

SHRnutí PŘÍNOSŮ ANALÝZ GENOTYPŮ POMOCÍ DARTSEQ PRO BRAMBORÁŘSKOU PRAXI

SUMMARY OF THE BENEFITS OF GENOTYPE ANALYSES USING DARTSEQ FOR POTATO PRACTICE

Jiří PTÁČEK¹, Oldřich TRNĚNÝ², Vladislav KLIČKA³, Jan ŠAFÁŘ⁴, Renata ŠVECOVÁ¹,
Jaroslava DOMKÁŘOVÁ¹, Miroslava ČEPLOVÁ¹

¹Výzkumný ústav bramborářský Havlíčkův Brod, s.r.o.

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

³VESA Velhartice, a.s.

⁴Ústav experimentální botaniky AV ČR, v.v.i., Olomouc

PTÁČEK, J. – TRNĚNÝ, O. – KLIČKA, V. – ŠAFÁŘ, J. – ŠVECOVÁ, R. – DOMKÁŘOVÁ, J. – ČEPLOVÁ, M.
SHRnutí PŘÍNOSŮ ANALÝZ GENOTYPŮ POMOCÍ DARTSEQ PRO BRAMBORÁŘSKOU PRAXI
Vědecké práce – Výzkumný ústav bramborářský Havlíčkův Brod, 2024, 30: 33–42

V rámci platformy NCKTN01000062 „Biotechnologické centrum pro genotypování rostlin“ byl koordinován ve VÚB Havlíčkův Brod Dílčí projekt TN01000062/04 – Brambory (ve spolupráci s ÚEB AV ČR, v.v.i., a VESA Velhartice, a.s.)

Řešení tohoto unikátního projektu probíhalo v letech 2019 až 2022, zde prezentujeme všechny dosažené výsledky.

DNA; DaT; genotypování; brambor; asociační analýza

ÚVOD

Hlízotvorné *Solanum*, které představuje kulturní brambor a jeho plané příbuzné druhy, je velice rozmanitá skupina (HAWKES, 1990). Většina těchto planých druhů obsahuje celou řadu neznámých systémů odolností vůči chorobám a škůdcům, mrazuvzdornosti, suchovzdornosti aj., které se v kulturních formách bramboru vůbec nevyskytují. V jihoamerické a v mexické oblasti existuje řada druhů s odolností k virózám bramboru, popř. s odolností vůči přenašečům těchto viróz, jsou známy druhy, které například nejsou napadány mandelinkou bramborovou. Výskyt genů zodpovědných za tyto vlastnosti je podmíněn především areálem výskytu jednotlivých druhů a podmínkami, jež zde panují.

Genová banka VÚB v Havlíčkově Brodě uchovává kromě odrůd bramboru a různých hybridních materiálů i velmi cenné genové zdroje planých druhů rodu *Solanum*. V tomto souboru je 22 druhů, které hlízy tvoří, a jeden druh, který hlízy netvoří. Druhy jsou udržovány ve více vzorcích a byly získány ze světových kolekcí genetických zdrojů bramboru

(HORÁČKOVÁ a DOMKÁŘOVÁ, 2003). Popis morfologických znaků, hodnocení růstové energie, vegetační doby, zdravotního stavu, výnosu a vybraných hospodářských vlastností je prováděn podle „Klasifikátoru pro genus *Solanum* L.“ (VIDNER *et al.*, 1987).

Brambor je jednou z nejdůležitějších potravinářských plodin na světě. Zlepšení jeho vlastností proto může mít významný dopad. Čtení genomové sekvence bramboru je však velmi složité, protože běžné brambory jsou tetraploidní, tzn. jsou složeny ze čtyř sad chromozomů, což mimo jiné ztěžuje správné poskládání referenčního genomu. Nedávný výzkum se týkal diploidního bramboru pouze s jedním genomem, tzv. homozygotem, který usnadňuje čtení a porovnání sekvencí DNA. Tato rostlina, Solyntus, byla vyprodukována jako součást hybridního šlechtitelského programu firmy Solynta.

V současnosti je již také dostupná referenční sekvence v sestavení po chromozomech ve verzi v4.04 (HARDIGAN *et al.*, 2016). Velmi přesná genomová sekvence umožňuje rychlejší a cílenější šlechtění, protože v DNA je snazší najít, která křížení s jinými odrůdami by mohla být zajímavá a kde by výměna genetického materiálu mezi „otcem“ a „matkou“ měla v ideálním případě probíhat. To znamená, že šlechtitelé vědí v rané fázi, zda brambor má požadované vlastnosti, jako je např. odolnost vůči specifickým chorobám.

GWA studie, GWAS

V genetice je hojně využívaným nástrojem *genome-wide association study*, tj. celogenomová asociční studie (studie GWA nebo GWAS), srovnávací studie genetických variant u různých jedinců, která určuje úroveň spojení mezi genetickou variantou a vlastností. GWAS se obvykle zaměřují na asociace mezi jednonukleotidovými polymorfismy (SNP) a fenotypovými vlastnostmi, jako jsou významná geneticky podmíněná lidská onemocnění, ale mohou být stejně aplikovány na jakékoli jiné genetické varianty jakýchkoli jiných organismů a jejich fenotypové znaky.

MATERIÁL A METODY

Z genové banky bramboru byl vybrán soubor fenotypově charakterizovaných genotypů bramboru s rozdílnými hodnotami pro barvu dužniny, ranost, obsah škrobu a rezistence k vybraným biotickým a abiotickým činitelům. Tyto rostliny byly průběžně pasážovány a vedeny jak v podmínkách *in vitro*, tak i *in vivo*. Tento soubor byl doplněn fenotypově částečně charakterizovanými šlechtitelskými materiály, u kterých jsou tato data v průběhu řešení projektu doplňována a současně srovnávána s genotypovými daty.

DNA byla extrahována z 376 genotypů bramboru na pracovišti ÚEB AV ČR Olomouc, v.v.i., kde byl čerstvý materiál nejprve lyofilizován, a poté homogenizován pomocí skleněných kuliček. Pro vlastní izolaci byl použit NucleoSpin Plant II kit (Macherey-Nagel). Byla zkontrolována kvalita DNA a poté každý vzorek naředěn na požadovanou koncentraci.

Izolovaná DNA byla zaslána do Diversity Arrays Technology Pty Ltd, Canberra, Austrálie, k vlastní analýze genotypu pomocí DArTSeq analýzy. Výsledkem DArTSeq analýzy jsou celogenomová genotypová data SNP polymorfismů pro 372 vzorků. Genotypová data byla dále filtrována na chybějící informace, frekvenci minoritní alely a mapována k aktuální verzi genomu bramboru.

Celogenomová asociační analýza byla provedena pro 8 vybraných fenotypových znaků (odolnost k rakovině bramboru, odolnost k hádátku, ranost, obsah škrobu, odolnost k Viru Y bramboru (PVY), odolnost k Viru svinutky bramboru (PLRV), výnos a suchovzdornost) podle dostupných dat hodnocení. Asociační analýza byla provedena pro genotypový soubor SNP polymorfismů s chybějícími údaji do 50 % a s frekvencí minoritní alely (MAF) do 5 %. Práce s genotypovými daty byla prováděna zejména pomocí balíku dartR, vcfR v R. Pro analýzu genetické diverzity nebyly dále použity vysoce korelované polymorfismy ve vazbě ($r^2 > 0.6$) do vzdálenosti 10 000 bp a také byl v rozmezí 500 bp ponechán vždy jen jeden polymorfismus s vyšší MAF za využití softwaru bcftools. Pro analýzu genetické diverzity nebyly také použity polymorfismy, které mapují na více pozicích v genomu (blastn evaluate < 0.00001). Asociační analýza byla provedena pomocí programu GAPIT implementovaného v jazyce R. Pro výpočet byl použit statistický výpočetní model Blink (HUANG *et al.*, 2019) s korekcí na genetickou strukturu populace, genetickou příbuznost a další matoucí efekty. Výběr počtu PCs, které sloužily ke korekci na genetickou strukturu uvnitř sledované populace, byl určen pomocí BIC s maximálním počtem 3 PCs. Pro zahrnutí vlivu příbuznosti jedinců byl použit koeficient příbuznosti dle VAN RADEN (2008). Pro imputaci byla použita majoritní alela. Výsledky byly korigovány pomocí *false discovery rate* FDR (BENJAMINI a HOCHBERG, 1995) pro mnohonásobné porovnání. Práh FDR pro signifikantní asociace byl zvolen na hladině 0,05. Spolehlivost výsledků asociace byla posouzena pomocí kvantil-kvantil Q-Q grafu při porovnání očekávaných výsledků proti získaným. Výsledky asociace polymorfismů a jejich fyzická pozice byly vyneseny do Manhattan grafů.

VÝSLEDKY A DISKUSE

Výsledky asociačních celogenomových analýz genotypů bramboru prokázaly vazbu mezi fenotypovým projevem a sekvenční variabilitou genotypů brambor. Znalost markerů asociovaných s požadovaným znakem umožňuje sledovat prostupnost genetické informace z rodičů na potomky napříč generacemi. U barvy dužniny můžeme například využít znalost asociovaného markeru k cílenému přikřížení znaku barvy pomocí zpětného křížení, tak abychom co nejlépe zachovali genetické pozadí obohacovaného genotypu. U jiných znaků může znalost asociovaného markeru umožnit šlechtitelům dříve a snáze rozpoznat tuto zájmovou vlastnost již v raných fázích vývoje potomků. Přiložené grafy reprezentují výsledky výše uvedené asociace jednotlivých vazeb mezi vybranými fenotypovými projevy

a genotypovým profilem. Podrobný popis je uveden v naší předešlé práci (PTÁČEK *et al.*, 2020), kde je popsán i význam jednotlivých grafů.

Přiložené grafy (Obr. 1) prezentují výše uvedené asociace, Q-Q graf výsledků asociace potvrzuje statistickou průkaznost výsledků. Výsledky asociace k odolnosti k hádátku nejsou na uvedeném obrázku prezentovány. Pro posouzení asociace byl použit statistický model Blink. Model Blink - Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (HUANG *et al.*, 2019) využívá nový postup opakovaného hodnocení asociace, který vybírá v hodnoceném lokusu v genomu to SNP, které nejlépe vyjadřuje změny v hodnotách hodnoceného znaku. Dále byly hodnoty asociace jednotlivých SNP korigovány podle genetické příbuznosti položek mezi položkami a podle genetické struktury, která byla již dříve rozpoznána v hodnocené kolekci brambor. Polymorfismy SNP označené touto metodou za nejvíce asociované v jednotlivých lokusech jsou obzvláště vhodné pro jejich použití v markery asistované selekci, neboť jsou očištěny od vlivu ostatních blízkých SNP. Paralelně s přicházejícími výsledky asocičních studií byl zahájen převod vybraných SNP u některých znaků na PCR markery, které budou použity ve šlechtění.

U vybraných sledovaných znaků (odolnost k hádátku, ranost, obsah škrobu, odolnost k PVY, odolnost k PLRV) jsme doplnili nově získaná fenotypová data šlechtitelských materiálů a analýza byla provedena opakovaně. Nyní již bylo zjištěno vazby mezi fenotypovým projevem a genetickým profilem statisticky průkazné. Statisticky průkazné vazby mezi fenotypovým projevem a genetickým profilem byly získány i u nově zařazených fenotypových znaků (výnos a suchovzdornost). U všech osmi testovaných fenotypových znaků byly prokázány statisticky průkazné vazby mezi fenotypovým projevem a genetickým profilem, což vedlo k získání souborů asociovaných markerů pro jednotlivé sledované znaky, které jsou vhodné k markery asistované selekci a k převodu na PCR markery. U znaků výnos, suchovzdornost, obsah škrobu, ranost a u znaků odolností k rakovině, k PVY a k PLRV byly nalezeny desítky statisticky průkazných SNP na úrovni významnosti $P = 0.05$ po korekci na mnohonásobné porovnání.

Následně byla analyzována fenotypová data u dalších znaků. Vhodnost brambor pro přípravu lupínků, hranolků a kaší. Hodnocení odolnosti k mechanickému poškození a odolnosti k napadení plísní bramboru. Též byly také asociovány s genetickou variabilitou chuťové parametry brambor.

Přiložené grafy (Obr. 2) prezentují nalezené asociace genotypových variant s fenotypovým projevem, Q-Q graf výsledků asociace potvrzuje statistickou průkaznost výsledků, zejména průkazné asociace lze sledovat u vhodnosti brambor k přípravě lupínků, hranolek a kaše. Výsledky asociace k odolnosti k hádátku nejsou na uvedeném obrázku prezentovány. Pro posouzení asociace byl použit statistický model Blink. Model Blink - Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (HUANG *et al.*, 2019) využívá postup opakovaného hodnocení asociace, který vybírá v hodnoceném lokusu v genomu to SNP, které nejlépe vyjadřuje změny v hodnotách hodnoceného znaku. Dále byly

hodnoty asociace jednotlivých SNP korigovány podle genetické příbuznosti položek mezi položkami a podle genetické struktury, která byla již dříve rozpoznána v hodnocené kolekci brambor. Polymorfismy SNP označené touto metodou za nejvíce asociované v jednotlivých lokusech jsou obzvlášť vhodné pro jejich použití v markery asistované selekci, neboť jsou očištěny od vlivu ostatních blízkých SNP. Paralelně s přicházejícími výsledky asocičních studií byl zahájen převod vybraných SNP u některých znaků na PCR markery, které budou použity ve šlechtění.

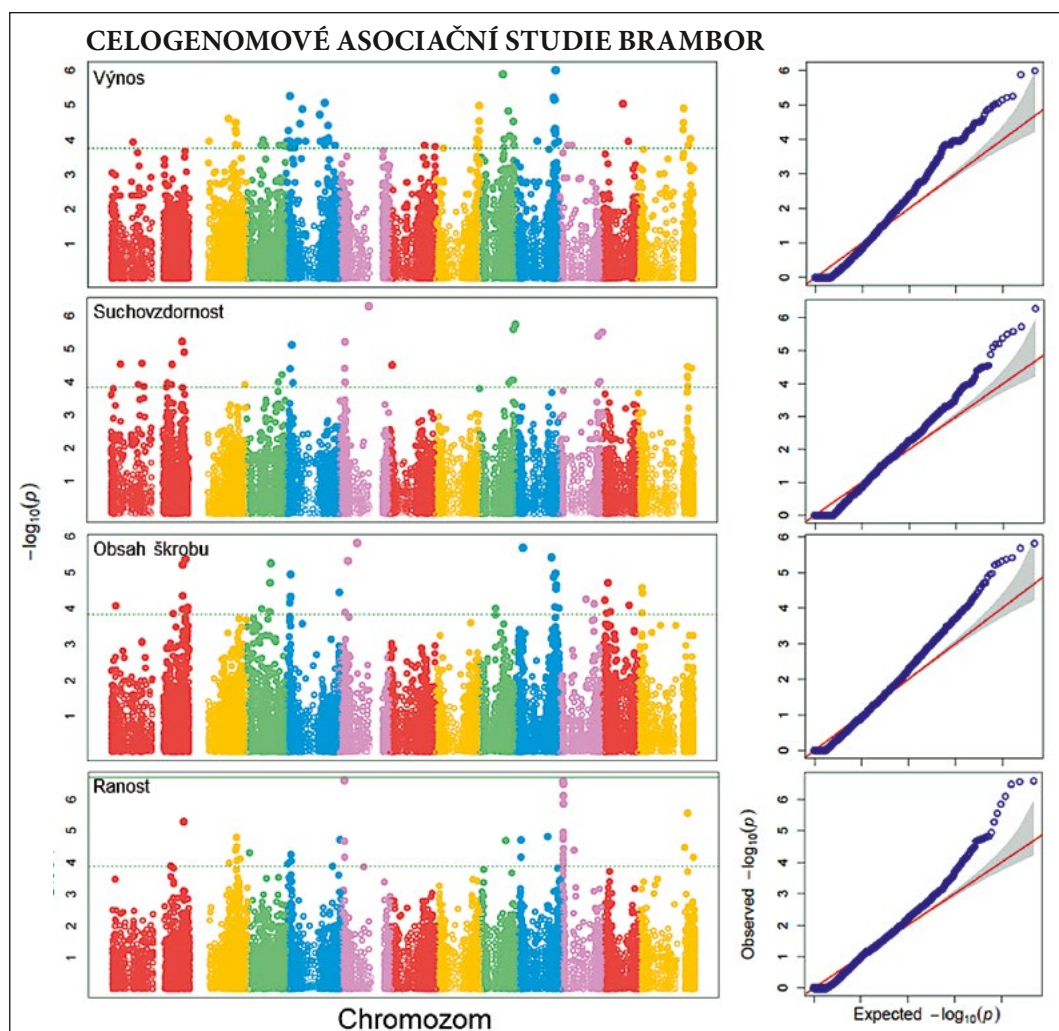
U těchto vybraných sledovaných znaků (vhodnost pro přípravu lupínků, hranolků a kaší, chuť hlíz po uvaření ve slupce v páře, odolnost k mechanickému poškození a odolnost k plísni bramboru) jsme doplnili nově získaná fenotypová data šlechtitelských materiálů a analýza byla provedena opakovaně. Nyní již bylo zjištění vazby mezi fenotypovým projevem a genetickým profilem statisticky průkazné u vhodnosti pro přípravu lupínků, hranolků a kaší, avšak pro chuť hlíz po uvaření ve slupce v páře, odolnost k mechanickému poškození a odolnost k plísni bramboru byly výsledky asociace statisticky neprůkazné, ale i přes to lze nejlepší výsledky těchto asociací použít pro navržení PCR primerů, jejichž diagnostické využití bude třeba ověřit v potomstvu. Statisticky průkazné vazby mezi fenotypovým projevem a genetickým profilem byly získány tedy u třech nově zařazených fenotypových znaků. Celkově jsme u jedenácti testovaných fenotypových znaků brambor našli desítky statisticky průkazných asociovaných SNP na úrovni významnosti $P = 0,05$ po korekci na mnohonásobné porovnání. Tyto SNP se vztahem k fenotypovému projevu byly vybrány pro tvorbu souboru molekulárních DNA markerů, které jsou vhodné k markery asistované selekci brambor a k převodu na PCR markery.

Doposud dosažené výsledky ukázaly, že by celogenomové genotypování a GWAS přístup mohly být efektivní vzhledem k nákladům a rychlosti získání výsledků nových molekulárních markerů pro selekci zájmových znaků.

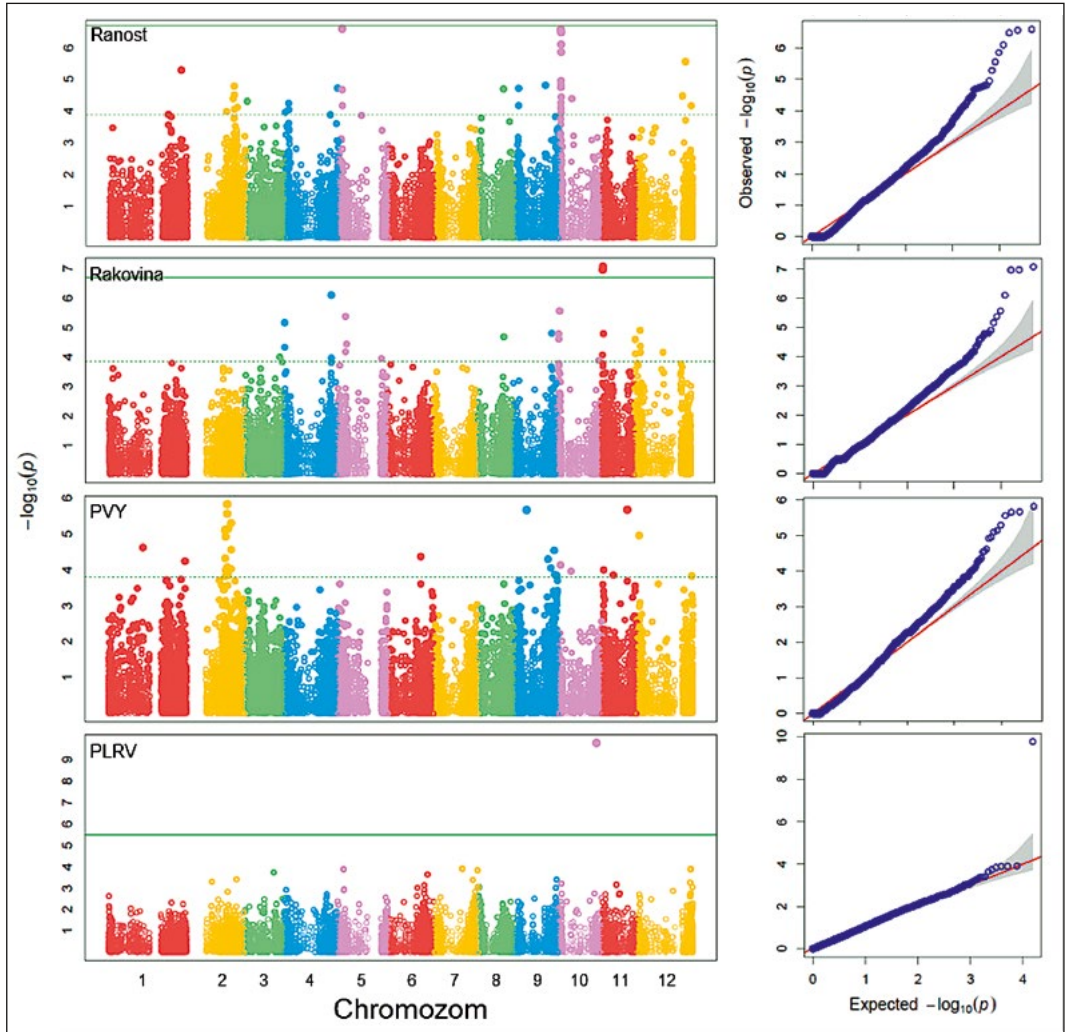
Dále si již musí šlechtitel z těchto souborů asociovaných markerů vybrat ty SNP, u nichž jsou jejich benefitní alely žádanou komponentou nově tvořených odrůd. Při výběru asociovaného SNP pro převod na selekční marker se nesmí opomenout také hledisko dostupnosti benefitních alel v donorových genotypyech v kolekcích genotypů určených pro šlechtění. Vybrané SNP lze převést na markery sériově detekovatelné v potomstvu celou řadou metod.

Doposud dosažené výsledky ukázaly, že by celogenomové genotypování a GWAS přístup mohly být efektivní vzhledem k nákladům a rychlosti získání výsledků.

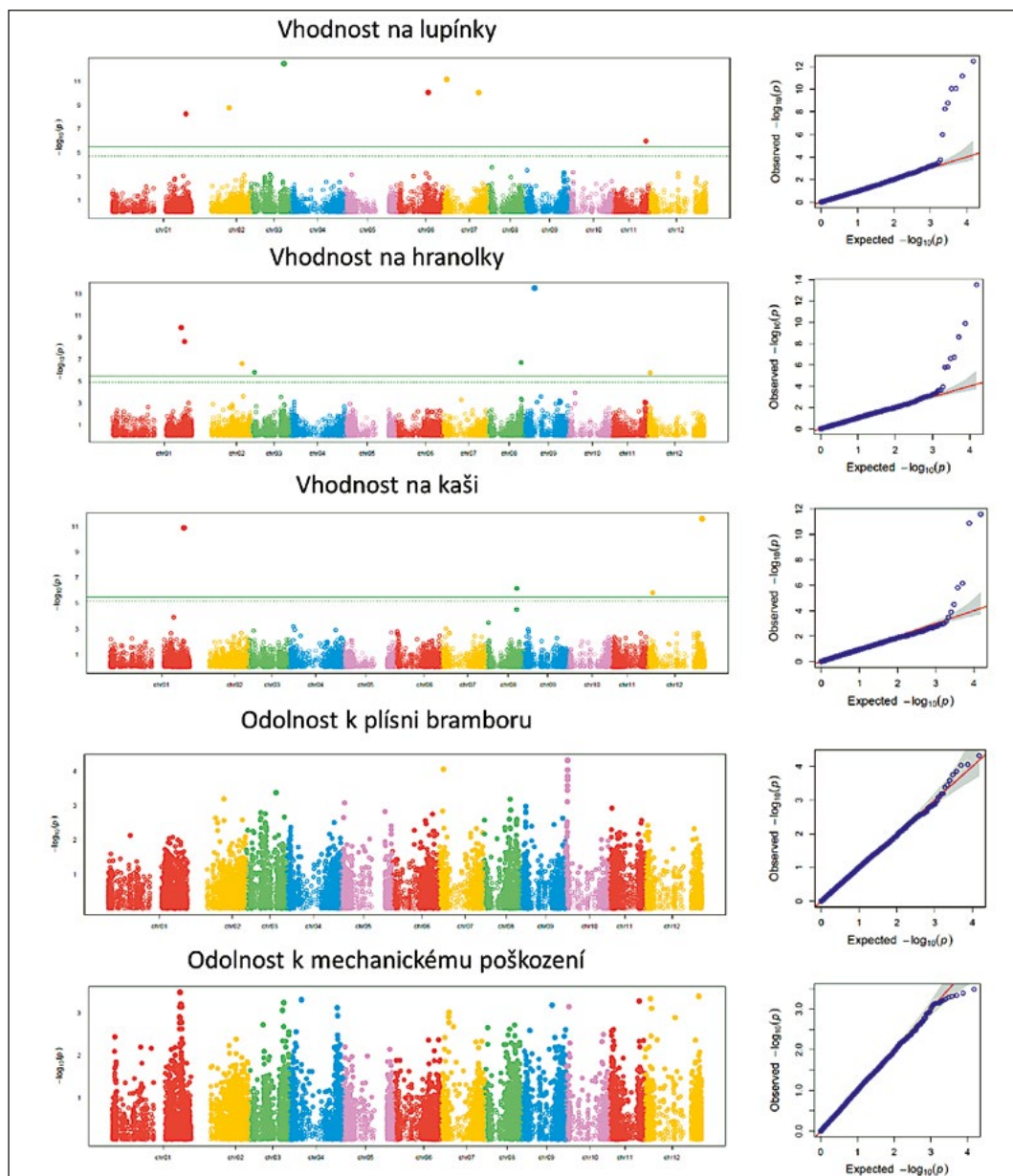
Obr. 1: V levé části obrázku jsou tzv. Manhattan grafy, které graficky znázorňují genotypovou asociaci k sledovanému znaku plodiny. Každý bod představuje jednonukleotidovou záměnu v řetězci DNA tzv. SNP. Úroveň asociace SNP je určena jeho vertikální pozicí v grafu. V horizontálním směru je znázorněna pozice SNP v genomu brambor. Barevně jsou rozlišeny jednotlivé chromozomy genomu lilku bramboru. Úroveň asociace je vyjádřena pomocí záporného dekadického logaritmu P-hodnoty statistického testu vypočtenou metodou BLINK. Zelená čára je hranice statistické významnosti P-hodnoty upravené na mnohonásobné testování na úrovni FDR = 0,05 (přerušovaná) nebo FDR = 0,01 (plná). V pravé části grafu jsou příslušné Q-Q grafy znázorňující distribuci hodnot úrovně asociace pozorovaných a očekávaných, červená čára kopíruje teoretickou hladinu očekávaných hodnot, velká odchylka vyšších pozorovaných hodnot od hodnot očekávaných značí vysokou naměřenou míru asociace mezi genetickým polymorfismem a sledovaným znakem.



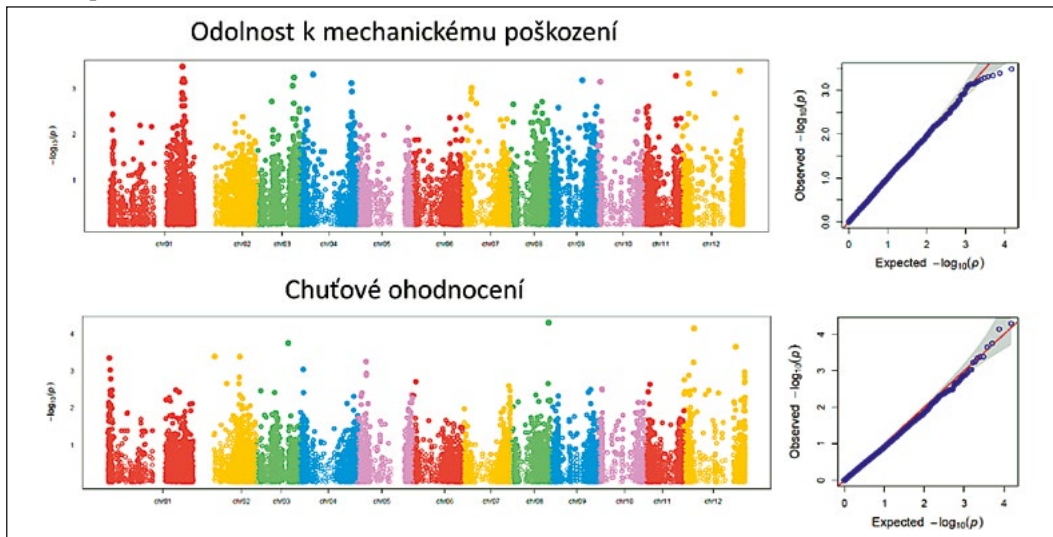
Obr. 1: pokračování



Obr. 2: Grafy celogenomových genetických asociací s vybranými popisnými znaky brambor. V levé části jsou zobrazeny Manhattan grafy síly asociace příslušného znaku s genetickými variantami SNP. Řazení variant SNP je podél referenčního genomu bramboru v4.04 (HARDIGAN *et al.*, 2016). V pravé části jsou v Q-Q grafech přiřazeny k očekávanému rozložení výsledných hodnot hodnoty pozorované, z čehož lze odvozovat statistickou sílu asociace vyjádřenou použitým modelem. Zelená čára označuje hranici hodnoty $P=0,05$ (čárkovaná) a $P=0,01$ (plná) po korekci na mnohonásobné testování.



Obr. 2: pokračování



PODĚKOVÁNÍ

Práce vznikla díky podpoře TA ČR na projekt NCKTN01000062 a Národnímu programu konzervace a využívání genetických zdrojů rostlin a agrobiodiversity MZE-62216/2022-13113/6.2.3.

LITERATURA

- BENJAMINI, Y. – HOCHBERG, Y. (1995): Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1): 289–300.
- HARDIGAN, M. A. – CRISOVAN, E. – HAMILTON, J. P. – KIM, J. – LAIMBEER, P. – LEISNER, C. P. – MANRIQUE-CARPINTERO, N. C. – NEWTON, L. – PHAM, G. M. – VAILLANCOURT, B. – YANG, X. – ZENG, Z. – DOUCHES, D. S. – JIANG, J. – VEILLEUX, R. E. – BUELL, C. R. (2016): Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *The Plant Cell*, 28: 388–405.
- HAWKES, J. G. (1990): *The potato: evolution, biodiversity and genetic resources*. London: Belhaven.
- HORÁČKOVÁ, V. – DOMKÁŘOVÁ, J. (2003): Biologický potenciál genofondu bramboru udržovaného v genové bance *in vitro*. *Vědecké práce – Výzkumný ústav bramborářský Havlíčkův Brod*, 14: 87–101.
- HUANG, M. – LIU, X. – ZHOU, Y. – SUMMERS, R. M. – ZHANG, Z. (2019): BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*, 8(2).
- PTÁČEK, J. – TRNĚNÝ, O. – KOPAČKA, V. – ŠAFÁŘ, J. – ŠVECOVÁ, R. – DOMKÁŘOVÁ, J. – ČEPLOVÁ, M. – KRPÁLKOVÁ, A. (2020): První přínosy celogenomového genotypování pro bramborářskou praxi. *Vědecké práce – Výzkumný ústav bramborářský Havlíčkův Brod*, 26: 9–18.
- VANRADEN, P. M. (2008): Efficient methods to compute genomic predictions. *J Dairy Sci.*, 91(11): 4414–4423.
- VIDNER, J. – DOBIÁŠ, K. – KONRÁD, J. – DĚDIČ, P. – BAREŠ, I. – SEHNALOVÁ, J. (1987): *Klasifikátor – genus Solanum L.* [Classifier – genus *Solanum L.*]. Praha: Výzkumný ústav rostlinné výroby. *Genové zdroje* (31). 45 s.

PTÁČEK, J. – TRNĚNÝ, O. – KLIČKA, V. – ŠAFÁŘ, J. – ŠVECOVÁ, R. – DOMKÁŘOVÁ, J. – ČEPLOVÁ, M.
SUMMARY OF THE BENEFITS OF GENOTYPE ANALYZES USING DARTSEQ FOR POTATO PRACTICE

Vědecké práce – Výzkumný ústav bramborářský Havlíčkův Brod, 2024, 30: 33–42

As part of the platform NCKTN01000062 "Biotechnological Center for Plant Genotyping", Subproject TN01000062/04 – Potatoes was coordinated at PRI Havlíčkův Brod (in cooperation with the IEB of the Academy of Sciences of the Czech Republic and VESA Velhartice.)

The solution of this unique project took place between 2019 and 2022, here we present all the achieved results.

DNA; DArT; genotyping; potato; association analysis

Kontakt:

RNDr. Jiří PTÁČEK, CSc.

Výzkumný ústav bramborářský Havlíčkův Brod, s.r.o.

Dobrovského 2366

580 01 Havlíčkův Brod

Česká republika

tel.: +420 569 466 244

e-mail: ptacek@vubhb.cz