



Fakulta zemědělská  
a technologická  
Faculty of Agriculture  
and Technology

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice

## **Metodika detekce genů odolnosti vůči antraknóze u luskovin**

Metodika byla vypracována jako výstup projektu FW10010461 „Moderní technologie ve šlechtění olejnin a luskovin – zavedení tvorby dihaploidů a molekulární identifikace donorů rezistence ve šlechtitelských programech olejnin a luskovin“

Autoři: Ing. Mgr. Ondřej Hejna, Ph.D.,  
Ing. Kristýna Beranová, Ing. Eva Jozová Ph.D., prof. Ing. Vladislav Čurn, Ph.D.

České Budějovice, 2025



**Metodika detekce genů odolnosti vůči antraknóze u luskovin**

Metodika byla vypracována jako výstup projektu FW10010461 „Moderní technologie ve šlechtění olejnin a luskovin – zavedení tvorby dihaploidů a molekulární identifikace donorů rezistence ve šlechtitelských programech olejnin a luskovin“

Ing. Mgr. Ondřej Hejna, Ph.D.  
Ing. Kristýna Beranová  
Ing. Eva Jozová Ph.D.  
prof. Ing. Vladislav Čurn, Ph.D.

České Budějovice, 2025

Metodika detekce genů odolnosti vůči antraknóze u luskovin

Hejna O. et al. 2025

hejna@fzt.jcu.cz

Katedra genetiky a biotechnologií, FZT JU v Českých Budějovicích, České Budějovice

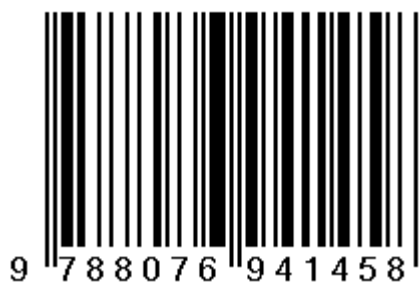
[www.fzt.jcu.cz](http://www.fzt.jcu.cz)

Vypracováno za podpory projektu FW10010461 „Moderní technologie ve šlechtění olejnin a luskovin – zavedení tvorby dihaploidů a molekulární identifikace donorů rezistence ve šlechtitelských programech olejnin a luskovin“

Recenzenty metodiky byli:

*Text: ©2025 Hejna O. a kol.*  
Vydáno bez jazykové úpravy  
ISBN: 978-80-7694-145-8

ISBN 978-80-7694-145-8



## Obsah

Uvedení problému a cíl metodiky .....	5
Vlastní popis metodiky .....	6
Úvod .....	6
Pěstování a sběr vzorků .....	8
Extrakce celkové RNA .....	8
Přípravná fáze.....	8
Příprava vzorků.....	8
Chloroformová fáze.....	9
Precipitační fáze .....	9
Promývací fáze .....	9
Rozpouštěcí fáze .....	9
Přečišťovací fáze (volitelné).....	10
Odstranění genomové DNA.....	10
Příprava sekvenačních knihoven .....	11
Kontrola integrity RNA a požadavky na vzorek.....	11
Příprava 1× fragmentační směsi pro eluci RNA .....	11
Čištění magnetických kuliček .....	11
Dvoustupňová separace mRNA.....	12
Eluce a fragmentace separované mRNA.....	13
Syntéza prvního řetězce cDNA .....	13
Syntéza druhého řetězce cDNA.....	14
Purifikace dvouřetězcové cDNA pomocí kuliček SPRIselect .....	14
Úprava konců cDNA knihovny .....	15
Ligace adaptorů .....	16
Zvýšení množství cDNA s adaptory pomocí PCR.....	16
Čištění PCR produktů pomocí kuliček SPRIselect.....	17
Bioinformatická analýza .....	18
Kontrola kvality a předzpracování RNA-seq dat.....	18
Mapování RNA-seq čtení na referenční genom .....	19
Identifikace polymorfizmů jednotlivých genotypů .....	19
Filtrování identifikovaných polymorfizmů.....	21

Odhad struktury populace .....	26
Praktické použití metodiky .....	29
Srovnání novosti postupů .....	34
Popis uplatnění metodiky .....	34
Ekonomické aspekty .....	35
Seznam použité literatury .....	37
Seznam publikací předcházející metodice .....	37

## Uvedení problému a cíl metodiky

Strupovitost bobu (antraknóza bobu) je závažné houbové onemocnění způsobené patogenem *Ascochyta fabae* (syn. *Didymella fabae*). Tento patogen napadá především listy, lodyhy a lusky, přežívá v infikovaném osivu a rostlinných zbytcích a při zvýšené vzdušné vlhkosti a vhodných teplotách se šíří velmi rychle. V podmínkách příznivých pro infekci může choroba výrazně snižovat výnosy, v některých případech až o 35–90 %, a v extrémních případech vést až k úplnému selhání porostu (Sudheesh et al., 2019; Adhikari et al., 2021).

Z pohledu ochrany rostlin je strupovitost bobu řazena mezi klíčové listové choroby bobu, resp. luskovin (Hampton, 1980; Adhikari et al., 2021). V dostupné genové zásobě je k dispozici převážně kvantitativně podmíněná rezistence, která v oblastech ČR často selhává; vysoká úroveň odolnosti dosud popsána nebyla a rezistence je víceméně polygenní (Sudheesh et al., 2019; Adhikari et al., 2021). Vysoce rezistentní genotypy proto v praxi chybí, a přestože byly identifikovány některé QTL a těsně vázané molekulární markery pro rezistenci k strupovitosti bobu, jejich rutinní využití v markerem asistované selekci je zatím omezené (Sudheesh et al., 2019). Šlechtění dále komplikuje nepravidelný výskyt přirozených epidemií a genetická i patogenní variabilita samotného patogena, což podtrhuje potřebu využití nových genomických a transkriptomických metod, které umožňují přesnější identifikaci genomových oblastí podmiňujících odolnost a jejich efektivnější zapojení do šlechtitelských programů (Adhikari et al., 2021).

Cílem této metodiky je poskytnout ucelený a praktický návod, jak u luskovin identifikovat kandidátní genomové oblasti, a v ideálním případě i konkrétní geny, podmiňující odolnost či toleranci k strupovitosti bobu způsobené výše popsaným patogenem *Ascochyta fabae*.

# Vlastní popis metodiky

## Úvod

Bob polní (*Vicia faba* L.) je významná luskovina pěstovaná jako zdroj bílkovin pro krmné i potravinářské využití. Jednou z klíčových houbových chorob tohoto druhu je strupovitost bobu (antraknóza bobu) způsobená patogenem *Ascochyta fabae*. Tato choroba je řazena mezi hlavní limitující faktory pěstování bobu, přičemž dostupná genetická rezistence je zatím jen částečná a bez podpory vhodných agrotechnických opatření, zejména fungicidů a správného osevního postupu, nedokáže spolehlivě zabránit významným výnosovým ztrátám (Karkanis et al., 2018; Adhikari et al., 2021).

Závažnost strupovitosti bobu dokládají odhadované ztráty výnosu, které mohou u listových chorob bobu včetně strupovitosti dosahovat přibližně 35–90 % v závislosti na intenzitě infekce a pěstitelských podmínkách (Hampton, 1980; Adhikari et al., 2021). Za mimořádně příznivých podmínek pro rozvoj choroby, zejména pokud k epidemii dojde v období květu a nasazování lusků, může dojít k velmi významnému poškození porostu až jeho selhání, podobně jako u dalších závažných listových chorob bobu (Adhikari et al., 2021).

*Ascochyta fabae* přežívá především v infikovaném osivu a napadených rostlinných zbytcích a za vlhkého počasí se šíří na listy, lodyhy a lusky rozstříkáním kapek deště. Infekce vyžaduje delší trvání vlhkosti kolem listů, nicméně při vhodné kombinaci teploty a vzdušné vlhkosti se choroba může objevit prakticky v jakékoli růstové fázi porostu. Přestože patogen preferuje vlhké podmínky, bob polní je v řadě pěstitelských oblastí pěstován spíše v sušších lokalitách, takže přirozený výskyt silných epidemií vhodný pro selekční hodnocení genotypů se objevuje jen ve vlhkých letech. Tento fakt je považován za jeden z hlavních faktorů komplikujících efektivní šlechtění na odolnost k hlavním listovým chorobám bobu včetně strupovitosti bobu (Adhikari et al., 2021).

Z hlediska genetické odolnosti je u pěstovaného bobu polního ke strupovitosti bobu k dispozici převážně jen střední úroveň rezistence; v dostupné genové zásobě nebyly popsány jednoznačně vysoce rezistentní genotypy. Rezistence je obecně hodnocena jako kvantitativní, polygenně podmíněná a je výsledkem kombinace více QTL s relativně malým individuálním účinkem, případně kombinace hlavního genu s dalšími minoritními lokusy (Díaz-Ruiz et al., 2009; Adhikari et al., 2021).

Na základě částečné odolnosti byla v několika pěstitelských oblastech registrována řada odrůd se střední úrovní rezistence ke strupovitosti bobu. V Austrálii jsou jako středně odolné k *Ascochyta fabae* uváděny například odrůdy Farah, Nura, PBA Rana, PBA Samira a PBA Amberley, které současně představují důležité zdroje

genů rezistence pro moderní šlechtitelské programy (Sudheesh et al., 2019; Adhikari et al., 2021).

Rozvoj molekulárních šlechtitelských nástrojů již u strupovitosti bobu pokročil. Pro rezistenci k *Ascochyta fabae* byly popsány některé QTL (Sudheesh et al., 2019; Adhikari et al., 2021). Šlechtění je však nadále komplikováno nepravidelným výskytem patogena a potřebou hodnocení míst na různých místech po několik let, takže samotná fenotypová selekce je jen omezeně efektivní a výraznější zlepšení úrovně rezistence bude pravděpodobně vyžadovat koordinovanější využití genomických metod (Adhikari et al., 2021).

Dalším limitujícím faktorem je variabilita patogena samotného. U *Ascochyta fabae* i dalších hlavních listových patogenů bobu, zejména *Botrytis fabae*, byla popsána výrazná morfologická a patogenní diverzita izolátů, a to jak z hlediska agresivity, tak spektra napadaných genotypů (Stoddard et al., 1999; Adhikari et al., 2021). Tato variabilita komplikuje stabilitu rezistence a vyžaduje testování kandidátních linií proti široké paletě izolátů a geograficky odlišných populací patogena.

Potenciální zdroje silnější odolnosti lze hledat také u planých příbuzných druhů, například *Vicia johannis* a *V. narbonensis*, které jsou obecně vysoce odolné a představují zajímavý genetický zdroj pro šlechtění bobu (Birch et al., 1985; Adhikari et al., 2021). Mezidruhové křížení je však velmi komplikované vzhledem k odlišnému počtu chromozomů a reprodukčním bariérám, a proto zatím nelze tyto druhy ve standardních šlechtitelských programech využít (Birch et al., 1985; Adhikari et al., 2021).

## **Pěstování a sběr vzorků**

Rostliny bobu polního (*Vicia faba*) se pro účely transkriptomických analýz běžně pěstují v laboratorní růstové komoře přibližně 4 týdny při 16hodinové fotoperiodě a intenzitě osvětlení kolem  $100 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ . Z každého genotypu se doporučuje pěstovat minimálně 5, ideálně však 10–15 rostlin, aby byl k dispozici dostatek biologické variability u druhu, který není čistě samosprašný. Pro odběr se sterilní pinzetou odebírají špičky třetího pravého listu (přibližně 100 mg pletiva na vzorek), které se ihned ukládají do 1,5ml mikrozkušavek, rychle se zamrazí v kapalném dusíku a dále se skladují při  $-80^\circ\text{C}$  až do doby izolace celkové RNA z hluboce zmraženého listového pletiva.

## **Extrakce celkové RNA**

### ***Přípravná fáze***

Před zahájením izolace RNA je nutné připravit veškeré potřebné chemikálie, plastový materiál a přístrojové vybavení. K extrakci RNA se použije komerční reagenční směs na bázi fenol–guanidinium thiokyanátu (např. TRI Reagent, TRIzol, RNazol, GENEzol apod.), chloroform, 2-propanol, roztok NaCl o koncentraci 1,2 M, 75% ethanol, DNase-free/RNase-free voda a RNase inhibitor. Dále je třeba mít připraveny sterilní mikrozkušavky o objemu 1,5 ml, sterilní plastové tloučky pro homogenizaci a sterilní špičky s filtrem pro pipetování. Reagenční směs pro extrakci RNA (TRI Reagent, TRIzol apod.) se předem předejde na  $60^\circ\text{C}$ , centrifuga se vychladí na  $4^\circ\text{C}$  a připraví se označené zkumavky s předem odebraným rostlinným materiálem určeným k extrakci.

Zkumavky s odebraným rostlinným materiálem musí být od okamžiku odběru až do samotného zpracování uchovávány buď v kapalném dusíku, nebo při teplotě  $-80^\circ\text{C}$ , aby se předešlo jakýmkoli změnám v expresním profilu genů a degradaci RNA. Všechny kroky se provádějí v podmínkách minimalizujících kontaminaci RNázami: použití RNase-free materiálu, čistý pracovní prostor ošetřený přípravkem například RNase Away.

### ***Příprava vzorků***

V tomto kroku se použijí mikrozkušavky o objemu 1,5 ml, které již obsahují 50–100 mg předem odebraného rostlinného materiálu. Ke každé zkumavce se přidá 800  $\mu\text{l}$  extrakční směsi předejde na  $60^\circ\text{C}$ . Obsah zkumavek se následně homogenizuje sterilním tloučkem tak dlouho, dokud není rostlinný materiál zcela rozmělněn a

vytvoří homogenní suspenzi. Po ukončení homogenizace se do každé zkumavky přidá dalších 200  $\mu$ l téže extrakční směsi, vzorky se krátce promíchají na vortexu a poté se zkumavky ponechají 5 minut stát při pokojové teplotě.

### ***Chloroformová fáze***

Ke každému vzorku se přidá 200  $\mu$ l chloroformu a zkumavky se důkladně promíchají na vortexu po dobu 15 sekund. Následně se vzorky nechají 10 minut stát při pokojové teplotě. Poté se zkumavky centrifugují ve vychlazené centrifuze (4 °C) při 12 000  $\times$  g po dobu 15 minut.

### ***Precipitační fáze***

Po předchozí centrifugaci se z každé zkumavky opatrně odebere přibližně 300  $\mu$ l horní čiré fáze a přenesse se do nové mikrozkušavky. Ke každému vzorku se následně přidá 150  $\mu$ l 2-propanolu a 150  $\mu$ l roztoku NaCl o koncentraci 1,2 M. Obsah zkumavek se důkladně promíchá na vortexu a vzorky se nechají 10 minut stát při pokojové teplotě. Poté se zkumavky centrifugují v předem na 4 °C vychlazené centrifuze při 12 000  $\times$  g po dobu 10 minut.

### ***Promývací fáze***

Po ukončení precipitace se supernatant opatrně odstraní a pelet RNA na dně zkumavky se promyje přidáním 1000  $\mu$ l 75 % etanolu. Vzorky se následně krátce promíchají na vortexu a centrifugují se ve vychlazené centrifuze (4 °C) při 7 500  $\times$  g po dobu 5 minut. Odstranění supernatantu a promytí 75 % ethanolem se ještě jednou zopakuje. Po druhém promytí peletu se RNA nechá sušit při pokojové teplotě přibližně 15 minut. Pelet musí zůstat mírně vlhký a stále průsvitný; nesmí zcela vyschnout, protože by se následně výrazně zhoršila rozpustnost nukleové kyseliny.

### ***Rozpouštěcí fáze***

K průhlednému RNA peletu se přidá 38  $\mu$ l RNase-free vody a 2  $\mu$ l RNase inhibitoru o koncentraci 20 U/ $\mu$ l. Pelet se rozpouští opakovaným nasáváním a vystřikováním. K tomuto účelu se používá 100  $\mu$ l pipeta a samotné promíchávání se provádí opatrně, aby nedocházelo ke zbytečným zlomům na řetězcích nukleových kyselin. Pro kompletní rozpuštění se vzorky inkubují 10 minut při teplotě 55 °C. V posledním

kroku je vhodné pomocí mikrospektrofotometru (například přístroj Nanodrop) určit koncentraci a kvalitu získané RNA.

### ***Přečišťovací fáze (volitelné)***

Případné znečištění extrakčními chemikáliemi lze řešit opakováním precipitační fáze. Ke vzorku o objemu 40 µl se přidá 60 µl RNase-free vody. Následné vysrážení RNA se provede přidáním 50 µl roztoku NaCl o koncentraci 1,2 M a 100 µl 2-propanolu. Po 20 minutách při pokojové teplotě se vzorky centrifugují v předem na 4 °C vychlazené centrifuze při 12 000 × g po dobu 30 minut. V dalším kroku se postupuje stejně jak je uvedeno v promývací a rozpouštěcí fázi.

### **Odstranění genomové DNA**

K odstranění zbytkové genomické DNA ze vzorků použijeme DNA-free DNA Removal Kit od společnosti Invitrogen. K získanému RNA vzorku o objemu 40 µl se přidá 4 µl 10× DNase I pufru a 1 µl rDNase I, poté se obsah zkumavky jemně promíchá opakovaným pipetováním nebo krátkým vortexováním. Směs se inkubuje při 37 °C po dobu 20 minut, aby došlo k úplné degradaci přítomné DNA.

Před ukončením inkubace se opakovaným pipetováním promíchá DNase Inactivation Reagent, aby se sediment rovnoměrně rozptýlil v roztoku. Po ukončení inkubace se ke každému vzorku přidá 4 µl tohoto resuspendovaného inaktivačního reagentu a promíchá pipetováním. Dále se vzorky inkubují 2 minuty při pokojové teplotě v průběhu kterých se vzorky promíchávají, aby zůstal inaktivační reagent rozptýlen.

V následujícím kroku se zkumavky centrifugují při 10 000 × g po dobu 2 minut. Po centrifugaci se opatrně odebere supernatant obsahující RNA a přenesení se do nové, sterilní RNase-free mikrozkušavky, aniž by došlo k narušení peletu z reagentu na dně zkumavky. Je nutné zabránit přenosu tohoto peletu do roztoků, které budou použity v následných enzymatických reakcích, protože inaktivační reagent váže dvojmočné kationty a může tak významně ovlivnit složení reakčního pufru a aktivitu enzymů.

## **Příprava sekvenačních knihoven**

Na přípravu knihoven je potřeba trojice kitů: NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490), NEBNext UltraExpress RNA Library Prep Kit (NEB# E3330) a NEBNext Multiplex Oligos for Illumina - 96 Unique Dual Index Primer Pairs (NEB# E6440). K těmto kitům je nutné mít také vlastní magnetické kuličky SPRIselect (Beckman) a magnetické stojánky pro 0,2 ml a 1,5 ml mikrozkušavky.

### ***Kontrola integrity RNA a požadavky na vzorek***

Před přípravou sekvenačních knihoven je nutné zhodnotit kvalitu a integritu izolované RNA. Integrita RNA se posuzuje pomocí mikročipové elektroforézy na přístroji Shimadzu MCE-202 MultiNA s použitím kitu pro total RNA. U kvalitní RNA se na elektroforeogramu objevují dva ostré píky odpovídající 18S a 28S rRNA, přičemž poměr intenzit 28S/18S by měl být přibližně 2:1. Za kvalitní RNA se obvykle považují vzorky s poměrem 28S/18S v rozmezí přibližně 1,8–2,2.

Dále je třeba, aby RNA byla zbavena solí (např.  $Mg^{2+}$  nebo guanidiniových solí), chelatačních činidel dvoumocných kationtů (např. EDTA nebo EGTA) a organických rozpouštědel (např. fenolu nebo etanolu). RNA musí být zároveň prostá zbytkové DNA odstraněna v předchozích částech DNA Removal kitem. Vstupní množství pro přípravu knihoven činí 25–250 ng DNA-free total RNA v 50  $\mu$ l, přičemž koncentrace se stanoví fluorimetricky pomocí přístroje Qubit (Qubit RNA Assay kit).

### ***Příprava 1× fragmentační směsi pro eluci RNA***

Před zahájením fragmentace RNA se nechá Fragmentation Master Mix (2×) zvolna rozmrazit na ledu. Pro každý vzorek se připraví 1× fragmentační směs smícháním 4  $\mu$ l reagentie NEBNext UltraExpress RNA Fragmentation Mix (fialová) a 4  $\mu$ l nuclease-free vody, tedy celkově 8  $\mu$ l směsi na vzorek. V našem případě, při zpracování 24 vzorků, je potřeba připravit 96  $\mu$ l NEBNext UltraExpress RNA Fragmentation Mix a 96  $\mu$ l nuclease-free vody. Připravená 1× fragmentační směs se uchovává po celou dobu na ledu až do okamžiku, kdy je purifikována mRNA.

### ***Čištění magnetických kuliček***

Pro přípravu k čištění a navázání RNA se nejprve zředí izolovaná RNA RNase-free vodou na konečný objem 50  $\mu$ l v 0,2 ml PCR zkumavce. Zkumavky s RNA se

uchovávají po celou dobu na ledu, dokud nejsou magnetické kuličky připraveny k použití.

Do 1,5 ml mikrozkušavky se následně přidá 480 µl magnetických kuliček NEBNext Oligo d(T)25. Kuličky se pečlivě promíchají jemným protřepáním nebo převrácením zkumavky, nedoporučuje se vortexování. Zkumavka se poté umístí do magnetického stojánku při laboratorní teplotě po dobu, dokud roztok nezprůhlední a kuličky nejsou zcela stažené k magnetu (cca 2 minuty). Supernatant se opatrně odsaje, aniž by došlo k narušení shromážděných kuliček na stěně zkumavky.

Po odstranění supernatantu se zkumavka vyjme z magnetického stojánku a ke kuličkám se přidá 1200 µl neředěného NEBNext RNA Binding Bufferu (2×). Kuličky se důkladně promyjí pipetováním nahoru a dolů (6×). Zkumavka se znovu umístí na magnet a inkubuje se při laboratorní teplotě, dokud roztok nezprůhlední (cca 2 minuty). Poté se supernatant opět opatrně všechen odstraní tak, aby kuličky zůstaly neporušené na stěně zkumavky. Následně se zkumavka opět sejme z magnetu, ke kuličkám se znovu přidá 1200 µl NEBNext RNA Binding Buffer (2×) a obsah se promíchá pipetováním, dokud nejsou kuličky homogenně suspendovány a připraveny k navázání RNA.

### ***Dvoustupňová separace mRNA***

Ke každému z 24 vzorků RNA (50 µl v 0,2 ml PCR zkumavce) se přidá 50 µl důkladně promíchaných magnetických kuliček NEBNext Oligo d(T)25. Vzorky se pečlivě promíchají jemným pipetováním nahoru a dolů (6×), aby došlo k efektivnímu kontaktu RNA s kuličkami. Tato první vazebná fáze slouží k odstranění většiny ne-mRNA frakcí z celkové RNA.

Následně se vzorky zahřejí za účelem denaturace sekundárních struktur RNA a usnadnění vazby poly(A) RNA na oligo(dT) kuličky. Zkumavky se umístí do předem přehřátého termocykleru s víkem nastaveným na  $\geq 90$  °C a inkubují se 2 minuty při 80 °C, poté 5 minut při 25 °C. Následně se zkumavky se vyjmou a umístí na magnetický stojánek při laboratorní teplotě přibližně na 2 minuty, dokud není roztok čirý a kuličky se zcela neshromáždí u magnetu. Supernatant, který obsahuje převážně ne-mRNA, se opatrně odsaje, aniž by došlo k narušení kuliček. Zkumavky zůstávají po tuto dobu na magnetickém stojánku. Kuličky se poté ještě na magnetu šetrně promyjí přidáním 200 µl NEBNext Wash Buffer, který odstraní zbytky nenavázané RNA. Po krátké inkubaci se supernatant opět zcela odsaje, aniž by byl narušen shluk kuliček.

Poté se zkumavky sejmou z magnetického stojánku a ke kuličkám nesoucím mRNA se do každé přidá 50 µl NEBNext Tris bufferu a 50 µl NEBNext RNA Binding Buffer (2×). Směs se důkladně, avšak šetrně promíchá pipetováním nahoru a dolů (6×), aby mohla poly(A) RNA znovu efektivně interagovat s kuličkami ve stejné zkumavce. Zkumavky se pevně uzavřou a znovu se inkubují v termocykleru 2

minuty při 80 °C s 5 minut při 25 °C. Následně se zkumavky vyjmou a opět se umístí na magnetický stojánek při laboratorní teplotě po dobu přibližně 2 minut, dokud roztok není čirý shromážděním kuliček u magnetu.

Po této druhé vazebné fázi se supernatant obsahující zbytky ne-mRNA opatrně a co nejúplněji odstraní, přičemž zkumavky zůstávají stále na magnetickém stojánku. Kuličky se ještě jednou na magnetu promyjí přidáním 200 µl NEBNext Wash Bufferu, který odstraní poslední stopy nenavázané RNA. Supernatant se následně zcela odsaje; v případě potřeby lze zkumavku krátce centrifugovat, opět ji umístit na magnetický stojánek a zbylý pufr odstranit pomocí špičky o objemu 10 µl. Je důležité odstranit veškerý promývací pufr, ale současně nestrhnout pelet kuliček obsahujících mRNA. Kuličky se nesmí nechat zcela vyschnout; po odstranění supernatantu je proto nutné přejít k dalšímu kroku. Zkumavky se poté sejmou z magnetického stojánku.

### ***Eluce a fragmentace separované mRNA***

K eluci mRNA z magnetických kuliček a současně fragmentaci RNA se ke kuličkám obsahujícím navázanou mRNA přidá 6,5 µl předem připravené 1× fragmentační směsi (Fragmentation Master Mix). Směs se důkladně, ale šetrně promíchá pipetováním nahoru a dolů (6×), aby došlo k úplné resuspendaci kuliček a rovnoměrnému kontaktu mRNA s fragmentačním pufrem. Zkumavky se poté umístí do předem předeřátého termocykleru s víkem nastaveným na 105 °C a inkubují se 15 minut při 94 °C. Jakmile teplota ve zkumavkách klesne přibližně na 65 °C, přenesou se ihned na led a nechají se 1 minutu zchladit.

Po krátkém zchlazení se zkumavky rychle centrifugují v mikrocentrifuze, aby se kapalina stáhla ze stěn dolů, a umístí se na magnetický stojánek, dokud roztok není čirý (zpravidla 1–2 minuty). Fragmentovaná mRNA se poté odebere tak, že se 5 µl supernatantu přenesou do nové, sterilní 0,2 ml PCR zkumavky (RNase-free). Pokud je objem supernatantu z jakéhokoli důvodu menší než 5 µl, doplní se na 5 µl přidáním odpovídajícího množství 1× fragmentační směsi. Při odběru je nutné vyhnout se přenosu jakýchkoli magnetických kuliček. Zkumavka s fragmentovanou mRNA se uchovává na ledu a bez prodlení se pokračuje dalšími kroky.

### ***Syntéza prvního řetězce cDNA***

Reakce pro syntézu prvního řetězce cDNA se připravuje na ledu přímo do zkumavky obsahující 5 µl připravené RNA z předchozího kroku. Nejdříve se připraví mix z 96 µl NEBNext UltraExpress Strand Specificity Reagent (hnědá) a 24 µl NEBNext UltraExpress First Strand Enzyme Mix (fialová), který se následně po 5 µl

rozpipetuje do zkumavek s RNA. Po přidání všech komponent se reakční směs důkladně promíchá pipetováním nahoru a dolů (alespoň 10×).

Zkumavky se poté umístí do termocykleru s víkem nastaveným na  $\geq 80$  °C a inkubují se 10 minut při 25 °C, 10 minut při 42 °C a 5 minut při 70 °C, následně je nastavena teplota na 4 °C. Po ukončení inkubace se bez prodlení pokračuje syntézou druhého řetězce cDNA.

### ***Syntéza druhého řetězce cDNA***

Před zahájením syntézy druhého řetězce cDNA se na ledu rozmrazený NEBNext UltraExpress Second Strand Master Mix (oranžová) promíchá převrácením (10×). Následně se krátce centrifuguje, aby se kapalina stáhla ze stěn dolů, a zkumavka se opět uloží na led. Do každé reakce s prvním řetězcem cDNA (10  $\mu$ l z předchozího kroku) se následně přidá 30  $\mu$ l NEBNext UltraExpress Second Strand Master Mix. Zkumavka se ponechá na ledu a reakční směs se důkladně promíchá pipetováním nahoru a dolů (alespoň 10×).

Takto připravené reakce se umístí do termocykleru s víkem nastaveným na teplotu maximálně 40 °C (případně s vypnutým ohřevem víka) a inkubují se 30 minut při 16 °C. Po ukončení inkubace se reakce ponechají v termocykleru při teplotě 4 °C do následujícího kroku.

### ***Purifikace dvouřetězcové cDNA pomocí kuliček SPRIselect***

Před zahájením purifikace se magnetické kuličky SPRIselect důkladně resuspendují krátkým vortexováním, aby vznikla homogenní suspenze. Do každé reakce z předchozího kroku se přidá 72  $\mu$ l resuspendovaných kuliček SPRIselect. Směs se pečlivě promíchá pipetováním nahoru a dolů (alespoň 10×); při posledním promíchání je důležité vytlačit veškerou tekutinu ze špičky zpět do zkumavky. Alternativně lze vzorky 3–5 sekund vortexovat. Následně lze vzorky krátce stočit, ale kuličky se nesmí začít usazovat. Po tomto kroku se vzorky inkubují při laboratorní teplotě minimálně 5 minut, aby se dvouřetězcová cDNA účinně navázala na kuličky.

Po inkubaci se zkumavky krátce centrifugují pro stažení kapaliny ze stěn a umístí se na magnetický stojánek. Jakmile roztok zprůhlední a kuličky se zcela stáhnou k magnetu, supernatant se opatrně odsaje, aniž by došlo k narušení kuliček nesoucích navázanou cDNA. Zkumavky necháme na magnetickém stojánku a kuličky se promyjeme přidáním 200  $\mu$ l čerstvě připraveného 80 % etanolu. Po přibližně 30 sekundách inkubace při laboratorní teplotě se etanol opatrně odstraní.

Tento promývací krok se zopakuje ještě jednou. Po druhém promytí je třeba odstranit veškerou viditelnou kapalinu; v případě potřeby lze zkumavku krátce centrifugovat, znovu ji umístit na stojánek a zbytky etanolu odstranit jemně pomocí 10  $\mu$ l špičky.

Kuličky se následně nechají na magnetickém stojánku na vzduchu oschnout, obvykle po dobu až 5 minut, přičemž víčka zkumavek zůstávají otevřená. Je však nezbytné zabránit jejich přesušení – eluce by měla být zahájena ve chvíli, kdy jsou kuličky stále tmavě hnědé a lesklé, ale ve zkumavkách již není patrná volná kapalina. Jakmile kuličky zesvětlají a začnou „praskat“, jsou nadměrně vysušené a může dojít ke snížení výtěžku cDNA.

Pro eluční krok se zkumavky sejmou z magnetického stojánku a ke kuličkám se přidá 22  $\mu$ l 0,1 $\times$  TE pufru. Směs se důkladně promíchá vortexováním nebo pipetováním nahoru a dolů (10 $\times$ ) a nechá se nejméně 2 minuty inkubovat při laboratorní teplotě. V případě potřeby lze zkumavku krátce centrifugovat, aby se kapalina stáhla ze stěn, a poté se zkumavka znovu umístí na magnetický stojánek. Po přibližně 5 minutách (nebo jakmile roztok zprůhlední) se 20  $\mu$ l eluátu obsahujícího purifikovanou dvouřetězcovou cDNA přenesou do nové 0,2 ml zkumavky. V této fázi lze vzorky krátkodobě uchovávat při  $-20$  °C (např. přes noc), nebo pokračovat v dalších krocích přípravy knihoven.

### ***Úprava konců cDNA knihovny***

Úprava konců dvouřetězcové cDNA se provádí na ledu přímo v eluátu z předchozího purifikačního kroku. Na ledu se nejprve připraví reakční mix složený z 2,5  $\mu$ l NEBNext UltraExpress End Prep Reaction Buffer (zelená) a 1,5  $\mu$ l NEBNext UltraExpress End Prep Enzyme Mix (zelená) na jeden vzorek. Při zpracování 24 vzorků je potřeba připravit celkem 60  $\mu$ l End Prep Reaction Buffer a 36  $\mu$ l End Prep Enzyme Mix, které se smíchají a následně se přidají 4  $\mu$ l této směsi k jednotlivým vzorkům s dvouřetězcovou cDNA.

Po přidání mixu se reakce důkladně promíchá pipetováním nahoru a dolů (alespoň 10 $\times$ ); doporučuje se použít pipetu nastavenou na 20  $\mu$ l a při posledním kroku vytlačit veškerou tekutinu zpět do zkumavky. Míchání musí být důkladné, přítomnost malého množství bublinek v reakci není na závadu. Následně lze zkumavky krátce stočit, aby se kapalina stáhla ze stěn. Zkumavky se poté umístí do termocykleru s víkem nastaveným na  $\geq 75$  °C a inkubují se 5 minut při 20 °C, 10 minut při 65 °C. Po ukončení inkubace se reakce ponechají v termocykleru při teplotě 4 °C a bez prodlení se pokračuje další částí – ligace adaptorů.

## ***Ligace adaptorů***

Před zahájením ligace adaptorů je nutné na ledu naředit reagensii z NEBNext oligos kitu NEBNext Adaptor (červená). Pro vstupní množství 25–250 ng celkové RNA se adaptor ředí 50×. Při zpracování 24 vzorků se připraví 50 µl naředěného adaptoru smícháním 1 µl NEBNext Adaptor a 49 µl Adaptor Dilution Buffer. Naředěný adaptor je po celou dobu uchováván na ledu.

Samotná ligace se sestavuje na ledu přímo do reakce připravené na konci přípravy cDNA konců knihovny. Do 24 µl reakční směsi se postupně přidají 2 µl naředěného adaptoru a 12 µl NEBNext UltraExpress Ligation Master Mix (červená), takže celkový objem ligace činí 38 µl. Ligation Master Mix a adaptor se předem spolu nepřimíchávají, ale vždy se přidávají přímo do reakce. Reakční směs se důkladně promíchá pipetováním přibližně 35 µl nahoru a dolů (alespoň 10×). Pokud je třeba, zkumavky se krátce stočí, aby se kapalina stáhla ze stěn. Ligation Master Mix je výrazně viskózní, proto je nutné věnovat zvýšenou pozornost pečlivému promíchání; malé množství bublinek v reakci nevadí. Zkumavky se následně umístí do termocykleru s víkem nastaveným na  $\leq 40$  °C (případně vypnutým ohřevem víka) a inkubují se 15 minut při 20 °C s doběhem na 4 °C.

Po ukončení ligace se do každé reakce přidají 2 µl NEBNext UltraExpress USER Enzyme (modrá) z NEBNext oligos kitu, čímž se celkový objem zvýší na 40 µl. Směs se opět jemně, ale důkladně promíchá pipetováním nahoru a dolů (cca 10×) při nastavení pipety na přibližně 35 µl. Zkumavky se umístí do termocykleru s víkem nastaveným na  $\geq 45$  °C a inkubují se 5 minut při 37 °C, poté s doběhem na 4 °C. Po tomto kroku se bez prodlení pokračuje další částí.

## ***Zvýšení množství cDNA s adaptory pomocí PCR***

Pro amplifikaci adaptory ligované dvouřetězcové cDNA se používají indexové primery z kitu NEBNext, dodávané ve formátu 96jamkové destičky. V této konfiguraci jsou forward a reverse primery (i5 a i7) již zkombinovány v jedné jamce; výsledná směs má celkovou koncentraci 10 µM (5 µM každý primer). PCR reakce se sestavuje tak, že se do zkumavky přidá 40 µl DNA s ligovanými adaptory z předchozího kroku, 50 µl NEBNext MSTC High Yield Master Mix (modrá) a 10 µl směsi indexových primerů. Celkový objem jedné PCR reakce tedy činí 100 µl. Po přidání všech složek se reakce důkladně promíchá jemným pipetováním nahoru a dolů (10×) a zkumavky se krátce stočí, aby se veškerá kapalina stáhla ze stěn dolů. Takto připravené reakce se umístí do termocykleru s víkem nastaveným na 105 °C a podrobí se PCR amplifikaci podle následujícího programu: počáteční denaturace 30 sekund při 98 °C, následovaná 12 cykly denaturace 10 sekund při 98 °C a nasedání/prodlužování 75 sekund při 65 °C. Po ukončení cyklů následuje

závěrečná prodlužovací fáze 5 minut při 65 °C. Reakce se poté udržuje při 4 °C, dokud nejsou vzorky připraveny k dalšímu zpracování.

### ***Čištění PCR produktů pomocí kuliček SPRIselect***

Čištění PCR produktů se provádí pomocí magnetických kuliček SPRIselect ve dvou navazujících fázích. Před použitím se kuličky důkladně resuspendují vortexováním, aby vznikla homogenní suspenze.

K PCR reakci o objemu přibližně 100 µl se nejprve přidá 70 µl resuspendovaných kuliček SPRIselect. Směs se pečlivě promíchá pipetováním nahoru a dolů (10×); při posledním promíchání je důležité vytlačit veškerou kapalinu ze špičky zpět do zkumavky. Alternativně lze použít krátké vortexování 3–5 sekund a následně zkumavky krátce stočit, než se začnou kuličky usazovat. Vzorky se následně inkubují minimálně 5 minut při laboratorní teplotě, aby došlo k navázání DNA na kuličky. Poté se zkumavky umístí na magnetický stojánek. Po přibližně 5 minutách, nebo jakmile roztok zprůhlední a kuličky se zcela stáhnou k magnetu, se supernatant opatrně odstraní, aniž by došlo k narušení peletu kuliček obsahujících cílovou DNA. V této fázi se kuličky nepromývají ethanolem.

Zkumavky se následně sejmou z magnetického stojánku a k peletu kuliček se přidá 50 µl 0,1× TE pufru, ve kterém se kuličky důkladně resuspendují pipetováním. Poté se přidá dalších 40 µl NEBNext Bead Reconstitution Buffer, takže celkový objem suspenze kuliček činí 90 µl. Směs se opět důkladně promíchá pipetováním nahoru a dolů (10×), případně krátkým vortexováním (3–5 sekund). Při případné centrifugaci je opět nutné zastavit dříve, než se kuličky začnou usazovat.

Vzorky se inkubují minimálně 5 minut při laboratorní teplotě a poté se znovu umístí na magnetický stojánek. Po cca 5 minutách, nebo jakmile roztok zprůhlední, se supernatant opatrně a odsaje, aniž by byl narušen pelet kuliček obsahujících cílovou DNA.

Kuličky se dále promyjí dvěma cykly 80 % etanolu. Zkumavky zůstávají na magnetickém stojánku a do každé se přidá 200 µl čerstvě připraveného 80 % etanolu. Po 30 sekundách inkubace při laboratorní teplotě se etanol opatrně odstraní. Tento promývací krok se zopakuje ještě jednou (celkem tedy dvě promytí). Po druhém promytí je nutné odstranit veškerou viditelnou kapalinu; v případě potřeby lze zkumavky krátce centrifugovat, znovu je umístit na magnet a zbytky etanolu odstranit jemně pomocí 10µl špičky, aniž by došlo k narušení peletu.

Následně se kuličky nechají na magnetickém stojánku oschnout na vzduchu, obvykle po dobu 5 minut, při otevřeném víčku. Je však nezbytné zabránit jejich přesušení – eluce by měla být provedena ve chvíli, kdy jsou kuličky stále tmavě hnědé a lesklé, ale ve zkumavce již není patrná volná kapalina. Jakmile kuličky zesvětlají a začnou „praskat“, jsou nadměrně vysušené a může dojít ke snížení výtěžku DNA.

Pro eluční krok se zkumavky sejmou z magnetického stojánku a ke kuličkám se přidá 23 µl 0,1× TE pufru. Směs se důkladně promíchá pipetováním nahoru a dolů (10×) nebo krátkým vortexováním a nechá se nejméně 2 minuty inkubovat při laboratorní teplotě. V případě potřeby lze vzorky krátce centrifugovat, aby se kapalina stáhla ze stěn, a poté se zkumavky znovu umístí na magnetický stojánek. Po přibližně 5 minutách, nebo jakmile roztok zprůhlední, se 20 µl eluátu obsahujícího finální amplifikovanou knihovnu přenesou do nových PCR zkumavek. Vzorky lze v této fázi uložit při -20 °C.

## Bioinformatická analýza

### ***Kontrola kvality a předzpracování RNA-seq dat***

Sekvenační data z RNA-seq se po demultiplexaci poskytovatelem sekvenačních služeb (např. SEQme) stahují ve formátu FASTQ a ukládají do strukturovaného adresáře projektu. Jednotlivé soubory se přejmenují do jednotného formátu podle označení vzorků a v případě potřeby se standardizují hlavičky čtení ve FASTQ souborech. Následně se provede základní kontrola kvality sekvenačních dat pomocí nástroje FastQC.

V dalším kroku se z čtení odstraňují adaptérové sekvence a nekvalitní úseky pomocí programu Trimmomatic. Typicky se používá nastavení ILLUMINACLIP (soubor s adaptéry adt.fa) s parametry ILLUMINACLIP:adt.fa:2:30:10 pro ořez adaptérových sekvencí, dále SLIDINGWINDOW:4:30 pro ořez na základě klouzavého okna o velikosti 4 bází s minimální průměrnou kvalitou 30 a MINLEN:30 pro odstranění čtení kratších než 30 nukleotidů. Po ořezu se kvalita dat znovu ověří pomocí FastQC a výsledné očištěné FASTQ soubory se komprimují a připraví pro navazující zpracování.

#### ***# Quality and adapters trimming with trimmomatic***

```
for i in $LOS; do
```

```
  trimmomatic PE -threads $TH \
```

```
  ${DIR}/data/${i}.fq.gz \
```

```
  ${DIR}/data/${i}.trim.fq \
```

```
  ILLUMINACLIP:${DIR}/data/adt.fa:2:30:10 SLIDINGWINDOW:4:30 MINLEN:30;
```

```
done
```

## **Mapování RNA-seq čtení na referenční genom**

Pro mapování RNA-seq čtení se používá referenční genom *Vicia faba* (soubor ve formátu FASTA) a odpovídající anotační soubor ve formátu GTF. Nejprve je nutné pro tyto referenční soubory vygenerovat index pro aligner **STAR** pomocí režimu **genomeGenerate**.

Ořezaná a zkontrolovaná data z předchozí části ve formátu FASTQ, komprimovaná gzip se následně mapují na referenční genom pomocí programu STAR. Pro každý vzorek je vhodné mít samostatný výstupní adresář nebo jednotné pojmenování výstupních souborů podle označení vzorku. Jako výstup se generují alignmentové soubory ve formátu BAM seřazené podle genomových souřadnic (outSAMtype BAM SortedByCoordinate).

Po dokončení mapování se výstupní BAM soubory přejmenují a uloží v jednotném formátu (např. <vzorek>.bam) do zvoleného adresáře pro mapování. Dočasné pomocné soubory a pracovní podadresáře generované STAR mohou být následně odstraněny, aby nedocházelo k nadbytečnému zaplňování diskového prostoru. Takto připravené, seřazené BAM soubory slouží jako vstup pro další kroky analýzy.

### **# Map reads to the reference**

```
for i in $LOS; do
```

```
cd ${DIR}/mapping/${i}
```

```
STAR --runThreadN $TH --outSAMtype BAM SortedByCoordinate --  
limitBAMsortRAM 250000000000 --genomeDir  
${DIR}/reference/star_genome_index --readFilesCommand zcat --readFilesIn  
${DIR}/data/${i}.trim.fq.gz
```

```
done
```

```
cd ${DIR}
```

## **Identifikace polymorfizmů jednotlivých genotypů**

Pro identifikaci genetických variant z mapovaných RNA-seq dat (snpy, indely) se jako výchozí používají BAM soubory z předchozí části. V prvním kroku se v jednotlivých vzorcích označí PCR duplikáty pomocí nástroje **GATK MarkDuplicates**, čímž se omezí vliv amplifikovaných artefaktů. Výsledné BAM

soubory s označenými duplikáty se následně indexují (např. pomocí samtools index). Současně je třeba připravit referenční genom ve formátu FASTA – ten se indexuje (samtools faidx) a generuje se pro něj také soubor se sekvenčním slovníkem **GATK CreateSequenceDictionary**, což je nutným vstupem pro další nástroje GATK.

### **# Mark duplicates with GATK**

```
for i in $LOS; do
```

```
gatk MarkDuplicates --INPUT ${DIR}/mapping/${i}.bam --OUTPUT  
${DIR}/mapping/${i}.dp.bam --METRICS_FILE ${DIR}/mapping/${i}.dp.stats
```

```
done
```

Protože RNA-seq čtení mohou spojovat exon–exonové úseky, je u mapovaných dat nutné provést tzv. Split'N'Trim krok pomocí nástroje **GATK SplitNCigarReads**. Ten rozdělí čtení s intronovými úseky (CIGAR s „N“) a upraví jejich zarovnání do formy vhodné pro identifikaci změn. Vzniklé BAM soubory se dále doplní o informace o čtecích skupinách (**GATK AddOrReplaceReadGroups**; parametry jako RGLB – knihovna, RGPL – platforma, RGSM – jméno vzorku), které jsou vyžadovány pro řadu GATK nástrojů a pro korektní zpracování více vzorků. Takto připravené BAM soubory s read groups se znovu indexují a slouží jako vstup pro následující kroky.

### **# Split'N'Trim with GATK**

```
for i in $LOS; do
```

```
gatk SplitNCigarReads --reference $REF --input ${DIR}/mapping/${i}.dp.bam --  
output ${DIR}/mapping/${i}.st.bam
```

```
done
```

Identifikace polymorfizmů se provádí pomocí nástroje **GATK HaplotypeCaller** v režimu GVCF. Pro každý vzorek se zvlášť generuje soubor ve formátu gVCF (--emit-ref-confidence GVCF), který obsahuje informace jak o variantních, tak referenčních pozicích. Tyto gVCF soubory se následně spojí do jednoho společného souboru pomocí nástroje **GATK CombineGVCFs**, čímž vznikne

podklad pro společné genotypování napříč všemi vzorky. Na kombinovaný gVCF se poté aplikuje nástroj **GATK GenotypeGVCFs**, který provede společné volání genotypů a vytvoří výstupní VCF soubor s „raw“ variantami.

### **# Call Variants with GATK**

```
for i in $LOS; do
```

```
gatk HaplotypeCaller --reference $REF --input ${DIR}/mapping/${i}.rg.bam --dont-use-soft-clipped-bases --emit-ref-confidence GVCF --output  
${DIR}/variants_calling/${i}.g.vcf.gz
```

```
done
```

### **# Combine all GVCFs into a joint variant calling set**

```
gatk CombineGVCFs \  
--reference $REF \  
--variant ${DIR}/variants_calling/samples.list \  
--output ${DIR}/variants_calling/combined.g.vcf.gz
```

### **# Genotype the combined GVCF**

```
gatk GenotypeGVCFs \  
--reference $REF \  
--variant ${DIR}/variants_calling/combined.g.vcf.gz \  
--output ${DIR}/variants_calling/raw_variants.vcf.gz
```

## ***Filtrování identifikovaných polymorfizmů***

Pro následnou asociační analýzu je vhodné sadu polymorfizmů výrazně zredukovat, protože celá sada obsahuje velké množství chybně identifikovaných změn, které by znemožnily signifikantně prokázat reálnou asociaci mezi genotypem a fenotypovým projevem. Toho dosáhneme pomocí následujícího vícekrokového filtrovacího schématu.

Nejprve se z výsledného VCF souboru oddělí bialelické SNPy a INDELY a pro jednotlivé varianty se dopočítají základní anotace pomocí nástroje bcftools +fill-tags (AC, AN, AF, NS, MAF, F\_MISSING). Zároveň se ověří přítomnost standardních

GATK anotací (QD, FS, SOR, MQ, ExcessHet, MQRankSum, ReadPosRankSum, DP) a z analýzy se vyřadí vzorky, pro které nejsou k dispozici fenotypové údaje. Tím vznikne pracovní SNP sada, která odpovídá pouze jedincům zahrnutým v následné asociační analýze GWAS.

### **# Normalize and separate biallelic SNPs and recalculate tags**

```
bcftools norm --threads $TH -f $REF -m -any -d all -Ou  
{DIR}/variants_filtering/raw_variants.vcf.gz \  
| bcftools view --threads $TH -v snps -m2 -M2 -c 1 -Ou \  
| bcftools +fill-tags -Ou -- -t AC,AN,AF,NS,MAF,F_MISSING \  
| bcftools view --threads $TH -Oz -o {DIR}/variants_filtering/snps.vcf.gz
```

```
tabix -@ $TH -f -p vcf {DIR}/variants_filtering/snps.vcf.gz  
bcftools stats {DIR}/variants_filtering/snps.vcf.gz | grep -E 'number of  
records:|^TSTV'
```

V dalším kroku se optimalizuje filtrování na úrovni genotypů. Systematicky se vyhodnocuje vliv různých kombinací minimální hloubky čtení (DP) a skóre kvality genotypu (GQ) na počet detekovaných SNP a poměr tranzic a transverzí (Ti/Tv). U RNA-seq dat se v praxi běžně používají prahové hodnoty přibližně  $DP \geq 8-10$  a  $GQ \geq 20-30$ , přičemž konkrétní nastavení představuje kompromis mezi velikostí výsledné SNP sady a její spolehlivostí. Doplnkově lze posuzovat i filtraci podle alelické rovnováhy (allelic balance, AB), tj. podílu čtení podporujících alternativní alelu, která může v některých případech napomoci identifikaci technických artefaktů.

### **# Mask low-quality genotypes to missing**

```
bcftools +setGT "$VCF" -Ob -o "$TMP1" -- -t q -n . -i 'FMT/DP<10 || FMT/DP=="."  
|| FMT/GQ<20 || FMT/GQ==".'"
```

### **# Keep SNPs and require at least one called genotype AND one ALT allele**

```
bcftools +fill-tags "$TMP1" -Ou -- -t AC,AN,NS \  
| bcftools view -v snps -i 'INFO/NS>0 && INFO/AC>0' -Ob -o "$TMP2"
```

### **# Compute tags on the FINAL set, then write gz VCF**

```
bcftools +fill-tags "$TMP2" -Oz -o "$OUT" -- -t AC,AN,AF,MAF,NS,F_MISSING  
tabix -f -p vcf "$OUT"
```

### **# Check the stats**

```
bcftools stats $OUT | grep -E 'number of records:|^TSTV'
```

Protože SNPy v bezprostřední blízkosti INDELů bývají náchylnější k chybám, odhaduje se pro každý SNP vzdálenost k nejbližšímu INDELu (NDIST) pomocí nástroje bedtools. Analýza Ti/Tv v jednotlivých vzdálenostních kategoriích ukazuje výrazně horší poměr u SNPů ležících v těsné blízkosti INDELů, a proto se z další analýzy vyloučí. Současně se z anotace GTF získají exonové oblasti a odstraní se SNPy mimo tyto oblasti.

### **# Remove SNPs near indels**

```
bcftools view -i 'INFO/NDIST>=16 || INFO/NDIST==".' -Ou "$INFONDIST" \  
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \  
| bcftools view -c 1 -v snps -Oz -o "$OUT"  
tabix -f -p vcf "$OUT"  
bcftools stats "$OUT" | grep -E 'number of records:|^TSTV'
```

V dalším kroku se nastavují filtry založené na metrikách spolehlivosti SNP, zejména na ukazatelích strand biasu a pozice čtení v rámci readu. Pro parametry FS (Fisher Strand), SOR (Strand Odds Ratio) a ReadPosRankSum se nejprve doporučuje prozkoumat distribuci hodnot a jejich vliv na poměr tranzicí a transverzí (Ti/Tv) při postupném zpříšňování prahů. V běžné praxi se u kvalitních datasetů používají středně přísné prahové hodnoty přibližně  $FS \leq 20-30$ ,  $SOR \leq 2-3$  a  $|ReadPosRankSum| \leq 3-4$ , které odstraňují varianty s výrazným strand biasem nebo systematickým posunem alternativní alely v rámci čtení, aniž by docházelo k zásadnímu snížení počtu SNP.

### **# Filter SNPs by FS**

```
bcftools view -i 'INFO/FS<=25 || INFO/FS=".' -v snps -c 1 -Ou $VCF \  
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \  
| bcftools view -c 1 -v snps -Oz -o "$OUT"
```

```
| bcftools view -Oz -o $OUT  
bcftools stats $OUT | grep -E 'number of records:|^TSTV'
```

```
tabix -f -p vcf $OUT
```

### **# Filter SNPs by SOR**

```
bcftools view -i 'INFO/SOR<=2.5 || INFO/SOR==".' -v snps -c 1 -Ou $VCF \  
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \  
| bcftools view -Oz -o $OUT
```

```
bcftools stats $OUT | grep -E 'number of records:|^TSTV'
```

```
tabix -f -p vcf $OUT
```

Parametr MQ (RMS Mapping Quality) slouží k hodnocení spolehlivosti zarovnání čtení k referenčnímu genomu. U kvalitních datasetů bývají hodnoty MQ silně soustředěny u vyšších hodnot a nízkých hodnot je minimum. V takovém případě se zpravidla neuplatňuje samostatný filtr na MQ ani na příbuznou metriku MQRankSum.

Na úrovni pozic lze sledovat také podíl chybějících genotypů (F\_MISSING). Pokud je podíl chybějících dat nízký a není patrný systematický vzor chybějících hodnot, má zavedení dodatečného filtru na F\_MISSING obvykle jen omezený přínos.

V některých případech se doplňkově testuje i vliv alelické rovnováhy (allelic balance, AB). Zatímco v DNA studiích se filtry na AB běžně používají k omezení technických artefaktů, u RNA-seq dat je alelický poměr silně ovlivněn expresí transkriptů a pokrytím, a proto je vhodné tyto filtry aplikovat spíše zdrženlivě a pouze po předchozím ověření jejich dopadu na kvalitu a velikost variantního souboru.

Na poziční úrovni se dále filtrují SNP s extrémní hloubkou pokrytí. Na základě distribuce hodnot INFO/DP se stanoví hranice 0,5 % a 99,5 % a SNP s hloubkou ležící mimo toto okno jsou ze sady vyloučeny. Tímto postupem se eliminují jak místa s podezřele nízkým pokrytím, tak pozice s extrémně vysokým pokrytím, aniž by docházelo k zásadnímu snížení celkového počtu identifikovaných SNP markerů.

### **# Filter SNPs by DP**

**# Extract and sort site DP values from the RPRS VCF**

```
bcftools query -f '%INFO/DP\n' "$VCF" \  
| awk 'tolower($1)!="nan" && $1!=".' \  
| sort -n > "$VALS"
```

```

# Compute 0.5 % and 99.5 % indices and cutpoints
N=$(wc -l < "$VALS")
lo_idx=$(awk -v N=$N 'BEGIN{lo=int(N*0.005); if(lo<1) lo=1; print lo}')
hi_idx=$(awk -v N=$N 'BEGIN{hi=int(N*0.995); if(hi<1) hi=1; if(hi>N) hi=N; print hi}')
read LO HI <<< "$(awk -v lo=$lo_idx -v hi=$hi_idx 'NR==lo{a=$1} NR==hi{b=$1} END{print a, b}' "$VALS)"

```

```

# Calculate how many sites below/above those cutpoints
low_rm=$(awk -v LO=$LO '$1<LO{c++} END{print c+0}' "$VALS")
high_rm=$(awk -v HI=$HI '$1>HI{c++} END{print c+0}' "$VALS")

```

```

echo "DP 0.5 % cutpoint = ${LO} (to drop: ${low_rm})"
echo "DP 99.5 % cutpoint = ${HI} (to drop: ${high_rm})"

```

```

# Apply the DP window [LO, HI]
bcftools view -i "INFO/DP>=${LO} && INFO/DP<=${HI}" -v snps -c 1 -Ou "$VCF" \
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \
| bcftools view -Oz -o $OUT

```

```
bcftools stats $OUT | grep -E 'number of records:|^TSTV'
```

```
tabix -f -p vcf $OUT
```

Pro parametr QUAL (Phred skóre kvality) se analogicky posuzuje vliv různých prahových hodnot na Ti/Tv a velikost SNP sady. V souladu s běžnou praxí se jako výchozí kritérium používá hranice  $QUAL \geq 50$ . Další možné filtry (např. ExcessHet) jsou testovány, avšak vzhledem k jejich minimálnímu dopadu na Ti/Tv a počet variant nejsou ve finálním filtračním schématu uplatněny.

### **# Filter SNPs by QUAL**

```

bcftools view -i 'QUAL>=50 || QUAL=="."' -v snps -c 1 -Ou $VCF \
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \
| bcftools view -Oz -o $OUT

```

```
bcftools stats $OUT | grep -E 'number of records:|^TSTV'
```

```
tabix -p vcf $OUT
```

V závěrečné fázi se obvykle aplikuje filtr založený na minoritní alelové frekvenci (MAF, Minor Allele Frequency). Pro několik kandidátních prahů (např. 0,01; 0,03; 0,05) se porovnává počet SNP a odpovídající poměr tranzic a transverzí a hodnotí se dopad na velikost a kvalitu výsledného SNP souboru. Cílem je nalézt kompromis mezi odstraněním extrémně vzácných alel, které bývají častěji zatíženy chybami a mají omezenou vypovídací hodnotu v asociačních analýzách, a zachováním dostatečné genetické variability. V praxi se pro GWAS často používají prahy v rozmezí  $MAF \geq 0,03-0,05$

### **# Applying MAF filtering**

```
bcftools view -i 'INFO/MAF>=0.03' -v snps -c 1 -Ou $VCF \  
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \  
| bcftools view -Oz -o $OUT1
```

```
bcftools stats $OUT1 | grep -E 'number of records:|^TSTV'
```

```
tabix -p vcf $OUT1
```

```
bcftools view -i 'INFO/MAF>=0.05' -v snps -c 1 -Ou $VCF \  
| bcftools +fill-tags -Ou -- -t AC,AN,AF,MAF,NS,F_MISSING \  
| bcftools view -Oz -o $OUT3
```

```
bcftools stats $OUT3 | grep -E 'number of records:|^TSTV'
```

```
tabix -p vcf $OUT3
```

## **Odhad struktury populace**

Pro analýzu populační struktury a následné asociační testy se nejprve připraví genotypová data ve formátu vhodném pro rychlé GWAS výpočty. Z finálního setu filtrovaných SNP sady se VCF soubor převede do binárního formátu nástrojem PLINK (soubory BED/BIM/FAM). V rámci převodu se standardizují identifikátory snp a vzorků tak, aby byly jednoznačné a kompatibilní s používanými GWAS nástroji.

### **# Convert VCF file into a fast GWAS format**

```
plink2 --vcf ${DIR}/variants_filtering/maf5.vcf.gz --double-id --set-missing-var-ids  
@:#:\$r:\$a --make-bed --out ${DIR}/variants_filtering/maf5.qc
```

Následuje základní kontrola kvality na úrovni vzorků. Pro každý genotyp se hodnotí podíl chybějících genotypů (per-sample missingness) a míra heterozygotnosti. Vzorky s nadměrným podílem chybějících dat nebo s extrémní heterozygotností (odlehle hodnoty mimo očekávaný interval, např. více než několik směrodatných odchylek od průměru) jsou považovány za problematické (nekvalitní DNA/RNA, technické artefakty) a z další analýzy se zpravidla vylučují. Tím se zajistí, že populační struktura a asociace budou odhadovány z biologicky i technicky spolehlivých vzorků.

### **# Validate per-sample missingness**

```
plink2 --bfile ${DIR}/variants_filtering/maf5.qc --missing sample-only --out  
${DIR}/variants_filtering/maf5.miss
```

```
awk 'NR>1 && $5>0.10 {print $1,$5}' ${DIR}/variants_filtering/maf5.miss.smiss
```

### **# Validate heterozygosity**

```
plink2 --bfile ${DIR}/variants_filtering/maf5.qc --het --out  
${DIR}/variants_filtering/maf5.het
```

```
awk 'NR>1{print $1,($3-$4)/$5}' ${DIR}/variants_filtering/maf5.het.het | awk  
'sqrt($2*$2)>0.3 {print $0}'
```

Pro odhad populační struktury se dále provede protřídění SNP na základě vazbové nerovnováhy (LD - Linkage Disequilibrium). Pomocí funkce indep-pairwise v PLINK se v klouzavém okně odstraňují blízko ležící vysoce korelované SNP. Výsledkem je podmnožina zhruba nezávislých markerů, která je vhodná pro odhad hlavních komponent a příbuznosti, aniž by byla analýza zkreslena bloky shluků SNP.

### **# Pruning the SNPs based on Linkage Disequilibrium**

```
plink2 --bfile ${DIR}/variants_filtering/maf5.qc --indep-pairwise 50 5 0.2 --threads $TH --out ${DIR}/variants_filtering/maf5.prune
```

```
plink2 --bfile ${DIR}/variants_filtering/maf5.qc --extract ${DIR}/variants_filtering/maf5.prune.prune.in --threads $TH --make-bed --out ${DIR}/variants_filtering/maf5.pruned
```

Na získaném setu SNP markerů se následně provede analýza hlavních komponent (Principal Component Analysis, PCA). Pro GWAS se typicky využívají první 3–5 hlavních komponent, které zachycují hlavní osu/osy populační struktury (subpopulace, geografická podobnost, šlechtitelské linie). Tyto komponenty se uloží do kovariačních souborů a v následujících asociačních modelech slouží jako fixní vlivy, které korigují za populační strukturu a omezují výskyt falešně pozitivních asociací.

### **# Perform Principal Component Analysis (PCA) from the pruned set**

```
plink2 --bfile ${DIR}/variants_filtering/maf5.pruned --pca 5 --threads $TH --out ${DIR}/variants_filtering/maf5.pca
```

Současně se ze stejného setu SNP odhaduje matice příbuznosti (kinship matrix, K). K tomu lze využít například program GEMMA, který z binárního PLINK formátu spočítá genomickou matici příbuznosti na základě sdílených alel mezi genotypy. Tato matice se používá jako náhodný efekt v lineárním smíšeném modelu (LMM), čímž se modeluje genetická příbuznost mezi jedinci a dále se snižuje riziko falešných asociací.

### **# Calculate Kinship matrix (K) from the pruned set**

```
gemma -bfile ${DIR}/variants_filtering/maf5.pruned -gk 1 -p ${DIR}/variants_filtering/dummy.pheno -o kinship_matrix
```

Finální GWAS se poté provádí v nástroji GEMMA. Jako vstup slouží původní set SNP, který je získán na konci filtračního schématu. Tato sada je v PLINK formátu. Dále se použije vypočtená matice podobnosti K a soubor z PCA, který zahrnuje 3,

4 nebo 5 komponent. Fenotypová data (např. index napadení) jsou zadána jako samostatný soubor. Lineární smíšený model testuje asociaci každého SNP s fenotypem při současném zohlednění populační struktury (PC) a příbuznosti (K).

### **# Run GWAS by GEMMA**

```
gemma -bfile ${DIR}/variants_filtering/maf5.qc \  
-k ${DIR}/variants_filtering/kinship.matrix \  
-c ${DIR}/variants_filtering/covariate.pc3 \  
-p ${DIR}/phenotype/phl.di \  
-lmm 4 -o phl
```

Výstupem jsou standardní asociační statistiky (beta, SE, p-hodnoty), které se dále zpracují a vizualizují do podoby Manhattan grafů v samostatných skriptech v prostředí R viz níže.

## **Praktické použití metodiky**

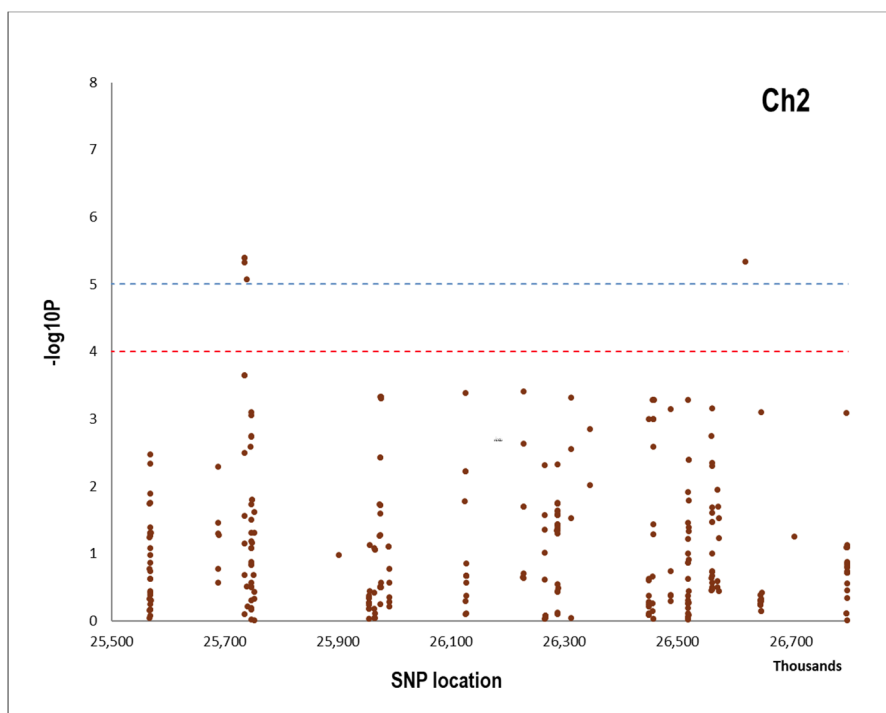
Metodika byla ověřena na sortimentu 72 genových zdrojů bobu polního (registrované odrůdy i šlechtitelské materiály). U každého genotypu byla hodnocena citlivost k antraknóze na sklizených semenech pomocí devítibodové škály (1 = silně napadená semena, 9 = zdravá semena). Pro každý genotyp byl z více rostlin vypočten průměrný index napadení, který následně sloužil jako kvantitativní fenotyp pro asociační analýzu.

Genotypová data ve formě SNP markerů na genových sekvencích byla po standardní kvalitativní filtraci použita k analýze asociací mezi markery a fenotypem napadení semen antraknózou. Výstupem analýzy jsou pro každý SNP odhad efektu na sledovaný znak a odpovídající P-hodnota, vizualizovaná pomocí Manhattanových grafů.

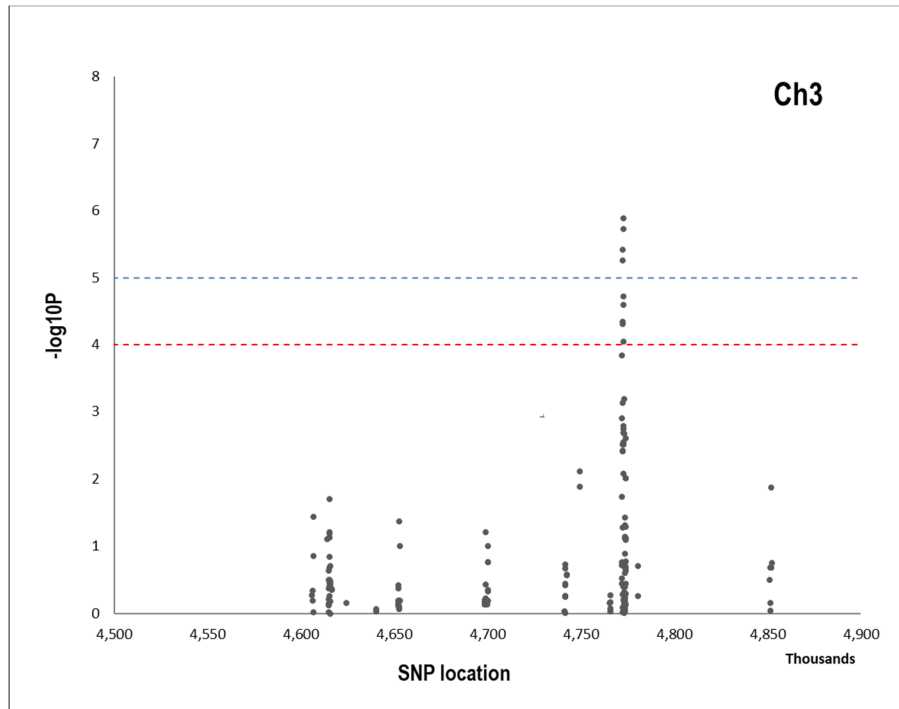
Tímto postupem lze identifikovat genomové regiony, v nichž se nacházejí polymorfismy významně asociované s nižší mírou napadení antraknózou. Tyto regiony představují kandidátní lokusy rezistence a obsahují geny vhodné pro detailnější funkční anotaci. Po nezávislém potvrzení v dalším materiálu mohou identifikované SNP sloužit jako markery pro markerem asistovanou selekci, podporu rozhodování při volbě rodičů křížení a pro sledování úrovně rezistence v šlechtitelských liniích.

Konkrétně prahovou hodnotu  $-\log_{10}(P)$  5 překročily SNP markery ve třech oblastech genomu. Na následujícím Manhattan grafu vidíme první potenciálně rezistentní

lokus na chromozomu 2. V levé části grafu, kolem 25,7 Mb, je patrný shluk SNP s hodnotami  $-\log_{10}(P)$  kolem 5–5,5. Několik markerů zde překračuje obě prahové hodnoty. To odpovídá potenciálnímu kandidátnímu lokusu – signál je nesen skupinou sousedních SNP v LD, nikoli jediným markerem.



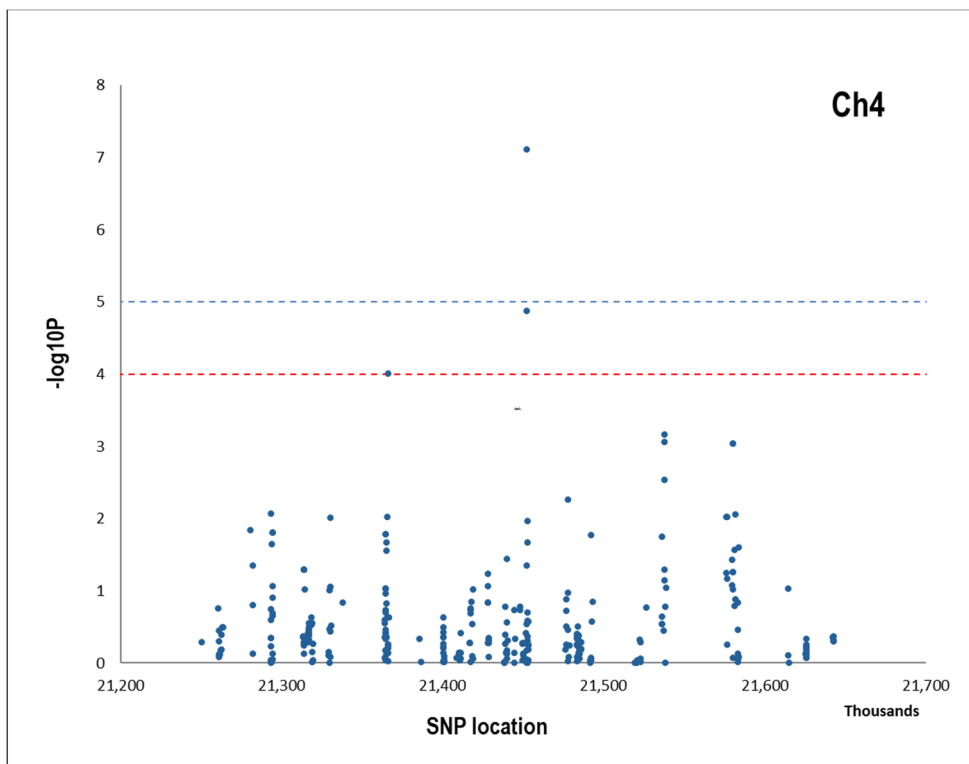
Druhou, z hlediska výsledků nejvýznamnější oblast představuje kandidátní lokus na chromozomu 3. Manhattanův graf ukazuje jeden dominantní vrchol asociace. V intervalu přibližně 4,74–4,78 Mb je patrný kompaktní shluk SNP markerů, z nichž řada překračuje jak práh  $-\log_{10}(P) = 4$ , tak  $-\log_{10}(P) = 5$ . Tento soubor vzájemně blízkých markerů ve výrazné vazbové nerovnováze (LD) ukazuje na úzce vymezenou genomovou oblast, která představuje nejzajímavější kandidátní QTL pro rezistenci k hnědé skvrnitosti (antraknóze) bobu polního.



Detailní výsledky asociace v kandidátní oblasti na chromozomu 3 shrnuje tabulka níže. Uvedeny jsou jednotlivé SNP v intervalu přibližně 4,76–4,78 Mb, včetně jejich fyzické pozice (sloupce chr, rs, ps), nukleotidového polymorfismu (allele1, allele0), frekvence alternativní alely (af), odhadu efektu na znak ( $\beta$ ) a směrodatné chyby (SE). Poslední tři sloupce obsahují  $-\log_{10}(P)$  pro tři různé testy implementované v programu GEMMA (LRT, Waldův test, score test); zeleně jsou zvýrazněny SNP s nejvyšší signifikancí.

V kandidátní oblasti je patrný souvislý shluk desítek SNP se zvýšenou asociací, přičemž nejvýznamnější markery dosahují hodnot  $-\log_{10}(P)$  okolo 5–6 ve všech třech testech. Tyto polymorfismy mají střední až vyšší frekvenci alternativní alely (zhruba 0,2–0,7) a velmi podobné odhady efektu i p-hodnot, což odpovídá jednomu genetickému bloku v nerovnováze vazby (LD). Tabulka tak ukazuje na úzký kandidátní region obsahující pravděpodobný QTL pro rezistenci k antraknóze.

chr	rs	ps	allele1	allele0	af	beta	se	log10(p_lrt)	log10(p_wald)	log10(p_score)
ch3	ch3:4765845:G:A	4765845	A	G	0.141	7.20E-01	7.76E+00	0.0354276	0.0331645	0.0351999
ch3	ch3:4765851:T:A	4765851	A	T	0.141	7.20E-01	7.76E+00	0.0354276	0.0331645	0.0351999
ch3	ch3:4765853:C:T	4765853	T	C	0.163	1.38E+00	7.38E+00	0.0741678	0.0692686	0.0736471
ch3	ch3:4766007:T:C	4766007	C	T	0.043	-5.27E+00	1.37E+01	0.16492	0.153319	0.163409
ch3	ch3:4772034:G:C	4772034	C	G	0.098	8.94E+00	9.14E+00	0.519162	0.477091	0.506715
ch3	ch3:4772036:T:C	4772036	C	T	0.62	1.73E+01	5.36E+00	2.89406	2.60161	2.39545
ch3	ch3:4772045:T:C	4772045	C	T	0.739	1.94E+01	4.98E+00	3.8443	3.44546	2.9728
ch3	ch3:4772093:G:C	4772093	C	G	0.109	1.10E+01	8.89E+00	0.716454	0.655685	0.691531
ch3	ch3:4772096:T:A	4772096	A	T	0.196	1.98E+01	8.62E+00	1.73672	1.57084	1.56403
ch3	ch3:4772111:G:A	4772111	A	G	0.065	-2.65E+01	2.03E+01	0.76715	0.701447	0.738199
ch3	ch3:4772120:C:T	4772120	T	C	0.065	9.82E+00	1.12E+01	0.447497	0.411977	0.438371
ch3	ch3:4772185:C:G	4772185	G	C	0.152	6.32E+00	1.04E+01	0.282675	0.261555	0.278956
ch3	ch3:4772252:T:C	4772252	C	T	0.12	6.37E-01	1.11E+01	0.0215966	0.0202352	0.0214613
ch3	ch3:4772258:A:C	4772258	C	A	0.62	2.18E+01	5.16E+00	4.34289	3.88771	3.24369
ch3	ch3:4772308:A:G	4772308	G	A	0.598	1.78E+01	5.27E+00	3.12496	2.8068	2.54371
ch3	ch3:4772321:T:G	4772321	G	T	0.674	1.52E+01	5.31E+00	2.42054	2.18041	2.07386
ch3	ch3:4772452:G:A	4772452	A	G	0.522	1.48E+01	5.17E+00	2.40995	2.17099	2.06639
ch3	ch3:4772464:C:G	4772464	G	C	0.609	1.02E+01	5.46E+00	1.27095	1.15427	1.18307
ch3	ch3:4772536:T:C	4772536	C	T	0.75	2.15E+01	5.12E+00	4.31586	3.86374	3.22952
ch3	ch3:4772544:G:A	4772544	A	G	0.696	2.30E+01	4.77E+00	5.25572	4.69679	3.6905
ch3	ch3:4772551:G:A	4772551	A	G	0.674	2.41E+01	4.90E+00	5.41827	4.8408	3.76402
ch3	ch3:4772554:C:A	4772554	A	C	0.065	2.74E+00	1.31E+01	0.0837037	0.0781327	0.083101
ch3	ch3:4772563:T:G	4772563	G	T	0.554	1.59E+01	5.43E+00	2.50891	2.25906	2.13574
ch3	ch3:4772721:T:A	4772721	A	T	0.554	1.59E+01	5.43E+00	2.50891	2.25906	2.13574
ch3	ch3:4772724:G:A	4772724	A	G	0.685	1.67E+01	5.37E+00	2.73507	2.46025	2.29017
ch3	ch3:4772727:T:C	4772727	C	T	0.554	1.59E+01	5.43E+00	2.50891	2.25906	2.13574
ch3	ch3:4772739:A:G	4772739	G	A	0.163	3.90E+00	8.39E+00	0.205503	0.190711	0.203365
ch3	ch3:4772742:T:C	4772742	C	T	0.674	2.00E+01	4.97E+00	4.05317	3.63076	3.08877
ch3	ch3:4772751:T:A	4772751	A	T	0.598	1.34E+01	5.21E+00	2.07247	1.87037	1.8214
ch3	ch3:4772799:C:T	4772799	T	C	0.717	2.35E+01	4.60E+00	5.72417	5.11177	3.89784
ch3	ch3:4772814:C:T	4772814	T	C	0.685	2.47E+01	4.74E+00	5.88885	5.25761	3.96752
ch3	ch3:4772823:G:A	4772823	A	G	0.728	2.24E+01	5.09E+00	4.5969	4.11292	3.3741
ch3	ch3:4772834:C:A	4772834	A	C	0.761	2.23E+01	4.98E+00	4.7252	4.22665	3.43813
ch3	ch3:4772840:T:G	4772840	G	T	0.217	5.84E+00	7.31E+00	0.398952	0.367779	0.391731
ch3	ch3:4772874:G:A	4772874	A	G	0.761	1.79E+01	6.05E+00	2.54103	2.28764	2.15801
ch3	ch3:4772901:T:C	4772901	C	T	0.75	1.97E+01	6.26E+00	2.78441	2.50412	2.32313
ch3	ch3:4772904:C:T	4772904	T	C	0.772	1.83E+01	5.97E+00	2.68271	2.41368	2.25491
ch3	ch3:4773007:G:A	4773007	A	G	0.793	2.02E+01	5.90E+00	3.17702	2.85305	2.5764
ch3	ch3:4773029:C:T	4773029	T	C	0.065	4.25E+00	1.40E+01	0.125725	0.117098	0.124704
ch3	ch3:4773034:A:T	4773034	T	A	0.207	-1.68E-01	7.76E+00	0.00800086	0.00750339	0.00795172
ch3	ch3:4773064:A:G	4773064	G	A	0.054	-1.05E+01	1.70E+01	0.289232	0.26756	0.285353
ch3	ch3:4773067:G:A	4773067	A	G	0.772	1.87E+01	6.12E+00	2.66967	2.40209	2.24608
ch3	ch3:4773101:C:G	4773101	G	C	0.054	-1.05E+01	1.70E+01	0.289232	0.26756	0.285353
ch3	ch3:4773102:C:A	4773102	A	C	0.076	-6.00E+00	1.63E+01	0.157395	0.146373	0.155987
ch3	ch3:4773106:G:T	4773106	T	G	0.217	-1.40E+00	7.79E+00	0.0710559	0.066374	0.0705609
ch3	ch3:4773128:T:A	4773128	A	T	0.207	-1.57E+00	7.62E+00	0.0823087	0.0768365	0.0817183
ch3	ch3:4773175:T:A	4773175	A	T	0.065	-9.71E+00	1.46E+01	0.317744	0.293651	0.313119
ch3	ch3:4773208:T:C	4773208	C	T	0.076	-1.39E+01	1.15E+01	0.6908	0.63251	0.667784
ch3	ch3:4773223:T:C	4773223	C	T	0.065	-9.71E+00	1.46E+01	0.317744	0.293651	0.313119
ch3	ch3:4773229:T:C	4773229	C	T	0.228	-4.20E+00	7.24E+00	0.267657	0.247793	0.264289
ch3	ch3:4773390:A:G	4773390	G	A	0.054	-1.05E+01	1.70E+01	0.289232	0.26756	0.285353
ch3	ch3:4773394:T:A	4773394	A	T	0.043	-4.42E+01	2.33E+01	1.31264	1.19162	1.21834
ch3	ch3:4773454:T:C	4773454	C	T	0.087	5.20E+00	1.06E+01	0.218395	0.202569	0.216031



Na chromozomu 4 byl zaznamenán poslední ze tří významných signálů. Manhattanův graf ukazuje převážně rozptýlené asociace s mírnou až střední významností, v centrální části oblasti ( $\approx 21,4\text{--}21,5$  Mb) vystupuje jeden SNP, který výrazně překračuje jak hodnotu  $-\log_{10}P = 4$  (červená), tak i hranici  $-\log_{10}P = 5$  (modrá) a dosahuje hodnoty 7,1. Vzhledem k tomu, že v těsném okolí tohoto markeru není patrný významný shluk dalších vysoce významných SNP, je třeba tento signál interpretovat opatrněji.

Souhrnně tato sada analýz ukazuje, že navržený postup – od extrakce RNA přes specifickou filtrovou sadu SNP až po GWAS ve smíšeném lineárním modelu – umožňuje u bobu polního identifikovat jak kandidátní QTL podporované shlukem sousedních markerů, tak i jednotlivé vysoce významné SNP, které už v této podobě mohou sloužit k marker-asistované selekci.

## Srovnání novosti postupů

Předkládanou metodiku s názvem “Metodika detekce genů odolnosti vůči antraknóze u luskovin“ lze hodnotit jako novou metodiku, neboť v současné době není ve šlechtitelské praxi k dispozici ucelená metodika popisující identifikaci rezistence k antraknóze u luskovin využívající moderní metody asociativní genomiky nebo transkriptomiky. Dosud dostupné informace jsou jen dílčí, a dokonce ani ve vědeckých publikacích není uvedeno jakékoli použití asociativní transkriptomiky u luskovin.

Identifikace genomových oblastí podmiňujících odolnost k antraknóze luskovin pomocí znalosti fenotypových dat, sekvenování transkriptomu a asociační analýzy založené na exonických SNP není v běžné šlechtitelské praxi dosud standardizovaným postupem. Většina programů stále vychází převážně z fenotypové selekce v polních či skleníkových pokusech, případně z omezeného počtu jednoduchých markerů, a to spíše pro jiné choroby. Předkládaná metodika představuje integrovaný přístup asociativní transkriptomiky, kdy se polymorfismy využitelné pro asociativní analýzu získávají přímo z přepisu (RNA-seq) a následného mapování na referenční genom.

Limitem takového přístupu je především potřeba kvalitních transkriptomických dat (standardizovaný odběr pletiv, izolace RNA, příprava knihoven, sekvenování) a dostupnost výpočetního zázemí a bioinformatického know-how pro zpracování (mapování, volání variant, filtrace SNP a GWAS ve smíšeném lineárním modelu). Metodika tyto bariéry částečně překonává tím, že podrobně popisuje celý postup – od pěstování rostlin, přes přípravu RNA-seq knihoven a zpracování sekvenačních dat až po konstrukci panelu SNP markerů a vlastní asociační analýzu. Výstupem je nejen identifikace kandidátních genomových oblastí a SNP vhodných pro markerem asistovanou selekci, ale také přenositelný postup, který lze po úpravě aplikovat i na další plodiny a jiné choroby.

## Popis uplatnění metodiky

Metodika je určena pro specializovaná výzkumná a aplikační pracoviště, zejména šlechtitelské a molekulárně-biologické laboratoře, které mají k dispozici základní vybavení pro práci s RNA, přípravu sekvenačních knihoven a bioinformatické zpracování NGS dat. V úvodní části metodika stručně shrnuje biologii patogena *Ascochyta fabae*, jeho význam jako původce antraknózy luskovin a hlavní důsledky pro pěstitelskou praxi a šlechtění. Následující kapitoly krok za krokem popisují praktické postupy od standardizovaného odběru pletiv a izolace RNA, přes přípravu a sekvenování RNA-seq knihoven až po bioinformatickou analýzu, která vede k

identifikaci SNP markerů a jejich využití marker asistované selekci k antraknóze u luskovin.

Navržený postup byl vyvinut a optimalizován s ohledem na potřeby šlechtitelských programů zaměřených na selekci genotypů s vyšší odolností či tolerancí k antraknóze. Metodika ukazuje, jak propojit dostupná fenotypová data (stupeň napadení antraknózou) s transkriptomickými daty a jak z nich odvodit oblasti a kandidátní lokusy spojené s nižší mírou napadení. Takto identifikované SNP markery mohou po nezávislém ověření sloužit jako nástroj pro markerem asistovanou selekci, podporu rozhodování při volbě rodičovských komponent v kříženích a monitorování úrovně rezistence ve šlechtitelských liniích.

Praktické uplatnění spočívá zejména v zavedení transkriptomicky podložených markerů do rutinní selekce, ve zpřesnění hodnocení genových zdrojů a ve strategickém plánování křížení s cílem zvýšit odolnost porostů k antraknóze a stabilizovat výnos v podmínkách infekčního tlaku patogena.

## **Ekonomické aspekty**

Analýzy popsané v této metodice mají výrazný ekonomický význam z pohledu šlechtitelských pracovišť. Vzhledem k závažnosti antraknózy luskovin, předpokládanému nárůstu infekčního tlaku a omezené dostupnosti odolných odrůd se šlechtění na odolnost či toleranci k tomuto biotickému stresu stává vysoce aktuálním cílem.

Pro úspěšné šlechtění je klíčová nejen obecná znalost genetického základu rezistence, ale také schopnost identifikovat konkrétní genomové oblasti spojené s nižší mírou napadení a převést tyto informace do podoby využitelných molekulárních markerů. Takto odvozené SNP markery pak mohou zásadně zefektivnit selekci rezistentních linií, zkrátit šlechtitelský cyklus a snížit náklady spojené s fenotypovým hodnocením rostlin.

Navržený metodický postup umožňuje na základě RNA-seq a následné asociační analýzy identifikovat genomové oblasti se SNP markery spojené s napadením antraknózou. Identifikované asociované polymorfismy mohou být dále převáděny do jednodušších laboratorních testů vhodných pro rutinní využití ve šlechtitelské praxi (PCR, sekvenování amplikonů).

Pro provedení analýz je nutné mít k dispozici soubor genetických zdrojů vybrané luskoviny, kvalitní fenotypová data, molekulárně-biologickou a bioinformatickou laboratoř. Při identifikaci nových genů/markerů je však nutné počítat s počátečními náklady na sekvenování, která představuje částku 4800 Kč na vzorek při zpracování aspoň 24 genotypů. Do ceny je ale třeba zahrnout zhruba týdenní práci laboranta a bioinformatika.

Vlastní detekce SNP markeru je založena na rutinní PCR analýze a sekvenování amplikonů, případně při využití snp-specifických markerů pouze na PCR analýze

v ceně cca 250 Kč na marker a genotyp. Pomocí tohoto postupu lze zefektivnit šlechtění a selekci genotypů, které nesou kandidátní geny a mají predispozici pro vyšší odolnost/toleranci k antraknóze.

Pro šlechtitelská pracoviště představuje navržený metodický postup významný přínos, protože umožňuje získat nové odrůdy s odolností nebo tolerancí vůči antraknóze bobu.

## Seznam použité literatury

Adhikari, K. N., Khazaei, H., Ghaouti, L., Maalouf, F., Vandenberg, A., Link, W., et al. (2021). Conventional and molecular breeding tools for accelerating genetic gain in faba bean (*Vicia faba* L.). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.744259

Birch, A. N. E., Tithecott, M. T., and Bisby, F. A. (1985). *Vicia johannis* and wild relatives of the faba bean: a taxonomic study. *Econ. Bot.* 39, 177–190. doi: 10.1007/BF02907843

Díaz-Ruiz, R., Satovic, Z., Ávila, C. M., Alfaro, C. M., Gutierrez, M. V., Torres, A. M., et al. (2009). Confirmation of QTLs controlling *Ascochyta fabae* resistance in different generations of faba bean (*Vicia faba* L.). *Crop Pasture Sci.* 60, 353–361. doi: 10.1071/CP08190

Gorfu, D. (1996). Morphological, cultural and pathogenic variability among nine isolates of *Botrytis fabae* from Ethiopia. *FABIS Newsl.* 38-39, 37–41.

Hampton, G. (1980). The significance of *Ascochyta fabae* in broad beans in the Manawatu, and methods for its control. *N. Z. J. Crop Hortic. Sci.* 8, 305–308. doi: 10.1080/03015521.1980.10426279

Karkanis, A., Ntatsi, G., Lapse, L., Fernández, J. A., Vågen, I. M., Rewald, B., et al. (2018). Faba bean cultivation – revealing novel managing practices for more sustainable and competitive European cropping systems. *Front. Plant Sci.* 9:1115. doi: 10.3389/fpls.2018.01115

Sudheesh, S., Kimber, R. B. E., Braich, S., Forster, J. W., Paull, J. G., and Kaur, S. (2019). Construction of an integrated genetic linkage map and detection of quantitative trait loci for ascochyta blight resistance in faba bean (*Vicia faba* L.). *Euphytica* 215:42. doi: 10.1007/s10681-019-2365-x

## Seznam publikací předcházející metodice

Hejna O., Havlickova L., He Z., Bancroft I., Curn V. (2019): Analysing the genetic architecture of clubroot resistance variation in *Brassica napus* by associative transcriptomics. *Molecular Breeding*, 39, 112.

Hejna, O., Kopecký, P., Čurn, V. (2019): Hledání genetických zdrojů rezistence k nádorovitosti košťálovin. *Úroda* 12, roč. LXVII, 2019, vědecká příloha, s. 121 – 125. ISSN 0139-6013.

Název: Hejna O. a kol. (2025): *Metodika detekce genů odolnosti vůči antraknóze u luskovin*

Autorský kolektiv: Ing. Mgr. Ondřej Hejna, Ph.D.  
Ing. Kristýna Beranová  
Ing. Eva Jozová Ph.D.  
prof. Ing. Vladislav Čurn, Ph.D.

Vydal: Jihočeská univerzita v Českých Budějovicích  
Fakulta zemědělská a technologická  
Studentská 1668  
370 05 České Budějovice

Vydáno bez jazykové úpravy

Metodika byla schválena Ministerstvem zemědělství ČR, Odborem rostlinných komodit – číslo Osvědčení MZE-5642/2026-13123, jako uplatněná metodika s doporučením pro její využití v zemědělské praxi.

Kontakt na autory: [hejna@fzt.jcu.cz](mailto:hejna@fzt.jcu.cz)

ISBN: 978-80-7694-145-8

ISBN 978-80-7694-145-8

