
PRVNÍ PŘÍNOSY CELOGENOMOVÉHO GENOTYPOVÁNÍ PRO BRAMBORÁŘSKOU PRAXI

FIRST BENEFITS OF WHOLE GENOME GENOTYPING METHODS FOR POTATO PRACTICE

Jiří PTÁČEK¹, Oldřich TRNĚNÝ², Viktor KOPAČKA³, Jan ŠAFÁŘ⁴, Renata ŠVECOVÁ¹,
Jaroslava DOMKÁŘOVÁ¹, Miroslava ČEPLOVÁ¹, Alena KRPÁLKOVÁ¹

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

²Zemědělský výzkum Troubsko, s.r.o.

³VESA Velhartice, a.s.

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

PTÁČEK, J. – TRNĚNÝ, O. – KOPAČKA, V. – ŠAFÁŘ, J. – ŠVECOVÁ, R. – DOMKÁŘOVÁ, J. –
ČEPLOVÁ, M. – KRPÁLKOVÁ, A.

PRVNÍ PŘÍNOSY CELOGENOMOVÉHO GENOTYPOVÁNÍ PRO BRAMBORÁŘSKOU PRAXI

Vědecké práce – Výzkumný ústav bramborářský Havlíčkův Brod, 2020, 26: 9–18

V rámci platformy NCKTN01000062 „Biotechnologické centrum pro genotypování rostlin“ koordinujeme 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 započalo v roce 2019 a zde prezentujeme zatím dosažené výsledky.

DNA; DArT; genotypování; brambor

ÚVOD

Hlízotvorné *Solanum*, které představuje kulturní brambor a jeho plané příbuzné druhy, je velice diverzní 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 jihoamerickém a v mexickém genovém centru existuje řada druhů s odolností k virózám bramboru, popř. s odolností vůči přenašečům těchto viróz, jsou známy druhy, které nejsou napadány mandelinkou bramborovou atd. Výskyt genů zodpovědných za tyto vlastnosti je podmíněn především areálem výskytu jednotlivých druhů a podmínkami, jež zde panují.

PVY virus bramboru je nejčastější příčinou komplexu chorob charakterizovaných jako „těžké mozaikové choroby a kadeřavost“ a je vážnou hrozbou v produkci bramboru. Zdroje rezistence i k PVY byly zjištěny v planých druzích a byly vneseny do genomu vypěstova-

ných brambor (VALKONEN *et al.*, 1994). Rezistence PVY jsou rozděleny do dvou hlavních skupin. Hypersenzitivita projevující se nekrotickou reakcí po infekci je způsobena N geny. Rostliny s hypersenzitivními geny vykazují lokální nebo systémové nekrózy, které zabráňují rozšiřování virů. Extrémní rezistence je způsobena R geny, které jsou účinné proti širokému okruhu PVY izolátů. Rostliny s extrémní rezistencí k virům nevykazují nekrotické symptomy po inokulaci PVY. Tyto geny pocházejí ze *Solanum demissum*, *Solanum stoloniferum*, *Solanum hougasii* a *Solanum tuberosum* subsp. *andigena* (HELDÁK *et al.*, 2007). Genová banka VÚB v Havlíčkově Brodě uchovává kromě odrůd bramboru 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 energie počátečního růstu, vegetační doby, zdravotního stavu, výnosu a vybraných hospodářských vlastností je prováděn podle „Klasifikátoru pro genus *Solanum* L.“ (VIDNER *et al.*, 1987).

To může v budoucnu vést k vyšlechtění bramboru, který je odolnější vůči teplu, suchu nebo má větší odolnost vůči chorobám.

Brambor je jednou z nejdůležitějších potravinářských plodin na světě. Zlepšení jeho vlastností proto může mít významný dopad. Čtení genomové struktury bramboru je však velmi složité, protože běžné brambory se skládají ze čtyř genomů, což ztěžuje určení polohy genů. Nedávný výzkum se týkal diploidního bramboru pouze s jedním genomem, tzv. homozygotem, který usnadňuje čtení a porovnání sekvence DNA. Tato rostlina, *Solyntus*, byla vyprodukována jako součást hybridního šlechtitelského programu fy *Solynta*.

V současnosti je již 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 genome-wide association study, tj. celogenomová asociační 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 jedno-nukleotidovými polymorfismy (SNP) a zvláštnostmi, jako jsou hlavní lidská onemocnění, ale mohou být stejně aplikovány na jakékoli jiné genetické varianty a jakékoli jiné organismy.

Manhattanský graf GWAS

Každá tečka grafu představuje SNP, přičemž osa X ukazuje genomickou polohu a osa Y ukazuje úroveň asociace.

Při aplikaci na údaje o lidech porovnávají studie GWA DNA účastníků, kteří mají různé fenotypy pro určitou vlastnost nebo nemoc. Těmito účastníky mohou být lidé s onemocněním (případy) a podobní lidé bez nemoci (kontroly), nebo to mohou být lidé s různými fenotypy pro určitou vlastnost, například krevní tlak. Tento přístup je známý jako první fenotyp, ve kterém jsou účastníci klasifikováni nejprve podle jejich klinických projevů, na rozdíl od genotypu první. Každá osoba dává vzorek DNA, ze které se pomocí polí SNP čtou miliony genetických variant. Je-li jeden typ varianty (jedna alela) častější u lidí s touto nemocí, říká se, že tato varianta je spojena s touto nemocí. Přidružené SNP jsou pak považovány za markery oblasti lidského genomu, která může ovlivnit riziko onemocnění.

Studie GWA zkoumají celý genom, na rozdíl od metod, které specificky testují malý počet předem určených genetických oblastí. Studie GWA identifikují SNP a další varianty v DNA spojené s onemocněním, nemohou však samy určit, které geny jsou příčinné.

První úspěšný GWAS publikovaný v roce 2002 studoval infarkt myokardu. Tento návrh studie byl poté implementován do významné studie GWA 2005, která zkoumala pacienty s makulární degenerací související s věkem, a byly zjištěny dva SNP s významně změněnou frekvencí alel ve srovnání se zdravými kontrolami. V roce 2017, přes 3 000 lidských GWA studií prozkoumalo přes 1800 nemocí a zvláštností a byly zjištěny tisíce SNP asociací. Obecně jsou tyto asociace velmi slabé a mohou mít dokonce zcela bezvýznamný význam, s výjimkou případů vzácných genetických chorob.

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 budou 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

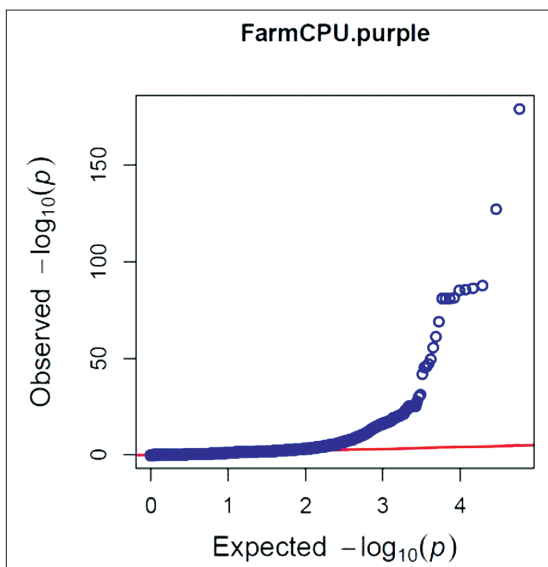
dále filtrována na chybějící informace, frekvenci minoritní alely a mapována k aktuální verzi genomu bramboru.

Celogenomová asociační analýza byla provedena pro vybrané fenotypové znaky (barva dužniny, odolnost k rakovině bramboru, odolnost k hádátku, ranost, obsah škrobu, odolnost k PVY, odolnost k PLRV u 97 až 372 genotypů, 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 R. Statistické výpočetní modely byly použity `MLM` a `FarmCPU` 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-HOCHBERG (1995)` pro mnohonásobné porovnání. Práh `FDR` pro signifikantní asociace byl zvolen na hladině 0,05. Spolehlivost výsledků asociace byla pousouzena pomocí kvantil-kvantil `Q-Q` grafu při porovnání očekávaných výsledků proti získaným. Výsledky asociace polymorfismů a jejich fyzická pozice byly vizualizovány do `Manhattan` grafu. Dále byly lokalizovány geny blízké asociovaným polymorfismům a pro vybrané geny byl navržen potencionální vliv na studovaný fenotypový znak dle literárních zdrojů.

VÝSLEDKY A DISKUSE

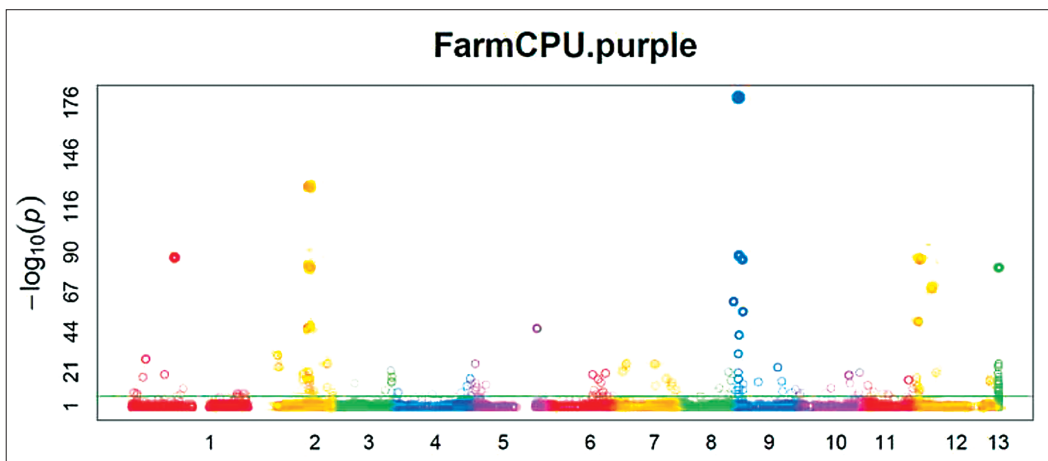
Prvotní výsledky asociačních analýz celogenomových genotypů bramboru prokázaly vazbu mezi fenotypovým projevem a genetickým pozadím pro barvu dužniny. Znalost markerů asociovaných s požadovaným znakem umožňuje sledovat prostupnost genetické informace mající efekt na výsledný fenotyp napříč generacemi. 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. Příložené grafy reprezentují výsledky výše uvedené asociace týkající se barvy dužniny u souboru 372 genotypů bramboru.

Obr. 1: Q-Q graf znázorňuje 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 barvy dužniny brambor.



Obr. 2: Manhattan graf asociace barvy dužniny brambor: úroveň asociace mapovaných genetických variant napříč 12 chromozomy genomu lilku bramboru. Varianty pod číslem 13 jsou genetické varianty nemapující na aktuální verzi genomu lilku bramboru. Úroveň asociace je vyjádřena pomocí záporného dekadického logaritmu P-hodnoty statistického testu vypočtenou metodou FarmCPU. Zelená čára je hranice statistické významnosti P-hodnoty upravené na mnohonásobné testování na úrovni FDR = 0,05. Nejvyšší úroveň asociace byla zjištěna na chromozomu 2 a chromozomu 9.

Dalším hodnoceným znakem byla odolnost k rakovině bramboru. Dva použité statistické

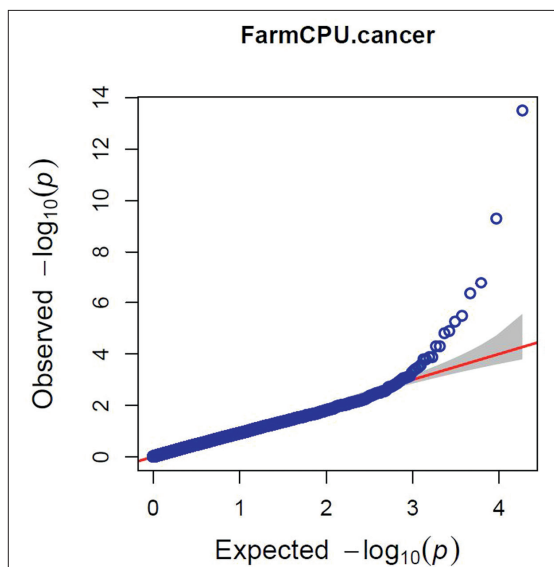


ké modely FarmCPU a MLM prokázaly vazbu mezi fenotypovým projevem a genetickým pozadím pro odolnost k rakovině bramboru lokalizovanou na 11 chromozomu (Obr. 4 a Obr. 5). Prokazatelně asociovaný polymorfismus umožní šlechtitelům zjistit dříve a sná-

ze odolnost šlechtitelského materiálu dle genotypového stavu a v jistých případech může v důsledku i nahradit stávající biologické testy.

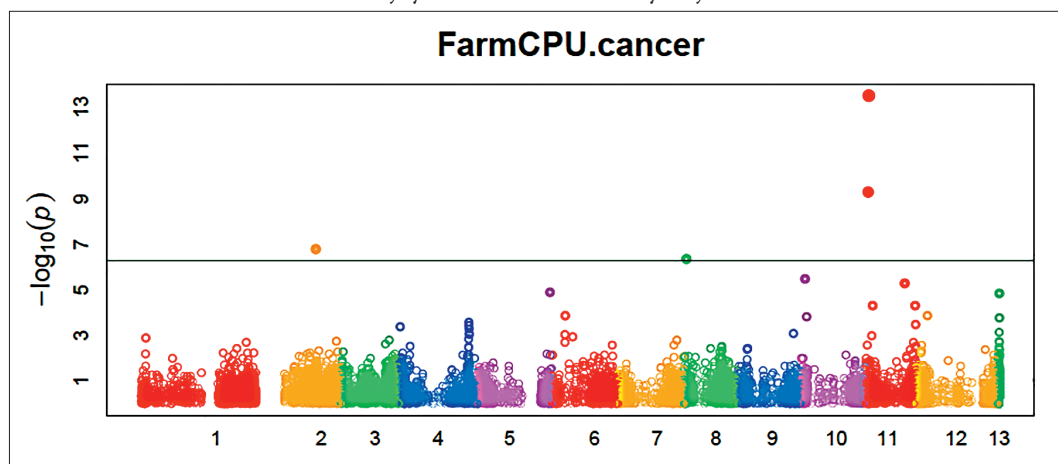
Příložené grafy prezentují výše uvedené asociace (Obr. 2, 4 a 5). Q-Q graf výsledků asociace potvrzuje statistickou věrohodnost výsledků (Obr. 1 a 3). Pro posouzení asociace byly použity dva statistické komplementární modely FarmCPU a Smíšený lineární model (MLM), které společně prokazují věrohodnost výsledků asociace. Místo pro výběr PCR markerů se nachází na chromozomu XI, což odpovídá literárním údajům. Dále byly hodnoceny genetické vzdálenosti jednotlivých položek a genetická struktura, kde byly nalezeny statisticky nejvýznamnější klastry u planých druhů a genotypů s barevnou dužninou.

U dalších sledovaných znaků (odolnost k hádátku, ranost, obsah škrobu, odolnost k PVY, odolnost k PLRV) bylo zjištění vazby mezi fenotypovým projevem a genetickým pozadím doposud statisticky neprůkazné, zejména z důvodu nekompletního popisu hodnoceného souboru jedinců. Avšak v roce 2020 doplníme nově získaná fenotypová data šlechtitelských materiálů a analýza bude provedena opakovaně.

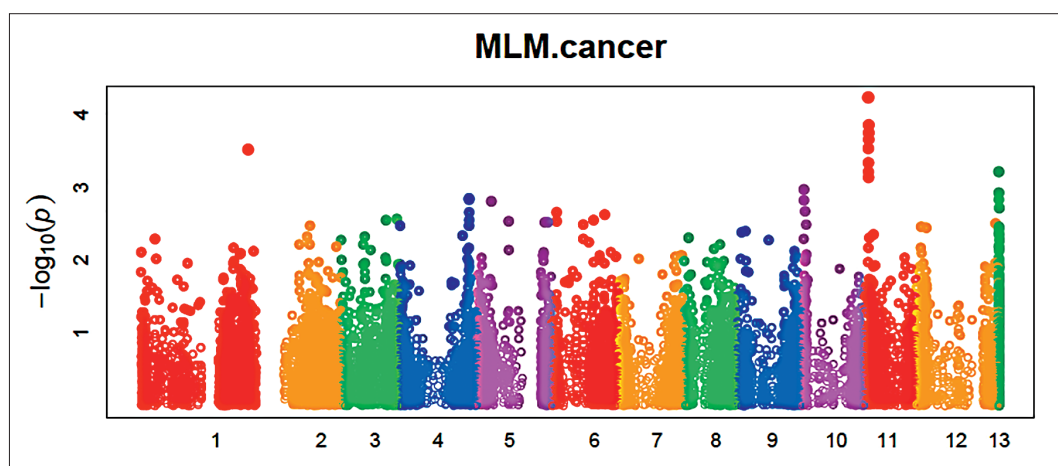


Obr. 3: Q-Q graf znázorňuje distribuci hodnot úrovně asociace pozorovaných a očekávaných, červená čára kopíruje teoretickou hladinu očekávaných hodnot, velká odchylka vyšších pozorovaných hodnot od hodnot očekávaných značí vysokou naměřenou míru asociace mezi genetickým polymorfismem a sledovaným znakem odolnosti k rakovině brambor.

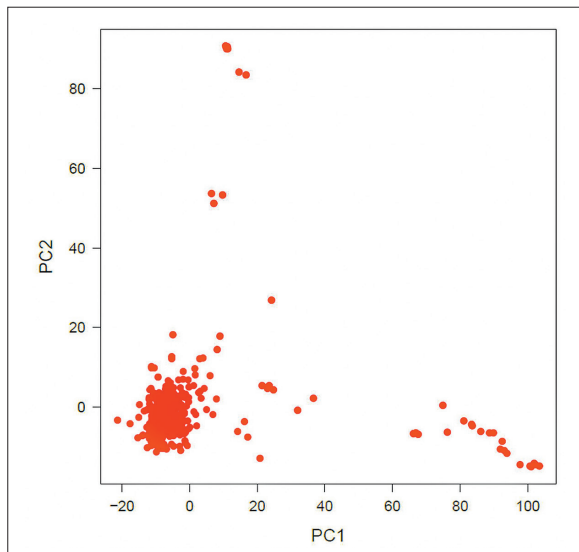
Obr. 4: Manhattan graf asociace odolnosti k rakovině bramboru: úroveň asociace mapovaných genetických variant napříč 12 chromozomy genomu lilku bramboru. Varianty pod číslem 13 jsou genetické varianty nemapující na aktuální verzi genomu lilku bramboru. Úroveň asociace je vyjádřena pomocí záporného dekadického logaritmu P-hodnoty statistického testu vypočtenou metodou FarmCPU. Zelená čára je hranice statistické významnosti P-hodnoty upravené na mnohonásobné testování na úrovni FDR = 0,05. Nejvyšší úroveň asociace byla zjištěna na chromozomu 11.



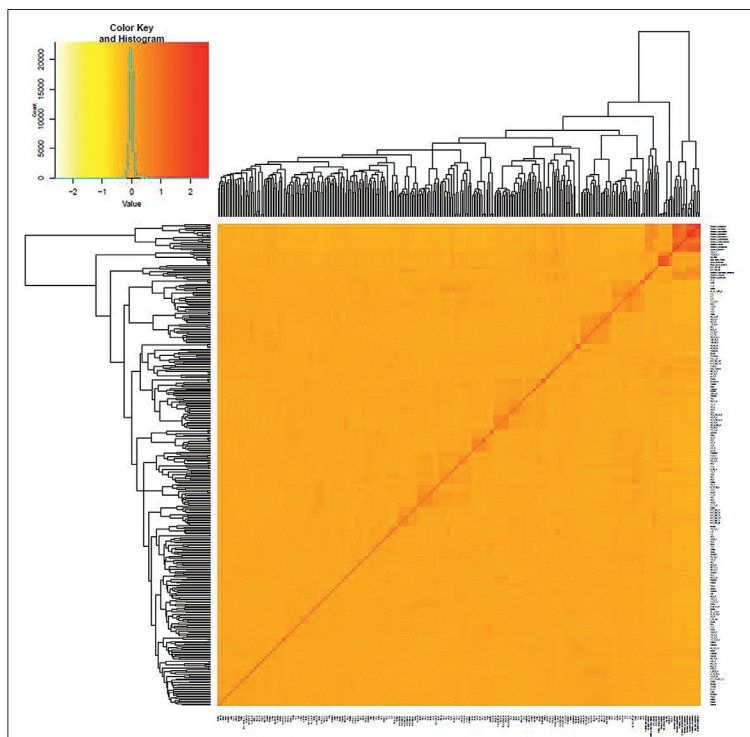
Obr. 5: Manhattan graf asociace odolnosti k rakovině bramboru: úroveň asociace mapovaných genetických variant napříč 12 chromozomy genomu lilku bramboru. Varianty pod číslem 13 jsou genetické varianty nemapující na aktuální verzi genomu lilku bramboru. Úroveň asociace je vyjádřena pomocí záporného dekadického logaritmu P-hodnoty statistického testu vypočtenou metodou MLM. Nejvyšší úroveň asociace byla zjištěna na chromozomu 11 stejně jako při statistickém testu metodou FarmCPU prezentovanou v Obr. 4.



Obř. 6: Genetická struktura sledovaného souboru 282 položek brambor.



A) Analýza hlavních komponent genetické variability



B) Shluková analýza položek podle genetické vzdálenosti znázorněna pomocí teplotní mapy a dendrogramu hierarchického shlukování

PODĚKOVÁNÍ

Práce vznikla díky podpoře TA ČR na projekt NCKTN01000062 a Národnímu programu konzervace a využívání genetických zdrojů rostlin a agrobiodiversity 51834/2017-17253/62.3.

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*. Belhaven Press, London, UK.
- HELDÁK, J. – BEŽO M. – ŠTEFŮNOVÁ V. – GALLIKOVÁ, A. (2007): Selection of DNA markers for detection of extreme resistance to potato virus Y in tetraploid potato (*Solanum tuberosum* L.) F1 progenies. *Czech J. Genet. Plant Breed.*, 43: 125–134.
- 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.
- VALKONEN, J.P.T. – XU, Y.-S. – PULLI, S. – PEHU, E. – ROKKA, V.-M. (1994): Transfer of resistance to potato leafroll virus, potato virus Y and potato virus X from *Solanum brevidens* to *S. tuberosum* through symmetric and designed asymmetric somatic hybridisation. *Ann. of Applied Biology*, 124(2): 351–362.
- 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. – KOPAČKA, V. – ŠAFÁŘ, J. – ŠVECOVÁ, R. – DOMKÁŘOVÁ, J. –
ČEPLOVÁ, M. – KRPÁLKOVÁ, A.

THE FIRST BENEFITS OF WHOLE GENOME GENOTYPING METHODS FOR POTATO PRACTICE

Vědecké práce – Výzkumný ústav bramborářský Havlíčkův Brod, 2020, 26: 9–18

Within the platform NCKTN01000062 „Biotechnological center for plant genotyping“ we coordinate in PRI Havlíčkův Brod Sub-project TN01000062 / 04 - Potatoes in cooperation with IEB CAS and VESA Velhartice. The project execution started in 2019 and here we present the results achieved so far.

DNA; DArT; genotyping; potato

Kontaktní adresa:

RNDr. Jiří PTÁČEK, CSc.

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

Dobrovského 2366, 580 01 Havlíčkův Brod

tel.: +420 569 466 244, e-mail: ptacek@vubhb.cz