów bliskosprzężonych z cechami użytkowymi łubinu wąskolistnego, które mogą uprościć selekcję materiału hodowlanego. Opracowane markery mogą być również wykorzystane w analizie syntenii genomów roślin motylkowatych.

PIŚMIENNICTWO

- Conesa A., Götz S., García-Gómez J.M., Terol J., Talón M., Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Ellwood S.R., Phan H.T.T., Jordan M., Hane J., Torres A.M., Avila C.M., Cruz-Izquierdo S., Oliver R.P. 2008. Construction of a comparative genetic map in faba bean (*Vicia faba* L.), conservation of genome structure with *Lens culinaris*. *BMC Genomics* 9: 380–391.
- Götz S., Arnold R., Sebastián-León P., Martín-Rodríguez S., Tischler P., Jehl M.-A., Dopazo J., Rattei T., Conesa A. 2011. B2G-FAR, a species centered GO annotation repository. *Bioinformatics* 27: 919–924.
- Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J., Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36: 3420–3435.
- Manly K.F., Cudmore R.H. Jr, Meer J.M. 2001. Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome* 12: 930–932.
- Nelson M.N., Moolhuijzen P.M., Boersma J.G., Chudy M., Leśniewska K., Bellgrad M., Oliver R.P., Święcicki W., Wolko B., Cowling W.A., Ellwood S.R. 2010. Aligning a new reference genetic map of *Lupinus angustifolius* with the genome sequence of the model legume, *Lotus japonicus*. *DNA Res.* 17: 73–83.
- Pazos-Navarro M., Dabauza M., Correal E., Hanson K., Teakle N., Real D., Nelson M.N. 2011. Next generation DNA sequencing technology delivers valuable genetic markers for the genomic orphan legume species, *Bituminaria bituminosa*. *BMC Genet.* 12: 104–117.
- Sato S., Isobe S., Asamizu E., Ohmido N., Kataoka R., Nakamura Y., Kaneko T., Samurai N., Okumura K., Klimentio I., Sasamoto S., Wada T., Watanabe A., Kohara M., Fujishiro T., Tabata S. 2005. Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). *DNA Res.* 12: 301–364.
- Voorrips R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93: 77–78.
- Zhu H., Choi H.-K., Cook D.R., Shoemaker R. C. 2005. Bridging model and crop legumes through comparative genomics. *Plant Physiol.* 137: 1189–1196.

K. KAMEL, M. KROC, W. ŚWIĘCICKI

GENETIC MAPPING OF STS MARKERS IN NARROW-LEAFED LUPIN (*LUPINUS ANGUSTIFOLIUS* L.) GENOME AND THEIR FUNCTIONAL ANNOTATION

Summary

Narrow-leafed lupin (NLL) is one of the three lupin crop species. Lupin farming gained recently much attention, due to the fact that lupin is a valuable source of protein. Unfortunately, our knowledge of the structure and organization of its genome is still rather limited. The main goal of this study was to saturate the most up-to-date reference genetic map of NLL with gene-based, sequence-tagged site markers. A total number of 85 polymorphic markers were generated and 81 of them were located on the narrow-leafed lupin genetic map. The total map length is now 2611,1 cM and the average distance between markers is 3,76 cM. The next step of our studies was the *in silico* functional analysis of 68 nucleotide sequences of lupin markers. Gene ontology (GO) terms were identified for 32 sequences and defined according to the molecular function, cellular component and biological processes.