



Podrobná analýza štruktúry populácie na základe genomických údajov a kvantifikácia vplyvu selekcie na genóm hospodárskych zvierat

Modul no. 2: Konzervovanie a udržateľné využívanie genetických zdrojov

Nina Moravčíková, Radovan Kasarda
Slovenská poľnohospodárska univerzita v Nitre
Fakulta agrobiológie a potravinových zdrojov

Obsah prednášky

- Kontrola kvality genomických údajov
- Prístupy a nástroje pre analýzu populačnej štruktúry
- Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat
- Funkčná anotácia oblastí signifikantne ovplyvnených selekčným tlakom



Kontrola kvality genomických údajov

SNP dáta (genotypovanie pomocou SNP čipov)

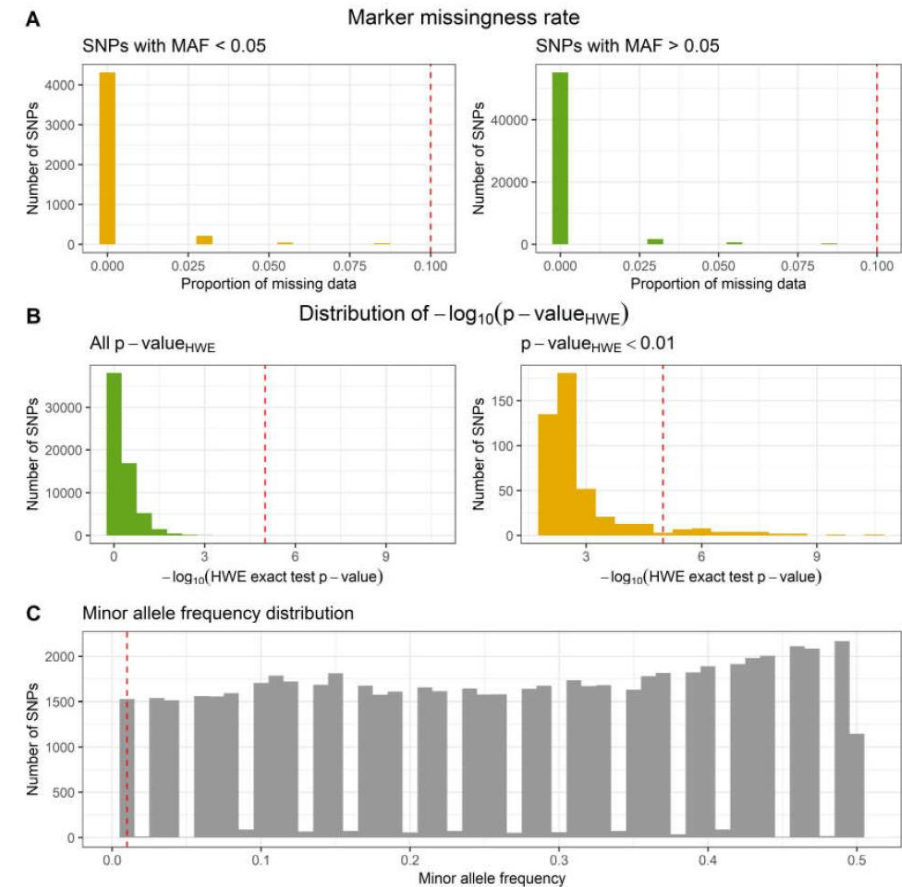
- QC je kľúčovým krokom na zabezpečenie presnosti a spoľahlivosti výsledkov následných analýz
- nesprávne alebo nekvalitné údaje môžu viesť k chybám v analýzach a interpretácii výsledkov
- ukazovatele kvality dát – úspešnosť genotypovania SNP markerov celkovo a v rámci jedincov, frekvencia minoritnej alely, odchýlka od Hardy-Weinbergovej rovnováhy, miera väzbovej nerovnováhy ...
- výber ukazovateľov použitých v QC závisí od cieľa/ typu následných analýz



Kontrola kvality genomických údajov

SNP dáta (genotypovanie pomocou SNP čipov)

- Štandardná QC v prípade analýzy populačnej štruktúry:
 - úspešnosť genotypovania SNP markerov celkovo – min. 90%
 - úspešnosť genotypovania SNP markerov v rámci jedincov – min. 90%
 - frekvencia minoritnej alely (MAF) – min. 1% v populácii
 - Hardy-Weinbergova rovnováha - 1×10^{-6}
- Softvérové nástroje: napr. PLINK



Obr. 1: Grafická vizualizácia kontroly kvality SNP dát (Moravčíková et al., 2020)



Prístupy a nástroje pre analýzu populačnej štruktúry

SNP dáta (genotypovanie pomocou SNP čipov)

Štúdium populačnej štruktúry pomocou genomických dát umožňuje:

- analyzovať genetickú diferenciáciu v rámci a medzi populáciami,
- hodnotiť mieru ich genetického premiešania ako aj zmeny v ich genofonde, ktoré vznikli napr. vplyvom selekcie, migrácie alebo genetického driftu,
- vyhodnotiť genomickú maticu príbuznosti a optimalizovať pripravovacie plány



Co-funded by
the European Union

Prístupy a nástroje pre analýzu populačnej štruktúry

SNP dáta (genotypovanie pomocou SNP čipov)

Najčastejšie typy analýz:

- výpočet Wrightovho F_{ST} indexu,
- výpočet genetických vzdialeností a matíc príbuznosti
- analýza hlavných komponentov
- diskriminačná analýza hlavných komponentov
- bayesovská analýza genetického premiešania a toku génov medzi populáciami
- konštrukcia fylogenetických stromov a genetických sietí



Co-funded by
the European Union

Prístupy a nástroje pre analýzu populačnej štruktúry

Wrightov F_{ST} index

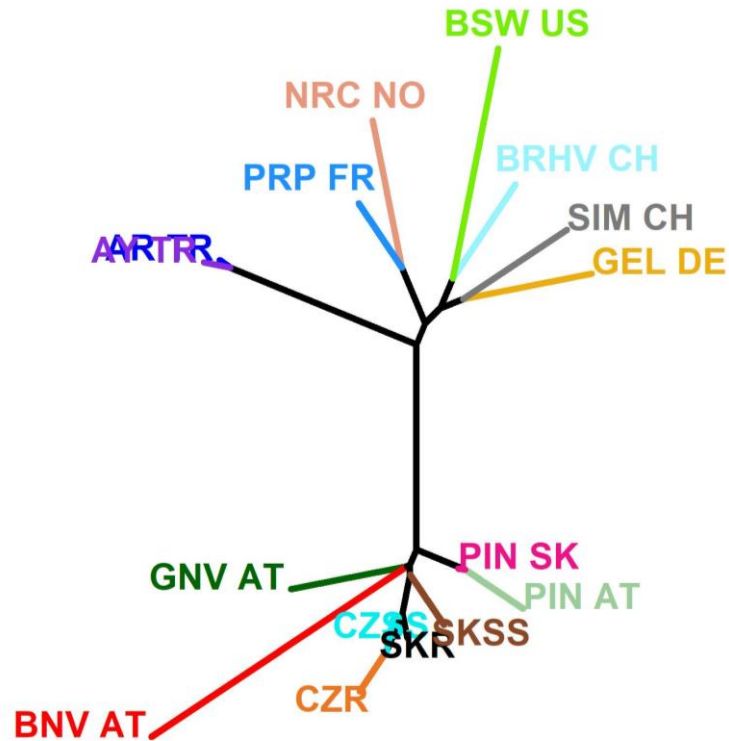
- je jedným z najčastejšie používaných ukazovateľov na stanovenie miery genetickej diferenciácie medzi populáciami
- jeho hodnota sa pohybuje od 0 (populácie sú geneticky identické) po 1 (populácie sú geneticky úplne diferencované)
- výhody: jednoduchosť interpretácie, rýchla výpočtová metóda pre detekciu diferenciácie medzi populáciami
- nevýhody: nie je možné ho využiť na individuálnej úrovni, nižšia spoľahlivosť v populáciách s nízkou úrovňou diverzity



Prístupy a nástroje pre analýzu populačnej štruktúry

Wrightov F_{ST} index

- Softvérové nástroje: Arlequin, Genepop a GenAlEx (limitovaný počet SNP markerov), R (napr. balíky StAMPP)



Obr. 2: Dendrogram zostrojený na základe F_{ST} matice znázorňujúci genetické vzťahy medzi 16 plemenami hovädzieho dobytku (nepublikované výsledky)

AR_TR - Anatolian Red, AY_TR - Anatolian Yellow, BNV_AT - Braunvieh (Austria), BNV_CH - Braunvieh (Switzerland), BSW_US - Brown Swiss, CZR - Czech Red, CZSS - Czech Spotted, GEL_DE - Gelbvieh, GNV_AT - Tyrol Grey, NRC_NO - Norwegian Red, PIN_AT - Pinzgau (Austria), PIN_SK - Slovak Pinzgau (Slovakia), PRP_FR - French Red Pied, SIM_CH - Simmental, SKR - Slovak Red, SKSS - Slovak Spotted

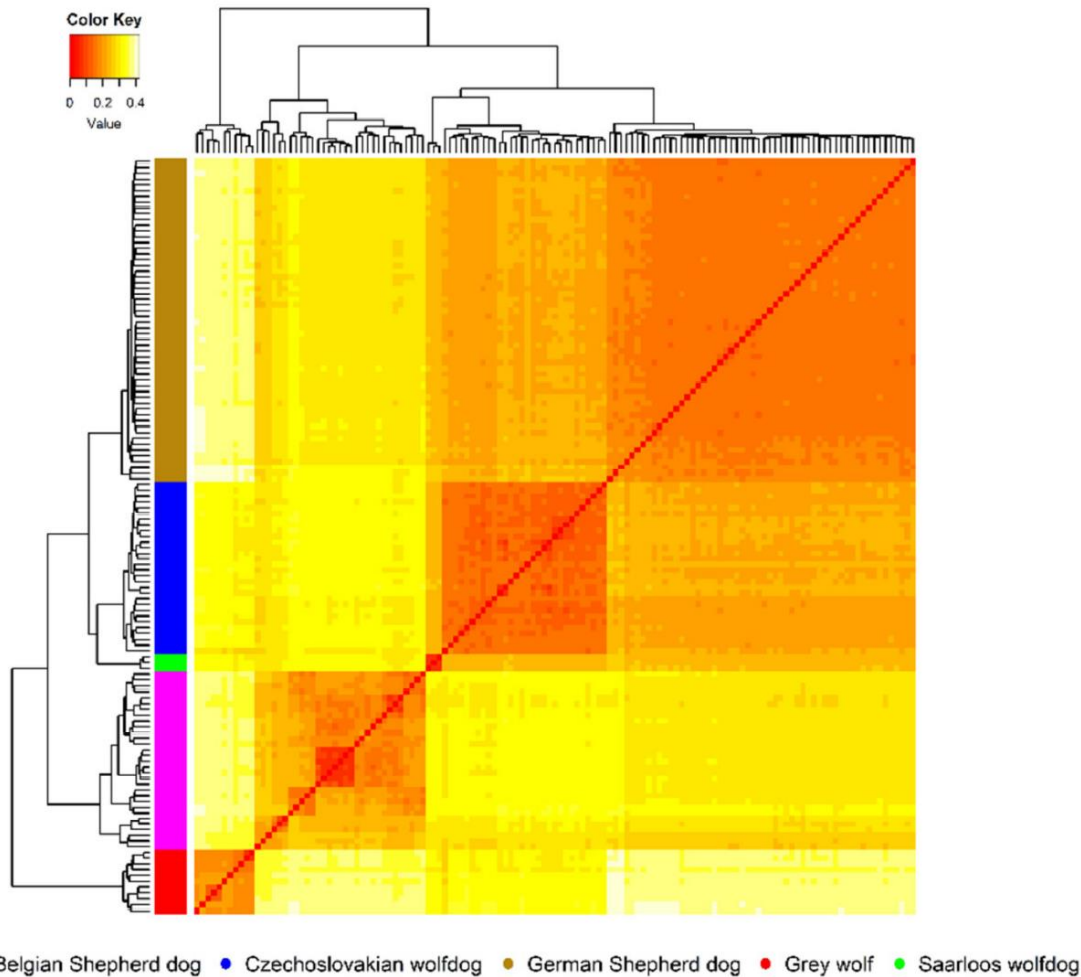
Prístupy a nástroje pre analýzu populačnej štruktúry

Maticie príbuznosti

- vyjadrujú genetické podobnosti a príbuznosť medzi jedincami v rámci populácie
- každý prvok matice predstavuje mieru genetickej podobnosti medzi dvojicou jedincov
- počítajú sa najčastejšie základe frekvencie alel, pričom samotný výpočet môže byť založený na rôznych prístupoch, napr. na výpočte IBD (identita podľa pôvodu) matice či Neiových genetických vzdialeností
- výhody: pomerne presné odhady príbuznosti, vhodné na štúdium vnútro populačných vzťahov
- nevýhody: v prípade veľkých databáz je výpočet časovo náročný



Prístupy a nástroje pre analýzu populačnej štruktúry



Obr. 3: Vnútro populačná genetická štruktúra v rámci 5 plemien psov odvodená z matice Neiových genetických vzdialeností (Moravčíková et al., 2021)

Maticie príbuznosti

- Softvérové nástroje: PLINK, R (napr. balík StAMPP)



Co-funded by
the European Union

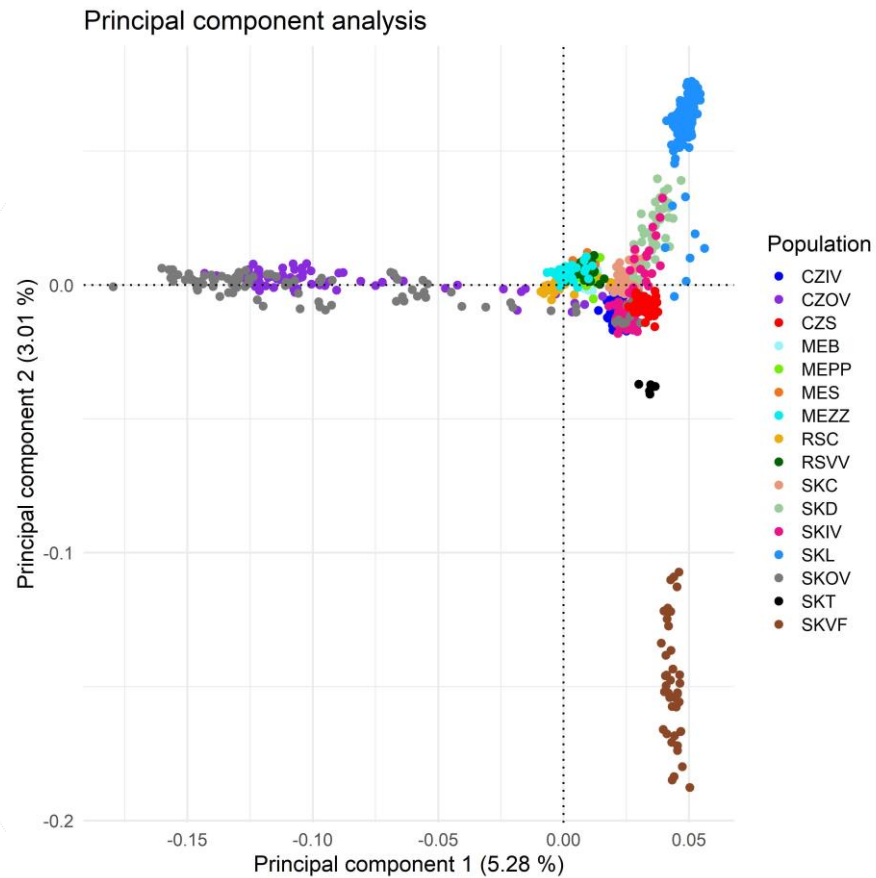
Prístupy a nástroje pre analýzu populačnej štruktúry

Analýza hlavných komponentov (PCA)

- viacrozmerná štatistická metóda, ktorá dekomponuje kovariančnú maticu genetických dát a extrahuje hlavné komponenty, ktoré reflektujú variabilitu dát v súbore
- prvé dva hlavné komponenty zachytávajú zvyčajne najväčší podiel variability
- poskytuje základné informácie o genetickej štruktúre, čo je užitočné pri testovaní databáz s veľkým počtom jedincov
- výhody: časovo nenáročná metóda na zhodnotenie stavu genetickej diferenciácie, jednoduchá a dobre interpretovateľná vizualizácia
- nevýhody: nízka citlivosť s pohľadom testovania genetického premiešania



Prístupy a nástroje pre analýzu populačnej štruktúry



Obr. 4: Genetická diferenciácia v rámci 16 plemien oviec na základe vizualizácie prvého a druhého hlavného komponenta PCA analýzy (nepublikované výsledky)

Analýza hlavných komponentov (PCA)

- Softvérové nástroje: PLINK, R (napr. balík adegenet)

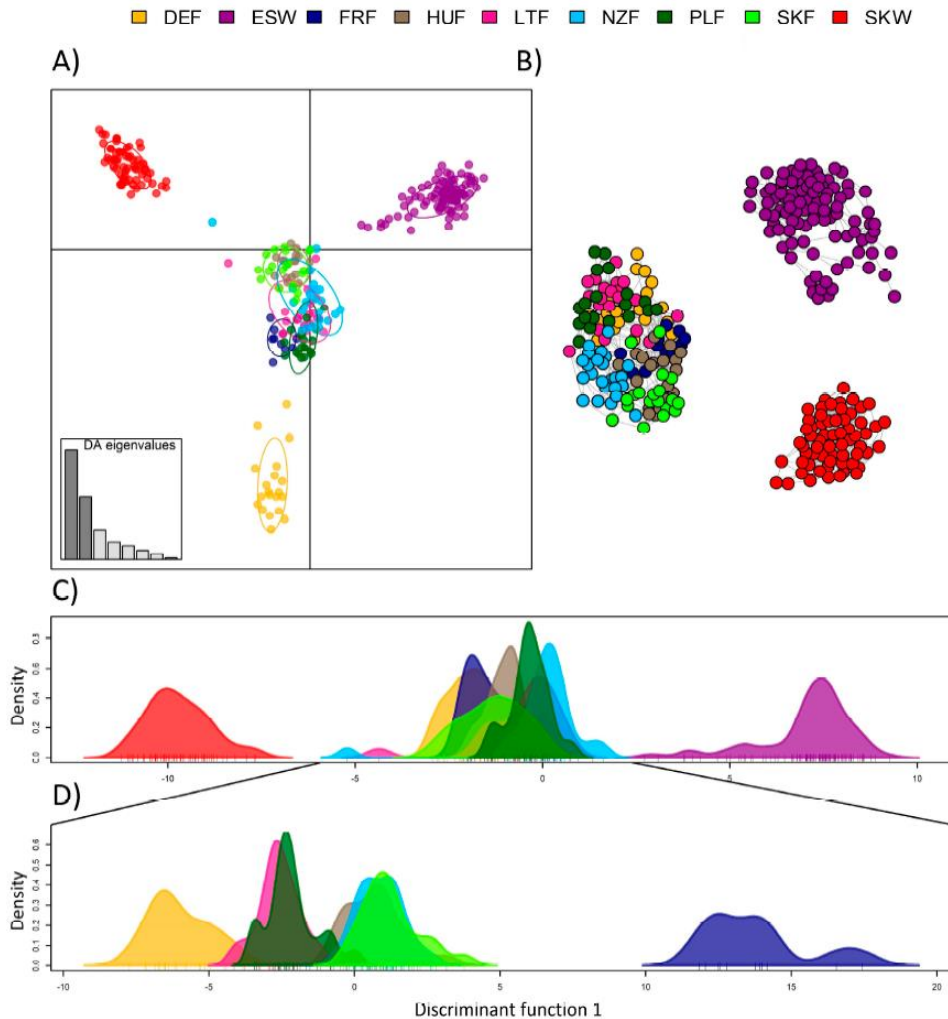
Prístupy a nástroje pre analýzu populačnej štruktúry

Diskriminačná analýza hlavných komponentov (DAPC)

- DAPC je metóda diskriminačnej analýzy zameraná na vizualizáciu genetickej štruktúry medzi preddefinovanými skupinami alebo klastrami. Využíva PCA na redukciu dimenzie dát a následne diskriminačnú analýzu na maximalizáciu rozlíšenia medzi populáciami
- presnejšie zobrazenie genetickej štruktúry medzi vopred definovanými klastrami, napríklad odlišných genetických subpopulácií alebo subpopulácií, ktoré vykazujú genetické premiešanie
- výhody: vysoká presnosť pri detekcii rozdielov medzi populáciami, jednoduchá a dobre interpretovateľná vizualizácia
- nevýhody: vyžaduje preddefinované skupiny, môže byť citlivá na nízku úroveň diverzity alebo vysokú mieru príbuznosti



Prístupy a nástroje pre analýzu populačnej štruktúry



Obr. 5: Genetické vzťahy medzi 7 farmovými a 2 voľne žijúcimi populáciami jeleňa lesného na základe prvej a druhej diskriminačnej funkcie DAPC analýzy (A), genetických vzdialeností (B) a prvej diskriminačnej funkcie samostatne za celý súbor (C) a farmové populácie (D) (Moravčíková et al., 2023)

Diskriminačná analýza hlavných komponentov (DAPC)

- Softvérové nástroje: R (balík adegenet)

Prístupy a nástroje pre analýzu populačnej štruktúry

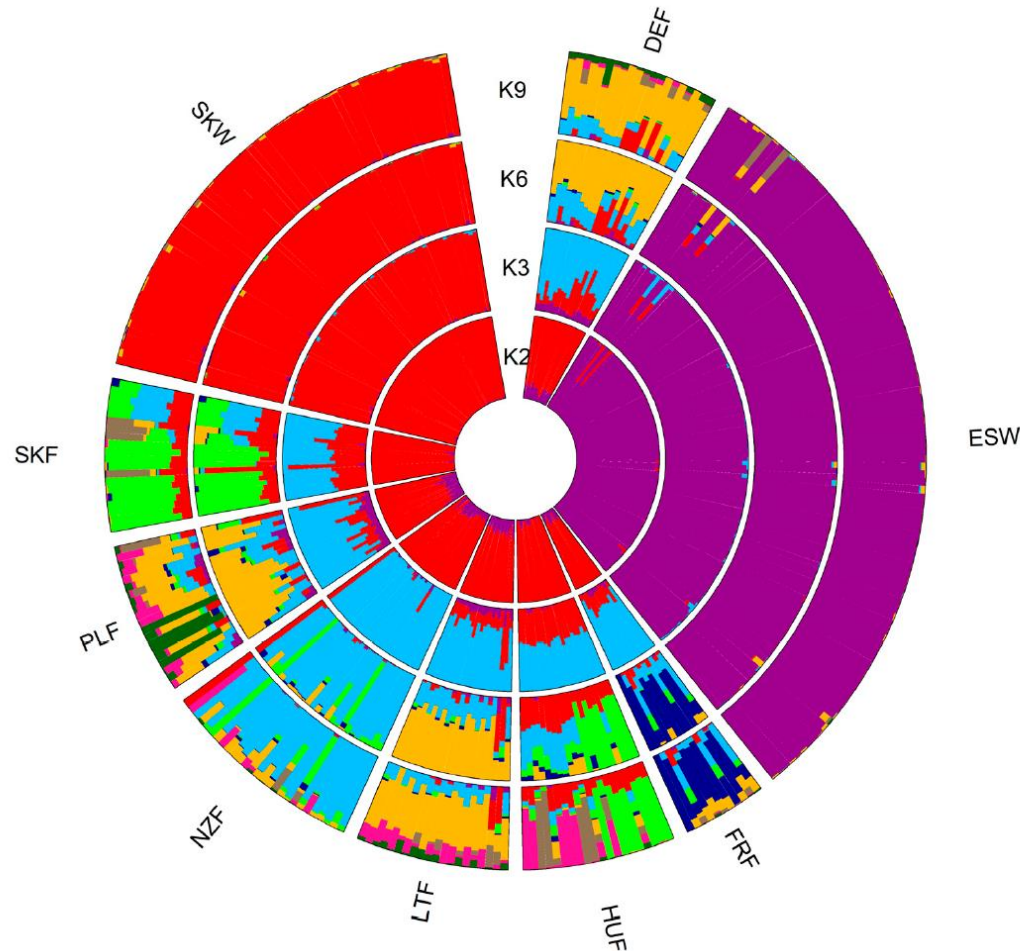
Analýza admixie – bayesovský prístup

- detekcia populačnej štruktúry a identifikácia genetických klastrov bez nutnosti preddefinovať skupiny
- umožňuje identifikáciu genetických skupín a mieru admixie v rámci jedincov, čo je užitočné pri štúdiu migrácie a diferenciácie
- výhody: presná identifikácia genetických klastrov, flexibilná metóda pre komplexné štruktúry
- nevýhody: časovo náročný výpočet, spoľahlivosť výsledkov závisí od počtu použitých SNP markerov a testovaných jedincov



Prístupy a nástroje pre analýzu populačnej štruktúry

ISAGREED



Analýza admixie – bayesovský prístup

- Softvérové nástroje: napr. STRUCTURE, ADMIXTURE, BAPS, FASTSTRUCTURE

Obr. 6: Reprezentatívne výsledky testovania genetického premiešania medzi farmovými a voľne žijúcimi populáciami jeleňa lesného pomocou bayesovského prístupu pre K=2, K=3, K=6 a K=9 (Moravčíková et al., 2023)

Prístupy a nástroje pre analýzu populačnej štruktúry

Tok génov medzi populáciami – program TreeMIX

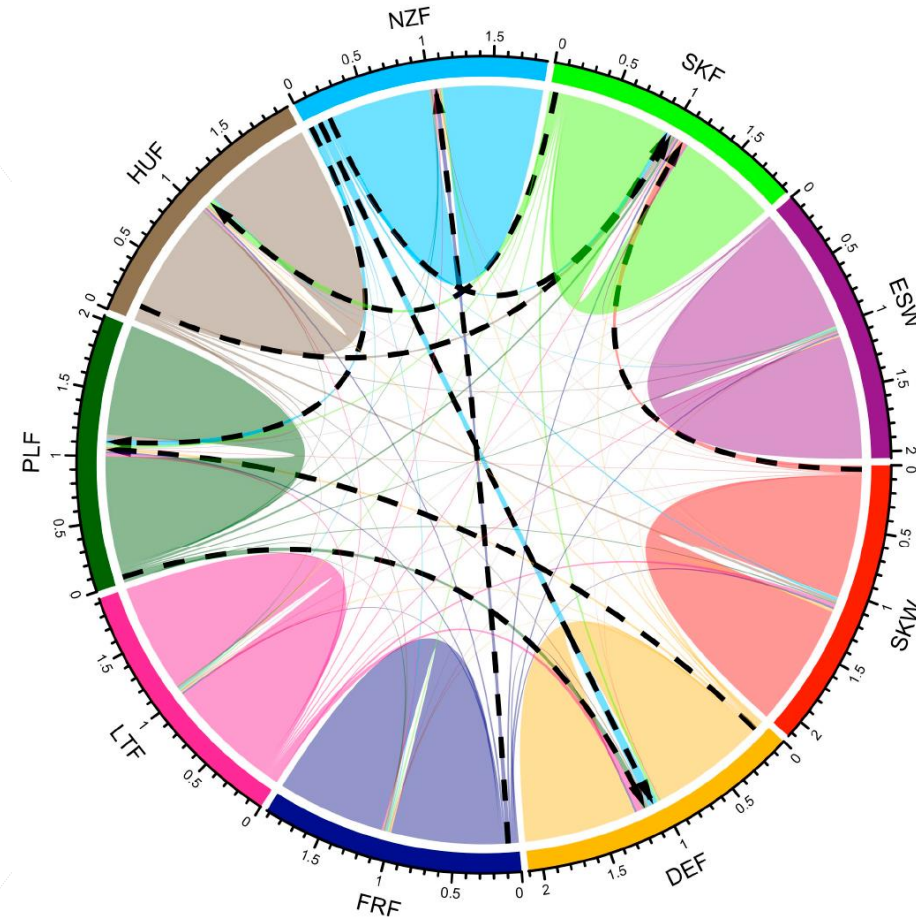
- na základe frekvencií alel vytvára fylogenetické stromy s možnosťou testovania intenzity migrácie medzi populáciami
- metóda je založená na maximálnej pravdepodobnosti a umožňuje simultánne odhadnúť fylogenetické vzťahy a migráciu medzi populáciami
- výhody: umožňuje detekciu intenzity migrácie a toku génov v minulosti
- nevýhody: spoľahlivosť výsledkov závisí od množstva dostupných genotypových dát ako aj spoľahlivosti odhadu frekvencií alel



Prístupy a nástroje pre analýzu populačnej štruktúry

Tok génov medzi populáciami – program Bayesass

- využíva bayesovský prístup
- umožňuje stanoviť intenzitu toku génov medzi populáciami aj vo vnútri nich



Obr. 7: Grafická vizualizácia intenzity toku génov medzi farmovými a voľne žijúcimi populáciami jeleňa lesného (Moravčíková et al., 2023)

Prístupy a nástroje pre analýzu populačnej štruktúry

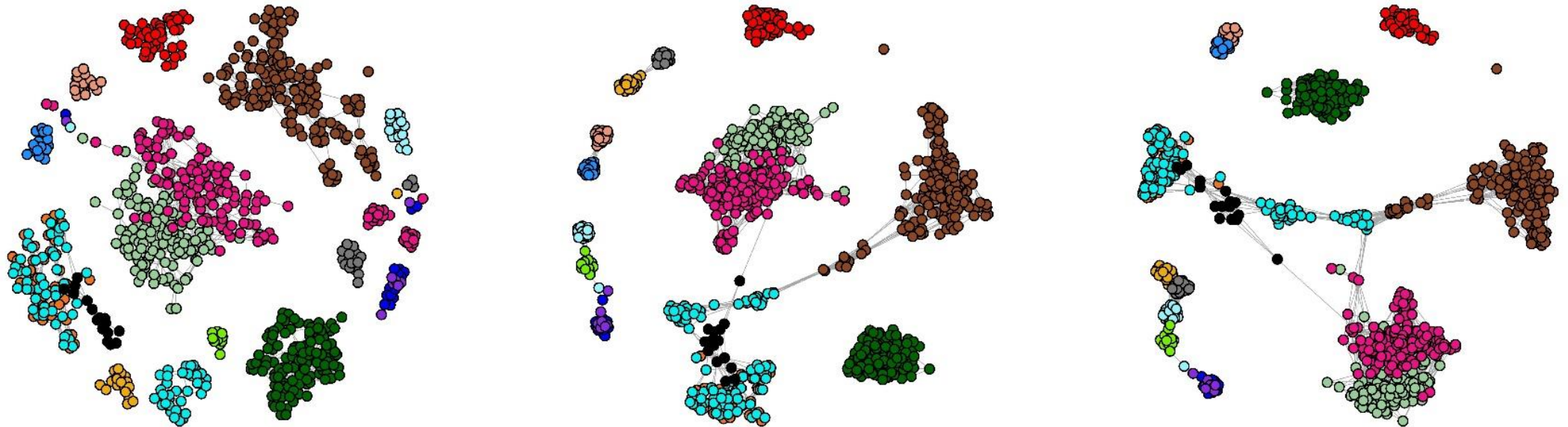
Konštrukcia genetických sietí – program Netview

- je vizualizačný nástroj, ktorý používa genetické siete na zobrazenie vzťahov medzi jedincami alebo populáciami
- vytvára genetické siete, ktoré zobrazujú príbuznosť a tok génov medzi populáciami
- je vhodný na zhodnotenie komplexných vzťahov ako aj vplyvu migrácie
- výhody: intuitívna vizualizácia, vhodné na zobrazenie admixie a diferenciácie
- nevýhody: obmedzené použitie v prípade populácií s veľkým počtom jedincov



Prístupy a nástroje pre analýzu populačnej štruktúry

Konštrukcia genetických sietí – program Netview



Obr. 8: Grafická vizualizácia výsledkov testovania 3 rozdielnych scénarov vývoja vnútro a medzipopulačných genetických vzťahov v rámci 16 plemien hovädzieho dobytku pomocou programu Netview (nepublikované výsledky)

Prístupy a nástroje pre analýzu populačnej štruktúry

Konštrukcia fylogenetických stromov

- grafické reprezentácie evolučných vzťahov medzi populáciami alebo druhmi na základe ich genetických dát
- používajú sa na vizualizáciu genealogických alebo genetických vzťahov, modelovanie evolučných procesov a sledovanie diferenciácie a migrácie populácií
- môžu byť vytvorené pomocou rôznych algoritmov a modelov, najčastejšie založených na genetických vzdialenostiach (napr. Neiova genetická vzdialenosť) alebo pravdepodobnostných modeloch (maximálna spoľahlivosť a Bayesovské metódy).

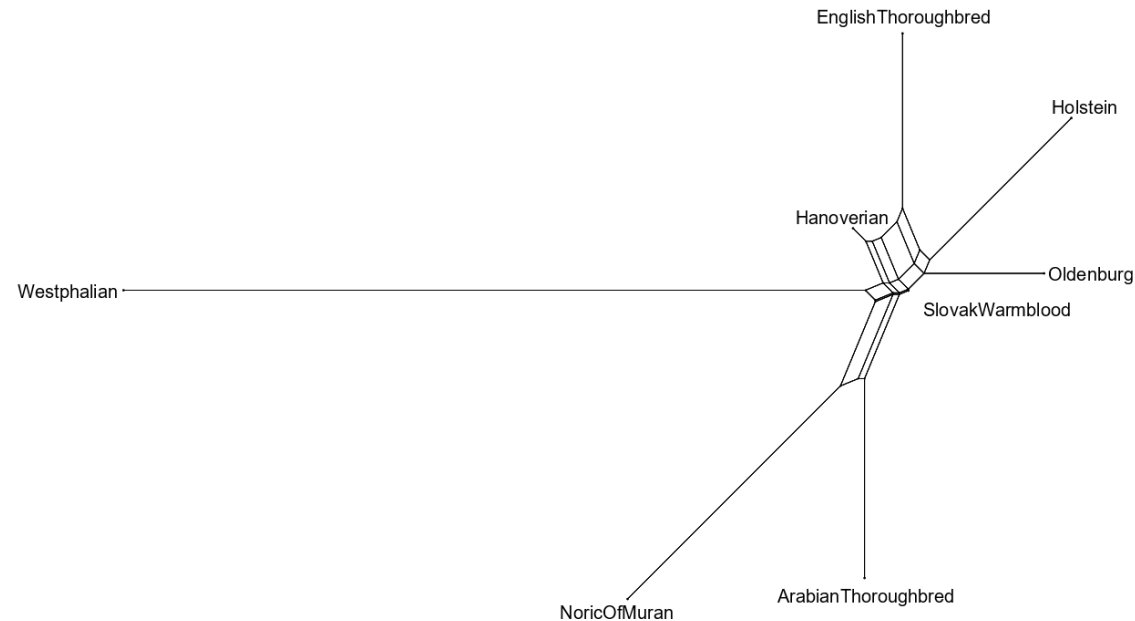


Co-funded by
the European Union

Prístupy a nástroje pre analýzu populačnej štruktúry

Konštrukcia fylogenetických stromov

- Softvérové nástroje: napr. MEGA, SplitsTree, rôzne balíky programu R



Obr. 9: Fylogenetický strom zostrojený na základe matice Neiových genetických vzdialeností reflektujúci vzájomné vzťahy medzi 8 plemenami koní (nepublikované výsledky)



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

- genomické oblasti pod významným selekčným tlakom – tzv. selekčné signály
- analýza selekčných signálov umožňuje lepšie pochopiť:
 - evolučné procesy a vplyv domestikácie
 - vplyv prírodnej a intenzívnej zámernej selekcie na špecifické genomické oblasti kontrolujúce preferované fenotypové znaky či už z pohľadu adaptability, odolnosti či úžitkovosti jedincov, populácií i druhov hospodárskych zvierat
- identifikovať genomické oblasti vykazujúce pokles genetickej variability resp. jej zvýšenie
- fenotypové informácie nie sú potrebné



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

- dve skupiny prístupov:
 - hodnotenie medzipopulačných/ medziľemenných rozdielov: napr.
 - Wrightov F_{ST} index na celogenómovej úrovni
 - Rozdiely vo väzbovej nerovnováhe (LD) – analýza na základe haplotypovej štruktúry
 - PCA analýza
 - hodnotenie na vnútropopulačnej úrovni: napr.
 - Distribúcia homozygotných úsekov
 - Distribúcia heterozygotných úsekov
 - Úroveň väzbovej nerovnováhy
 - RDA analýza
 - Tajimova D štatistika



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medzipliesmenných rozdielov - Wrightov F_{ST} index

- jeden z najčastejšie využívaných prístupov
- selekčné signály sú identifikované na základe rozdielov v alelických frekvenciách medzi populáciami, ktoré vznikli v dôsledku napr. rozdielneho chovného cieľa alebo plemenného štandardu
- dva základné typy signálov - rozdielnemu typu selekcie zodpovedajú úseky (reprezentované viacerými lokusmi resp. SNP markermi v tesnej blízkosti) s vysokou hodnotou F_{ST} indexu a naopak úseky s nízkou hodnotou reprezentujú genomické oblasti, ktoré u daných plemien podliehali rovnakému typu selekcie



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov - Wrightov F_{ST} index

- hraničná hodnota definujúca signál – napr. 1% najvyšších hodnôt
- výhody: relatívne jednoduchý spôsob výpočtu a široké využitie v populačnej genetike
- nevýhody: nie je možné využiť pre identifikáciu signálov na vnútro populačnej úrovni

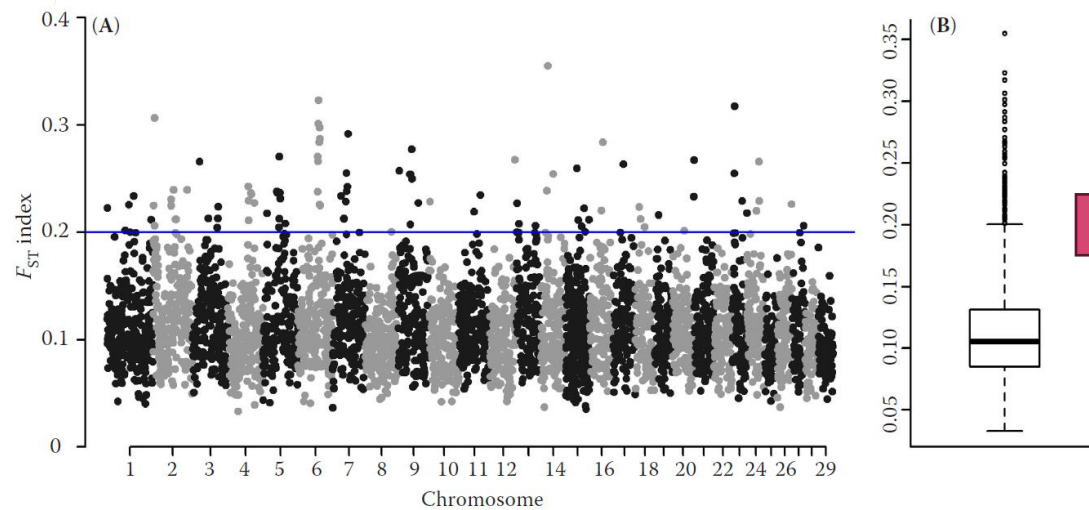


Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov - Wrightov F_{ST} index

- Softvérové nástroje: napr. PLINK



Obr. 10: Distribúcia hodnôt F_{ST} indexu v autozomálnom genóme mäsových plemien dobytky (A) a celkovo pomocou krabicového grafu (B) (Moravčíková et al., 2019)

Tab. 1: Popis vybraných selekčných signálov (Moravčíková et al., 2019)

| BTA | Start (Mb) | End (Mb) | Length (Mb) | Genes n | QTL in region |
|-----|------------|----------|-------------|-----------|--|
| 1 | 62.046 | 75.434 | 13.388 | 164 | live weight at age of 365 days, carcass weight, BSE resistance |
| | 80.063 | 91.358 | 11.295 | 132 | |
| 2 | 5.277 | 5.664 | 0.387 | 6 | marbling score, milk production, carcass weight |
| | 61.736 | 62.266 | 0.530 | 9 | |
| 3 | 70.560 | 77.289 | 6.729 | 45 | marbling score, length of production life, birth weight, IMF |
| | 84.966 | 88.207 | 3.241 | 28 | |
| 4 | 70.452 | 71.554 | 1.101 | 10 | LMA (<i>Longissimus thoracis cut</i>) |
| | 80.303 | 81.434 | 1.131 | 2 | |
| 5 | 48.460 | 49.399 | 0.939 | 11 | dressing percentage, birth weight, LMA, FSH concentration, EBV for backfat thickness |
| | 55.652 | 56.618 | 0.966 | 38 | |
| 6 | 57.470 | 60.557 | 3.086 | 154 | growth, strength, body frame, ham thickness, claw angle, claw quality, tits placement, udder quality, udder depth, fat and protein content |
| | 67.644 | 69.185 | 1.541 | 21 | |
| 7 | 70.001 | 72.907 | 2.906 | 35 | ovulation intensity, SCS |
| | 74.085 | 76.210 | 2.125 | 6 | |
| 9 | 38.904 | 39.327 | 0.423 | 8 | fat content |
| | 45.439 | 48.909 | 3.469 | 70 | |
| 9 | 51.021 | 53.468 | 2.447 | 67 | milk yield, fat and protein yield, ham angle, conformation traits 2 and 6 |
| | 40.401 | 43.145 | 2.744 | 35 | |
| 9 | 46.531 | 69.144 | 22.613 | 148 | |



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medzipliedenných rozdielov – integrované skóre haplotypov (iHS)

- selekčné signály sú odvodené od zmeny väzbovej nerovnováhy v genóme hodnotených plemien a vzniku špecifických haplotypov v dôsledku väzbovej nerovnováhy
- hodnota iHS môže byť definovaná jednoducho ako miera vyjadrujúca, ako neobvyklý je haplotyp pozostávajúci zo špecifických SNP markerov, v porovnaní so zvyškom genómu



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov – integrované skóre haplotypov (iHS)

- iHS je obzvlášť citlivá metóda na detekciu efektu nedávnej selekcie, ktorá vedie k zvýšeniu frekvencie určitého alelického variantu v populácii, ale ešte nestihla eliminovať iné varianty v danom lokuse
- analýza začína výpočtom hodnoty EHH (extended haplotype homozygosity), ktorá kvantifikuje pokles homozygotnosti haplotypu od určitého SNP markera a následne pokračuje výpočtom hodnoty iHS, ktorý je založený na logaritme pomeru integrovaných hodnôt EHH pre dva alelické varianty



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov – integrované skóre haplotypov (iHS)

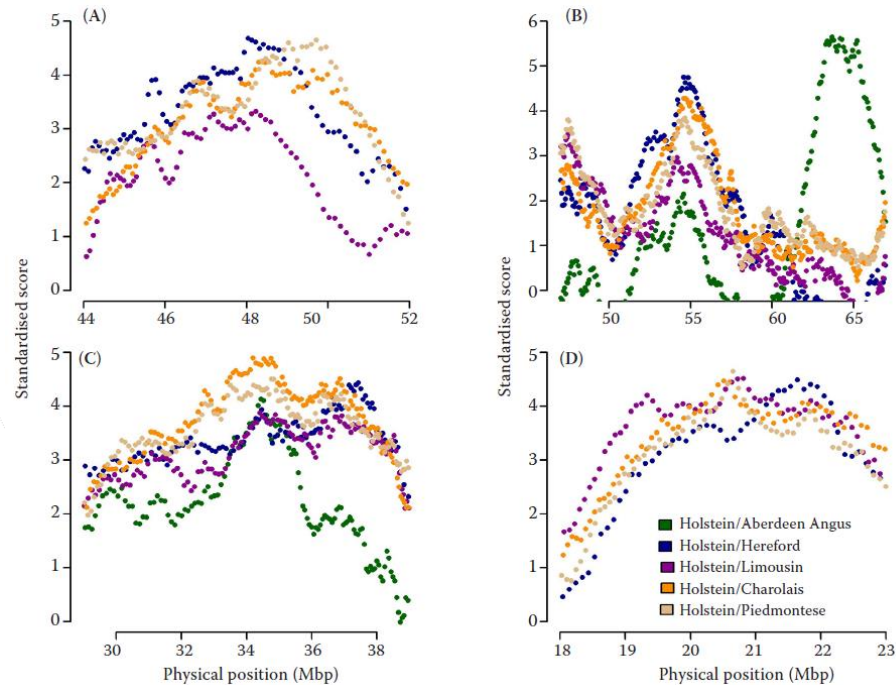
- iHS môže nadobúdať pozitívne hodnoty (haplotyp nesúci jednu alelu je dlhší a má vyššiu EHH, čo indikuje signifikantný vplyv pozitívnej selekcie) alebo negatívne hodnoty (alternatívna alela má vyššiu EHH, čo môže takisto odrážať selekciu, ale v opačnom smere)
- hraničná hodnota definujúca signál – napr. 1% najvyšších pozitívnych hodnôt
- výhody: vhodný prístup pre detekciu efektu nedávnej selekcie a identifikáciu signálov vznikajúcich napr. v dôsledku adaptácie, možnosť analýzy haplotypovej štruktúry
- Nevýhody: potreba kvalitných a robustných genomických dát, potreba definovania resp. určenie frekvencií haplotypov



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medzipliesmenných rozdielov – integrované skóre haplotypov (iHS)

- Softvérové nástroje: napr. balík programu REHH, Haploview



Obr. 11: Rozdiely vo väzbovej nerovnováhe na chromozómoch 7 (A), 13 (B), 20 (C) a 26 (D) medzi holštajnským plemenom a mäsovými plemenami dobytky (Moravčíková et al., 2019)

Tab. 2: Autozomálne oblasti pod selekčným tlakom identifikované na základe variability vo väzbovej nerovnováhe (Moravčíková et al., 2019)

| Breed | BTA | Start (Mb) | End (Mb) | Region length (Mb) | Genes <i>n</i> | QTL in region |
|----------------|-----|------------|----------|--------------------|----------------|--|
| Aberdeen Angus | 13 | 62.825 | 65.859 | 3.034 | 82 | dressing percentage |
| Hereford | 7 | 47.201 | 49.331 | 2.130 | 39 | dressing percentage |
| | 13 | 54.208 | 55.559 | 1.350 | 70 | |
| Limousin | 26 | 19.088 | 22.122 | 3.034 | 47 | dressing percentage, IMF, milk yield |
| Charolaise | 20 | 32.621 | 37.325 | 4.704 | 39 | |
| Piedmontese | 7 | 48.308 | 50.202 | 1.894 | 53 | milk production, fat and protein content dressing percentage, IMF, milk yield |
| | 20 | 33.418 | 35.197 | 1.778 | 7 | |
| | 26 | 20.572 | 21.075 | 0.503 | 11 | |
| Romagnola | 6 | 38.791 | 41.800 | 3.009 | 25 | birth weight, growth, strength, milk yield, fat and protein content |



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov – PCA analýza

- predpokladá, že selekčné signály v genóme vznikli v dôsledku lokálnej adaptácie jedincov na podmienky prostredia
- alternatívna metóda pre identifikáciu selekčných signálov k F_{ST} indexu
- detekcia selekčných signálov vychádza z predpokladu existencie korelácie medzi genetickými variantmi a hlavnými komponentmi, ktorá reflektuje lokálnu adaptáciu jednotlivých populácií na produkčné resp. chovateľské prostredie



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov – PCA analýza

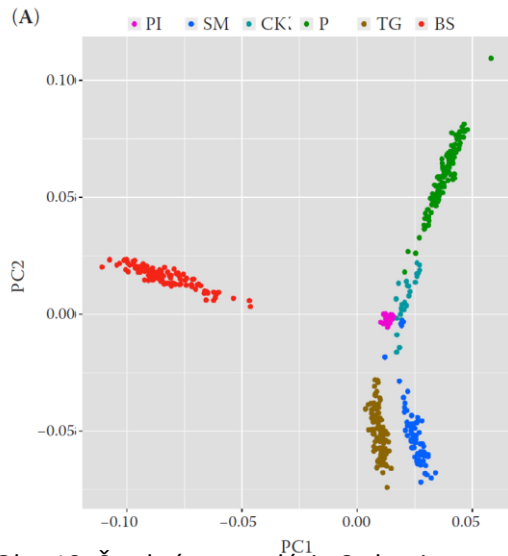
- na identifikáciu selekčných signálov sa používa napr. Mahalanobisov štatistický test, ktorý hodnotí vzdialenosť bodov od priemeru - identifikácia SNP markerov vykazujúcich asociáciu s pozitívnou selekciou je potom založená na zostavení vektora z-skóre, získaného regresnou analýzou vzťahu medzi SNP markermi a hlavnými komponentmi K
- hraničnú hodnotu definujúci signál možno stanoviť napr. na základe FDR testu
- výhody: efektívna vizualizácia komplexných genetických dát
- nevýhody: alternatívny, nie až tak často využívaný prístup, komplikovaná interpretácia



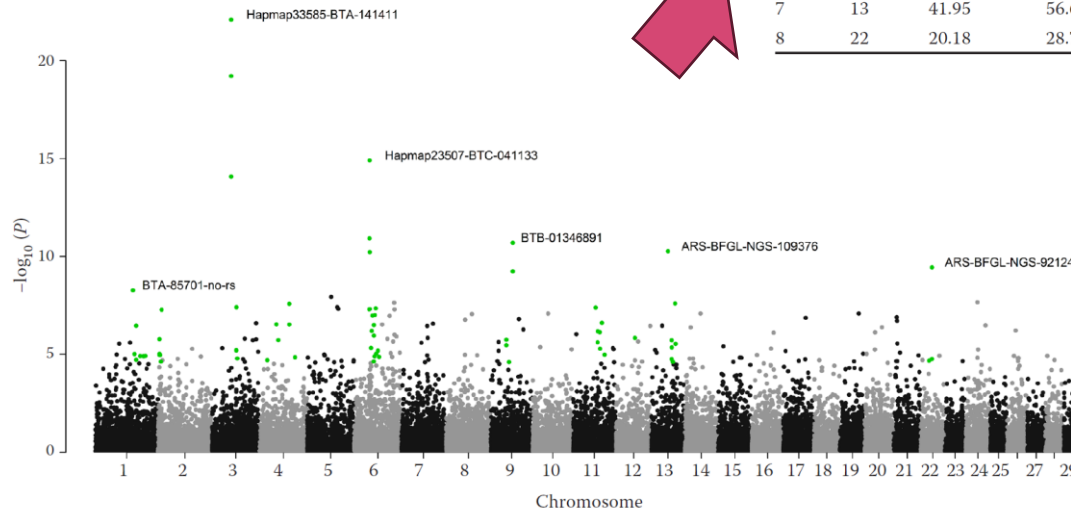
Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie medzipopulačných/ medziľemenných rozdielov – PCA analýza

- Softvérové nástroje: R balík PCAdapt



Obr. 12: Štruktúra populácie 6 plemien hovädzieho dobytku na základe PCA analýzy (Moravčíková et al., 2018)



Obr. 13: Selekčné signály identifikované prostredníctvom PCA analýzy (Moravčíková et al., 2018)

Tab. 3: Genomické oblasti vykazujúce najsilnejší selekčný signál (Moravčíková et al., 2018)

| Region | BTA | Start position (Mb) | End position (Mb) | No. of SNPs | QTL traits | No. of genes |
|--------|-----|---------------------|-------------------|-------------|--|--------------|
| 1 | 1 | 94.59 | 133.94 | 11 | Resistance to BSE; Birth weight; Adjusted weaning and yearling weight | 244 |
| 2 | 2 | 4.06 | 7.49 | 15 | Yearling weight; Kidney, pelvic and heart fat | 41 |
| 3 | 3 | 43.69 | 64.87 | 13 | Marbling score; Estimated kidney, pelvic and heart fat | 159 |
| 4 | 6 | 37.87 | 62.69 | 21 | Longissimus muscle area; Hot carcass weight; Birth weight; Yearling weight; Marbling score | 95 |
| 5 | 9 | 40.12 | 56.41 | 10 | Marbling score; Canonical conformation trait 2 | 110 |
| 6 | 11 | 57.13 | 72.84 | 14 | Yield grade | 92 |
| 7 | 13 | 41.95 | 56.69 | 9 | Canonical conformation trait 9 | 226 |
| 8 | 22 | 20.18 | 28.71 | 8 | | 23 |

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – Distribúcia homozygotných úsekov (ROH)

- tento prístup predpokladá, že oblasti v genóme vykazujúce silné selekčné signály sú výsledkom zvýšenia lokálnej homozygotnosti v dôsledku intenzívneho šľachtenia na znaky definované v plemennom štandarde každého plemena
- ROH úseky tvoriace selekčné signály lokalizované v genóme v tesnej blízkosti sú tvorené alelami pochádzajúcimi od spoločných predkov, ktoré sa dedia z generácie na generáciu v nezmennej forme



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – Distribúcia homozygotných úsekov (ROH)

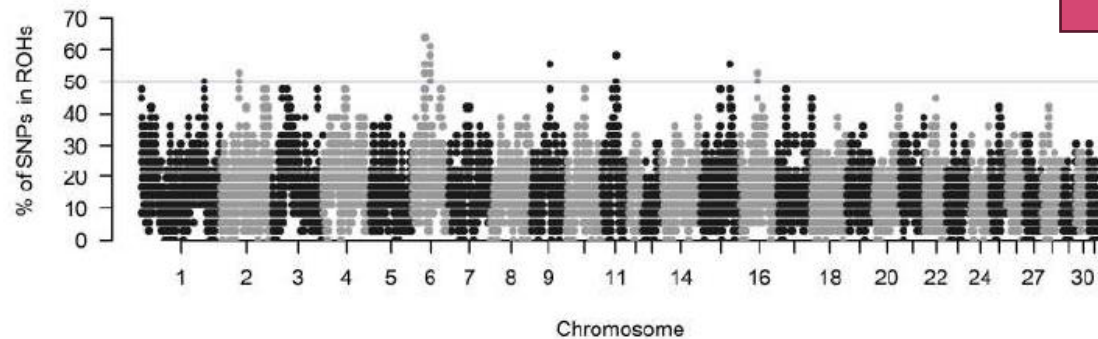
- selekčné signály sa stanovujú na základe frekvencie výskytu SNP markerov v ROH v konkrétnej oblasti naprieč jedincami v populácii
- hraničná hodnota definujúca signál – napr. 1% najvyšších hodnôt
- výhody: umožňuje detekovať oblasti v ktorých došlo k zníženiu diverzity, dobrý indikátor efektu pozitívnej selekcie
- nevýhody: potreba kvalitných a robustných genomických dát



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – Distribúcia homozygotných úsekov (ROH)

- Softvérové nástroje: napr. PLINK, R balík detectRUNS



Obr. 14: Prekrývajúce sa ROH úseky v autozomálnom genóme slovenského teplotkrvníka (Moravčíková et al., 2020)

Tab. 4: Selekčné signály v genóme slovenského teplotkrvníka vykazujúce extrémne hodnoty frekvencie SNP markerov v ROH úsekoch (Moravčíková et al., 2020)

| ECA | Start position (Mb) | End position (Mb) | Region size (Mb) | Protein-coding genes |
|-----|---------------------|-------------------|------------------|---|
| 1 | 147.61 | 148.06 | 0.44 | STARD9, TTBK2, CDAN1, HAUS2, LRRC57, SNAP23, ZNF106, CAPN3, GANC, VPS39 |
| 2 | 42.74 | 43.12 | 0.38 | RERE, SLC45A1, ERRF1I, PARK7, TNFRSF9 |
| 6 | 29.36 | 30.02 | 0.66 | CACNA2D4, DCP1B, LRTM2, ADIPOR2 |
| | 41.18 | 42.71 | 1.54 | LRP6, MANSC1, BORCS5, DUSP16, CREBL2, GRP19, CDKN1B, APOLD1, DDX47, GRPC5A, GRPC5D, HEBP1, FAM234B, GSG1, EMP1, GRIN2B |
| 9 | 44.25 | 44.53 | 0.27 | UQCRB, MTERF3, PTDSS1, SDC2 |
| 11 | 32.32 | 33.58 | 1.26 | MSI2, CCDC182, MRPS23, CUEDC1, VEZF1, SRSF1, DYNLL2, EPX, MKS1, LPO, MPO, TSPOAP1, MIR142, RNF43, SUPT4H1, HSF5, MTMR4, TEX14, RAD51C PPM1E, TRIM37, SKA2 |
| 15 | 67.26 | 67.83 | 0.57 | LBH, YPELP |
| 16 | 39.90 | 40.62 | 0.71 | SLC26A6, TMEM89, UQCRC1, MIR711, PFKFB4, SHISA5, TREX1, ATRIP, CCDC51, PLXNB1, FBXW12, SPINK8, NME6, ECATH-3, ECATH-2, CDC25A, MAP4 |

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – Distribúcia heterozygotných úsekov (HRR)

- táto metóda slúži na detekciu oblastí vykazujúcich vysokú mieru heterozygotnosti, ktoré môžu mať význam z pohľadu adaptability, odpovede na environmentálne zmeny či výskyt patogénov – heterozygotné jedince majú vyšší fitness ako homozygotné
- vysoká úroveň heterozygotnosti môže byť výsledkom pôsobenia stabilizačnej selekcie – uchovávanie genetickej diverzity v rámci populácie
- podobný prístup ako pri ROH – avšak selekčné signály tvoria rôzne dlhé nepretržité úseky heterozygotných genotypov v genóme



Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – Distribúcia heterozygotných úsekov (HRR)

- selekčné signály sa stanovujú na základe frekvencie výskytu SNP markerov v HRR v konkrétnej oblasti naprieč jedincami v populácii
- hraničná hodnota definujúca signál – napr. 1% najvyšších hodnôt
- výhody: umožňuje detekovať oblasti v ktorých je zvýšený podiel heterozygotných genotypov – indikátor oblastí dôležitých z hľadiska adaptácie či evolučného potenciálu
- nevýhody: potreba kvalitných a robustných genomických dát

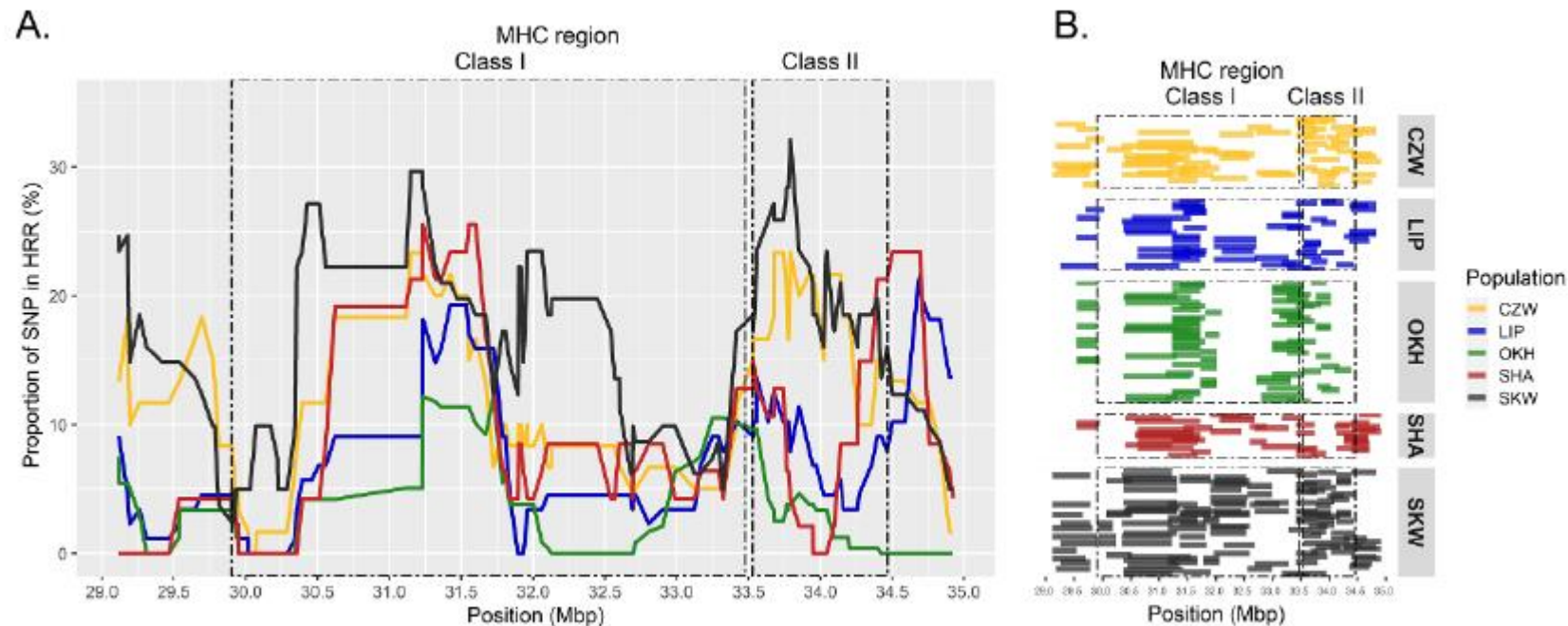


Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútro populačnej úrovni – Distribúcia heterozygotných úsekov (HRR)

- Softvérové nástroje: R balík detectRUNS



Obr. 15: Grafická vizualizácia frekvencie výskytu SNP markerov v HRR úsekoch (A) a distribúcia HRR úsekov v oblasti hlavného histokompatibilného komplexu (B) (Moravčíková et al., 2024)

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – RDA analýza

- RDA (Redundancy analysis) testuje vzťahy medzi genetickou variabilitou a environmentálnymi faktormi – vplyv prírodnej selekcie
- jedná sa o metódu hodnotenia asociácií medzi genotypom a prostredím (GEA), ktorá súčasne hodnotí percento genomickej variability vysvetlené environmentálnymi premennými a detekuje lokusy pod selekčným tlakom
- dvojkroková analýza v ktorej sú genetické a environmentálne dáta hodnotené pomocou viacrozmernej lineárnej regresie
- výhody: komplexný prístup na hodnotenie vzťahov medzi genetickou variabilitou v rámci populácie a faktormi prostredia
- nevýhody: potreba kvalitných a robustných dát
- Softvérové nástroje: napr. R balík vegan, DeepGenomeScan



Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútropopulačnej úrovni – Tajimova D štatistika

- hodnotí populačnú diverzitu a môže slúžiť ako jeden z indikátorov stabilizačnej selekcie
- pozitívne hodnoty indikujú stabilizačnú selekciu - frekvencie alel majú strednú hodnotu, čo má za následok väčšiu variabilitu ako by sa očakávalo na základe teórie neutrálnej evolúcie
- negatívne hodnoty sú naopak asociované s efektom pozitívnej selekcie, ktorá vedie k zníženiu genetickej diverzity a nukleotidovej variability
- hraničná hodnota definujúca signál – napr. 1% najvyšších pozitívnych hodnôt
- výhody: umožňuje detekovať oblasti v ktorých je zvýšený podiel heterozygotných genotypov – indikátor oblastí dôležitých napr. z hľadiska adaptácie
- nevýhody: potreba kvalitných a robustných genomických dát

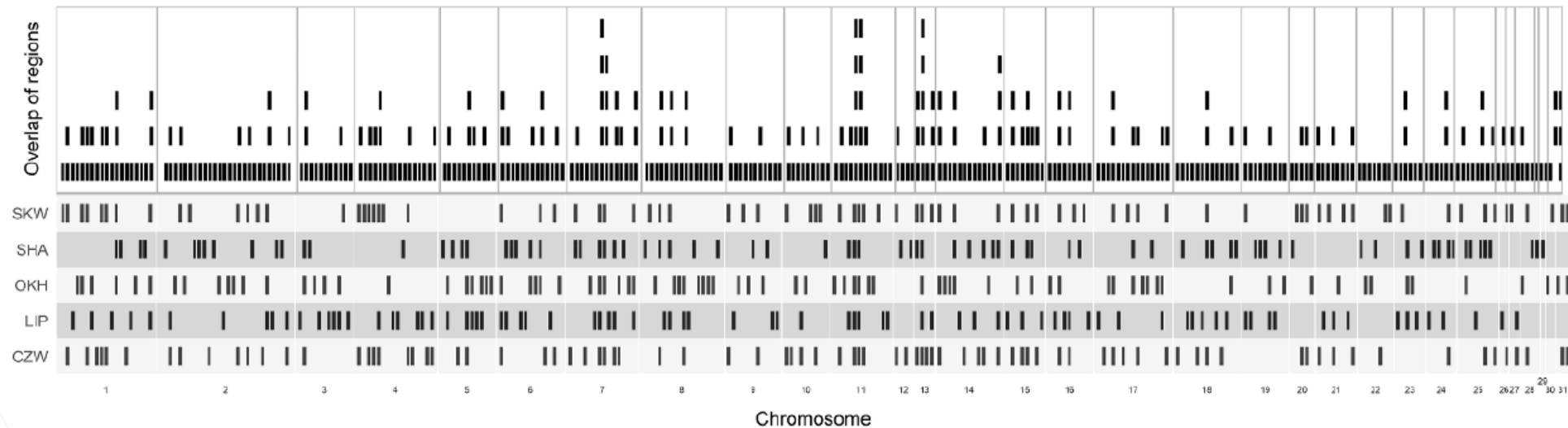


Co-funded by
the European Union

Prístupy a nástroje pre vyhodnotenie vplyvu selekcie na genóm hospodárskych zvierat

Hodnotenie na vnútro populačnej úrovni – Tajimova D štatistika

- Softvérové nástroje: napr. VCFtools, R balík SnpR



Obr. 16: Signály stabilizačnej selekcie v genóme 5 plemien koní identifikované na základe Tajimovho D indexu (Moravčíková et al., 2024)

Funkčná anotácia oblastí signifikantne ovplyvnených selekčným tlakom

- vyhľadávanie lokusov pre kvantitatívne vlastnosti (QTL) v okolí selekčných signálov
- vyhľadávanie génov lokalizovaných priamo v oblasti alebo v blízkosti selekčných signálov – potenciálne kandidátne gény pre ďalšiu analýzu
- funkčná anotácia: identifikácia biologických funkcií génov pomocou nástrojov ako GO (Gene Ontology) alebo KEGG (Kyoto Encyclopedia of Genes and Genomes)



Funkčná anotácia oblastí signifikantne ovplyvnených selekčným tlakom

Databázy pre identifikáciu QTL a génov

- Animal QTLdb - databáza informácií o QTL asociovaných s významnými fenotypovými vlastnosťami rôznych druhov hospodárskych zvierat
- BioMart – Ensembl – jednoduchý webový nástroj pre získavanie údajov o genóme (napr. pozície génov) bez potreby znalosti programovania

Funkčná analýza génov

- DAVID: webový nástroj na funkčnú anotáciu a analýzu úlohy génov, vrátane klasifikácie podľa biologických procesov a dráh



Co-funded by
the European Union

Funkčná anotácia oblastí signifikantne ovplyvnených selekčným tlakom

- výhody: detailná analýza oblastí v genóme signifikantne ovplyvnených selekciou umožňuje identifikáciu konkrétnych génov a biologických dráh zodpovedných za fenotypové vlastnosti, môže pomôcť pri ďalšom výskume génov a QTL lokusov s potenciálnym využitím v šľachtiteľských programoch
- nevýhody: prekryv medzi oblasťami selekčných signálov a funkčnými oblasťami nemusí vždy znamenať kauzálny vzťah, informácie v dostupných databzách sú obmedzené na aktuálne poznatky a nie vždy musia pokrývať všetky relevantné gény resp. QTL lokusy



Zoznam použitej literatúry

- MORAVČÍKOVÁ, Nina - SIMČIČ, Mojca - MESZÁROŠ, G. - SÖLKNER, Johann - KUKUČKOVÁ, Veronika - VLČEK, Michal - TRAKOVICKÁ, Anna - KASARDA, Radovan. Genomic response to natural selection within alpine cattle breeds. In *Czech journal of animal science*. ISSN 1212-1819, 2018, vol. 63, iss. 4, s. 136-143. Dostupné na internete: <http://dx.doi.org/10.17221/62/2017-CJAS>.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - VOŠTRÝ, Luboš - KRUPOVÁ, Zuzana - KRUPA, Emil - LEHOČKÁ, Kristína - OLSANŠKA, Barbora - TRAKOVICKÁ, Anna - NADASKÝ, Rudolf - ŽIDEK, Radoslav - BELEJ, Ľubomír - GOLIAN, Jozef. Analysis of selection signatures in the beef cattle genome. In *Czech journal of animal science*. ISSN 1212-1819, 2019, vol 64, no. 12, s. 491-503. Dostupné na internete: <https://doi.org/10.17221/226/2019-CJAS>.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - HALO, Marko - KADLEČÍK, Ondrej - TRAKOVICKÁ, Anna - LEHOČKA, Kristína - OLSANŠKA, Barbora - CANDRAK, Juraj. Genome-wide distribution of autozygosity islands in slovak warmblood horse. In *AGROFOR*. ISSN 2490-3434, 2020, vol. 5, no. 1, s. 77-84.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - ŽIDEK, Radoslav - VOŠTRÝ, Luboš - VOŠTRÁ VYDROVÁ, Hana - VASEK, Jakub - ČILOVÁ, Daniela. Czechoslovakian wolfdog genomic divergence from its ancestors canis lupus, german shepherd dog, and different sheepdogs of european origin. In *Genes*. ISSN 2073-4425 online, 2021, vol. 12, iss.6, article 832, p. 1-18. Dostupné na internete: <https://doi.org/10.3390/genes12060832>.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - ŽIDEK, Radoslav - MCEWAN, John Colin - BRAUNING, Rudiger - LANDETE-CASTILLEJOS, Tomas - CHONCO, Louis - ČIBEREJ, Juraj - POKORADI, Jaroslav. Traces of Human-Mediated Selection in the Gene Pool of Red Deer Populations. In *Animals*. ISSN 2076-2615, 2023, vol. 13, iss. 15, art. no. 2525 [17] s. Dostupné na internete: <https://doi.org/10.3390/ani13152525>.
- MORAVČÍKOVÁ, Nina - KASARDA, Radovan - VOŠTRÁ VYDROVÁ, Hana - VOŠTRÝ, Luboš - KARÁSKOVÁ, Barbora - CANDRAK, Juraj - HALO, Marko. Genomic variability of the MHC region: Empirical evidence from five horse breeds. In *Livestock science*. ISSN 1871-1413, 2024, vol. 284, art. no. 105480, [10] s. Dostupné na internete: <https://doi.org/10.1016/j.livsci.2024.105480>.





Siedlce University
of Natural Sciences
and Humanities

Partneri:



Czech University
of Life Sciences Prague



Ďakujem za pozornosť!

This presentation has been supported by the Erasmus+ KA2 Cooperation Partnerships grant no. 2021-1-SK01-KA220-HED-000032068 "Innovation of the structure and content of study programs in the field of animal genetic and food resources management with the use of digitalisation - Inovácia obsahu a štruktúry študijných programov v oblasti manažmentu živočíšnych genetických a potravinových zdrojov s využitím digitalizácie". The European Commission support for the production of this presentation does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.



Nina Moravčíková



nina.moravcikova@uniag.sk



Co-funded by
the European Union