

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
Fakulta agrobiologie, potravinových a přírodních zdrojů
Katedra genetiky a šlechtění

**Analýza vybraných vlastností prasat a možného využití SNP
pro jejich zlepšení**

.....
doktorská dizertační práce

Autor: Ing. Jan Calta

Školitel: prof. Ing. Luboš Vostrý, Ph.D.

Konzultant: Ing. Eliška Žáková, Ph.D. (Svaz chovatelů prasat)
doc. Ing. Jaroslav Čítek, Ph.D. (ČZU v Praze)

Praha 2024

Čestné prohlášení

Prohlašuji, že jsem předkládanou doktorskou dizertační práci „Analýza vybraných vlastností prasat a možného využití SNP pro jejich zlepšení“ vypracoval samostatně pod vedením uvedeného školitele a konzultantů. Při psaní dizertační práce byly dodrženy standardní postupy pro citování informačních zdrojů a nebyla porušena autorská práva třetích osob.

V Praze dne 30.07.2024

Poděkování

V první řadě děkuji své rodině, která mi umožnila takto dlouho studovat. Ovšem co se týče samotné dizertační práce, největší uznání zasluhují dvě osoby, bez jejichž nepřetržité podpory by nemohla vzniknout. První je prof. Ing. Luboš Vostrý, Ph.D., jenž vytvořil po odborné i lidské stránce skvělé pracovní prostředí a vždy si věděl rady, ačkoliv tvrdil opak. Druhou je Ing. Eliška Žáková, Ph.D., která se mnou strávila násobně více času, než musela, přičemž po celou dobu zůstala ohromně trpělivá a pečlivá. Dále bych rád vyjádřil poděkování: doc. Ing. Jaroslavu Čítkovi, Ph.D. a Ing. Kateřině Zadinové, Ph.D. z katedry chovu hospodářských zvířat za ochotné konzultace už od bakalářského studia; Ing. Jakubu Vaškovi, Ph.D., Ing. Karolíně Dvořákové Machové, Ph.D. a Ing. Ladislavu Tichému, Ph.D. z katedry genetiky a šlechtění za několikahodinové debaty o všem možném; Ing. Emilu Krupovi, Ph.D. a Ing. Zuzaně Krupové, Ph.D. z Výzkumného ústavu živočišné výroby za poskytnutí nezbytných materiálů včetně souvisejících znalostí a prof. Ing. Bc. Josefu Příbylovi, DrSc. za moudré rady nejen ve šlechtění, díky nimž jsem nemusel přešlapovat na místě; Ing. Janu Stibalovi za velmi vstřícnou podporu ze strany Svazu chovatelů prasat; Assoc. Prof. Priv.-Doz. Dr. Gáboru Mészárosovi z Univerzity přírodních zdrojů a biologických věd ve Vídni za možnost nabýt vědomosti přímo na jeho institutu.

Abstrakt

Šlechtění hraje klíčovou roli v zajištění efektivity živočišné výroby, a tedy v zajištění potravinové bezpečnosti. Moderní šlechtitelské postupy jako genomická selekce jsou založeny na dostupnosti desítek až stovek tisíc molekulárněgenetických markerů, přesněji jednonukleotidových polymorfizmů (SNP), o jejichž využití se v Česku snaží i národní šlechtitelský program prasat. Cílem předkládané dizertační práce proto bylo posouzení přínosu SNP pro genetické hodnocení prasat. Hlavní část práce se týkala návrhu nového statistického modelu pro současné hodnocení průměrného denního přírůstku, výšky hřbetního tuku a výšky svalu. Pro tento účel byla využita užitková data od téměř 124 tisíc jedinců mateřských i otcovských plemen a také vzorky z DNA čipů, ze kterých prošlo kontrolou kvality 1810 jedinců se zhruba 30 tisíci SNP. Výběr vhodného animal modelu v rámci nejlepší nevychýlené lineární předpovědi (BLUP) byl založen na parametrech dopředné validace, které vyjadřují prediktivní výkonnost. Takto byly rovněž srovnány tradiční animal modely s jedнокrokovou genomickou předpovědí (ssGBLUP). Zahrnutí SNP pomocí matice realizované příbuznosti mělo ve studovaných populacích zatím nízký účinek kvůli omezenému počtu genotypovaných zvířat. V hodnocení mateřských plemen ovšem lze tento účinek považovat za kladný, neboť se projevila převaha ssGBLUP v individuálních přesnostech předpovědí plemenných hodnot, přičemž celková prediktivní výkonnost modelu byla v případě českého bílého ušlechtilého zachována a v případě české landrase dokonce zlepšena. Ve sdruženém hodnocení otcovských plemen nebylo možné učinit jednoznačný závěr, jelikož výsledky patrně zkreslila nevhodná struktura dat. Na základě získaných výsledků bylo možné podpořit hypotézu o tom, že molekulárněgenetické informace přispívají ke zpřesnění genetického hodnocení. Nicméně dostupná literatura napovídá, že praktický dopad genomické selekce na genetický zisk nastane až poté, co budou nashromážděna genotypová data od několikanásobně většího množství jedinců. Zároveň se projevila nutnost pravidelně revidovat vhodnost statistického modelu včetně rozdělení dat nebo populační genetické parametry.

Klíčová slova: plemenná hodnota, genetický marker, statistický model, genomická selekce

Abstract

Breeding plays a crucial role in ensuring the efficiency of animal production and, consequently, food security. Modern breeding practices, such as genomic selection, are based on the availability of tens to hundreds of thousands of molecular genetic markers, specifically single nucleotide polymorphisms (SNPs). The utilization of SNPs is also an aim for the Czech pig breeding program. The objective of this presented dissertation was to assess the contribution of SNPs to pig genetic evaluation. The main part of the thesis focused on designing a new statistical model for simultaneous evaluation of average daily gain, backfat thickness, and loin muscle depth. Phenotypic data from nearly 124,000 individuals from maternal and paternal breeds were analysed, along with 1810 DNA microarray samples with roughly 30,000 SNPs (after quality control). The selection of an appropriate animal model within the best unbiased linear prediction (BLUP) framework was based on forward validation methods, which manifest model's predictive performance. Traditional animal models were compared with single-step genomic models (ssGBLUP). In the studied populations, the inclusion of SNPs through the realized relationship matrix had a limited effect due to the small number of genotyped animals. However, in the evaluation of maternal breeds, this effect could be considered positive, as ssGBLUP showed higher individual accuracies of breeding values, while overall predictive performance was maintained in Czech Large White and even increased in Czech Landrace. In the joint evaluation of paternal breeds, no clear conclusion could be drawn due to unsuitable data structure. Based on the obtained results, it is possible to support the hypothesis that molecular genetic information contributes to improving genetic evaluation. Nevertheless, available literature suggests that the practical impact of genomic selection on genetic gain will occur only after accumulating genotypes from several times higher number of individuals. Regular revision of the suitability of the statistical model, including data distribution, and population genetic parameters, is also essential.

Key words: breeding value, genetic marker, statistical model, genomic selection

Obsah

1.	Úvod	7
2.	Literární přehled	8
2.1.	Užitkovost prasat	8
2.2.	Šlechtění prasat v Česku	9
2.3.	Matematicko-statistické základy šlechtění	10
2.3.1.	Rozložení proměnlivosti	10
2.3.2.	Populační genetické parametry	11
2.3.3.	Smíšený model	12
2.3.4.	REML	16
2.3.5.	BLUP	19
2.4.	Výběr vhodného modelu	22
2.4.1.	Metody pro srovnání animal modelů	22
2.4.2.	Modely pro výšku tuku a svalu	27
2.5.	Využití SNP ve šlechtění	32
2.5.1.	Genotypovací metody	32
2.5.2.	Markery asistovaná selekce	34
2.5.3.	První metody genomické selekce	36
2.5.4.	Jednokroková GBLUP	38
2.5.5.	Genomická selekce u prasat	40
3.	Cíl práce a vědecká hypotéza	43
3.1.	Hlavní a vedlejší cíl práce	43
3.2.	Vědecká hypotéza	43
4.	Metodika	44
4.1.	Asociační analýza polymorfizmu v <i>MC4R</i>	44
4.1.1.	Fenotypová data	44
4.1.2.	Genotypování	45
4.1.3.	Statistické vyhodnocení	45
4.2.	Genetické hodnocení výšky tuku a svalu	46
4.2.1.	Výběr BLUP-AM: data	46
4.2.2.	Výběr BLUP-AM: vyhodnocení	47
4.2.3.	Porovnání BLUP a ssGBLUP: data	49
4.2.4.	Porovnání BLUP a ssGBLUP: vyhodnocení	52
5.	Výsledky a diskuze	55
5.1.	Asociační analýza polymorfizmu v <i>MC4R</i>	55
5.1.1.	Významnost modelu	55
5.1.2.	Významnost genotypu	55
5.2.	Genetické hodnocení výšky tuku a svalu	60
5.2.1.	Navržený BLUP-AM	60
5.2.2.	Populační genetické parametry	65
5.2.3.	Porovnání BLUP a ssGBLUP: mateřská plemena	67
5.2.4.	Porovnání BLUP a ssGBLUP: otcovská plemena	71
5.2.5.	Efekty SNP	73
6.	Závěr	82
7.	Seznam použité literatury	84
8.	Seznam použitých zkratk	105
9.	Seznam příloh	108

1. Úvod

Potravinová bezpečnost je celosvětově aktuálním tématem, které se dotýká dostupnosti a kvality potravin, welfare zvířat či klimatických změn (Garnett et al. 2013). V období 2022–2031 se spotřeba potravin má každoročně zvýšit o 1,4 %, přičemž hlavní příčinou bude růst lidské populace (OECD/FAO 2022). Na zvýšenou poptávku musí reagovat také výroba masa, která zajišťuje nepostradatelný zdroj bílkovin. Celosvětově v roce 2021 průměrná spotřeba čtyř hlavních druhů masa na obyvatele a rok činila téměř 34 kg, z čehož drůbeží představovalo 44 %, vepřové 31 %, hovězí 19 % a skopové 5 %. V zemích Evropské unie byla spotřeba masa dvojnásobná (69 kg) a nejvíce konzumovaným druhem zůstává vepřové s podílem 48 % (OECD/FAO 2022).

Vzhledem ke stoupající poptávce ve spojení s environmentálními nároky je pro živočišnou výrobu jedním z hlavních cílů efektivita (Neeteson-van Nieuwenhoven et al. 2013). Chov prasat názorně ukazuje, jakých pokroků lze v tomto směru dosáhnout. Kupříkladu během let 1998–2018 v českých populacích vzrostl přírůstek do konce testu ze 604 g na 710 g a počet odchovaných selat na prasnici a rok ze 17,5 na 29,3 (SCHP 2021). Klíčovou úlohu v tomto směru hraje šlechtění (Neeteson-van Nieuwenhoven et al. 2013), jehož rozvoj jde ruku v ruce s inovacemi v zootechnice, genetice, biologii, sběru dat, matematicko-statistických metodách, výpočetních technologiích nebo biotechnologiích (Dekkers et al. 2011; Gianola & Rosa 2015; Ibanez-Escriche & Simianer 2016).

Biotechnologie se do moderního šlechtění výrazně promítly technikami jako umělá inseminace, *in vitro* fertilizace či embryotransfer. Další kapitoly otevřely snadno zjiřitelné molekulární markery (Landgridge 2019). Během dvou dekad kolem přelomu století se využití markerů vyvinulo z hledání malého množství příčinných mutací na sběr desítek až stovek tisíc informací napříč celým genomem pomocí DNA čipů. Markery z DNA čipů byly zapojeny do nejlepší lineární nevychýlené předpovědi (BLUP), a tak začala éra genomické selekce (Meuwissen et al. 2016). Takzvaný genomický BLUP animal model se už více jak 10 let využívá u všech hlavních druhů hospodářských zvířat, včetně prasat (Van Eenennaam et al. 2014; Bermann et al. 2022).

V českých populacích prasat zatím probíhá sběr genomických dat (Krupa et al. 2019a) a vznikají první studie pro zavedení genomické selekce (Krupová et al. 2024). Zároveň Svaz chovatelů prasat řeší pokles výšky svalu, který může být důsledkem způsobu odhadu zmasilosti v kombinaci s nízkým selekčním tlakem vůči této vlastnosti. Navrhovaným řešením je nahradit zmasilost v selekčním indexu jejími složkami, tedy výškou hřbetního tuku a výškou svalu. Proto hlavním cílem této práce je navržení vhodného modelu pro přímé genetické hodnocení výšky tuku a svalu a následné rozšíření tohoto modelu o genomická data.

2. Literární přehled

2.1. Užítkovost prasat

Úroveň chovu prasat je hodnocena přibývajícím množstvím ukazatelů. Stále větší důraz je kladen na zdraví a welfare, kam se řadí znaky jako vitalita, robustnost, dlouhověkost, stav končetin, odolnost vůči chorobám nebo poruchy chování. Zároveň jsou sledovány environmentální dopady, jako uhlíková stopa (Kanis et al. 2005; Merks et al. 2012). Tyto znaky tak doplňují tradiční ukazatele užítkovosti, které se dělí na reprodukční a produkční (Stupka et al. 2009) a které zůstávají v popředí zájmu pro svou ekonomickou významnost (Krupa et al. 2019a).

Reprodukční užítkovost zahrnuje počet ovulovaných oocytů, přežitelnost embryí, počet (živě) narozených selat, počet odstavených selat, počet vrhů za rok, zabřezávání, počet struků, mléčnost, libido, kvalita spermatu aj. Produkční užítkovost pak popisují ukazatele výkrmnosti a jatečné hodnoty. Výkrmnost (efektivita růstu) je vyjádřena průměrným denním přírůstkem a spotřebou krmiva na jednotku přírůstku (konverzí). Jatečná hodnota se dále dělí na kvantitativní ukazatele (jatečná výtěžnost, podíl libového masa či podíl hlavních masitých částí) a kvalitativní ukazatele (barva, šťavnatost či tučnost masa; Pulkrábek 2005; Stupka et al. 2009).

Praktická část této práce věnuje pozornost podílu libového masa (také podílu svaloviny nebo zmasilosti, LMP). Tato vlastnost má v Evropské unii zvláštní důležitost, neboť se na jejím základě v rámci systému SEUROP zpeněžují jatečně upravená těla (JUT) prasat (Evropský parlament a Rada EU 2013; Font-i-Furnols et al. 2016).

Kvůli časové a peněžní náročnosti však měření LMP v komerčních podmínkách neprobíhá přímo. Detailní disekce JUT slouží pouze pro navržení regresních rovnic a kalibraci přístrojů, které dokážou s dostatečnou přesností LMP odhadnout. Nejčastěji se proto používají různá měření výšky hřbetního tuku nebo výšky svalu jako *gluteus medius* a *longissimus thoracis et lumborum* (Causer et al. 2003; Pomar et al. 2009; Scholz et al. 2015; Font-i-Furnols et al. 2016; Uniporc 2021). Pro klasifikaci jatečných kusů v Česku bylo schváleno použití šesti metod, přičemž každá z nich obsahuje rovnici se záporným koeficientem pro hřbetní tuk a kladným koeficientem pro sval (Evropská komise 2013). Stejný princip se používá také u živých čistokrevných zvířat na konci unifikovaného testu, kdy je LMP odhadnut na základě dvou sonografických měření pro výšku hřbetního tuku (špeku) a jednoho měření pro výšku svalu (kotlety; SCHP 2010). Tento způsob je dále rozveden v kapitole 4.2.1.

Zjištěné hodnoty sledovaných vlastností, identifikační či doprovodné údaje a rodokmen jsou zaznamenávány do databáze kontroly užítkovosti, kterou spravuje Svaz chovatelů prasat (SCHP). Kontrola užítkovosti tak tvoří podklad pro hodnocení a výběr zvířat (Krupa et al. 2019a).

2.2. Šlechtění prasat v Česku

Pro dosažení efektivní výroby vepřového masa slouží šlechtitelský program (Krupa et al. 2019a). Ve šlechtění prasat se uplatňují dva typy programů: selekční a hybridizační. Selekční program v rámci čistokrevných plemen vytváří přípravný plán tak, aby docházelo ke kumulaci genetického zisku. Hybridizační program pak využívá víceplemenné křížení pro odchov finálních hybridů, neboť se díky celkovému efektu křížení maximalizuje produkce masa. Proto šlechtitelský program prasat stojí na horizontálním a vertikálním členění populace (Whittemore 2006; Stupka et al. 2009).

Horizontálním členěním se rozumí šlechtění mateřských a otcovských plemen. Mateřská vynikají ukazateli reprodukce a výkrmnosti, zatímco u otcovských je kladen důraz na jatečnou hodnotu při zachování velmi dobré výkrmnosti (Stupka et al. 2009). Za posledních 10 let byla v české databázi kontroly užítkovosti evidována dvě mateřská plemena: české bílé ušlechtilé (CLW; 57 % databáze) a česká landrase (CLA; 26 %); a tři otcovská plemena: bílé otcovské (LWS; 6 %), duroc (DU; 4 %), peitrain (PE; 3 %) a jejich hybridní linie (4 %; na základě dat od SCHP).

Vertikální členění tradičně charakterizuje pyramida o třech úrovních. Na jejím vrcholu v nukleových chovech (cca 3 % populace prasnic) probíhá čistokrevná plemenitba mateřských (A, B) a otcovských linií (C, D). Střední část (8 %) představují rozmnožovací chovy $A \times B$ s účelem produkce F1 prasniček. Třetí a největší úroveň (89 %) jsou užítkové chovy využívající křížení $(A \times B) \times (C/D)$ pro produkci finálních hybridů (Whittemore 2006; Stupka et al. 2009). V praxi ovšem dochází ke slučování, podrobnějšímu dělení nebo vyčlenění jednotlivých úrovní. Na úpravách tradiční struktury se výrazně podílejí komerční společnosti (Dekkers et al., 2011).

Šlechtění prasat v Česku zastřešuje národní program CzePig, koordinovaný SCHP. Poradním orgánem je Rada plemenné knihy, který stanovuje, jaké vlastnosti se budou sledovat a jaké jsou jejich cíle. Selektce zvířat pak probíhá na základě selekčního indexu (celkové plemenné hodnoty, CPH) a dalších vlastností či znaků jako exteriér. Současný index pro mateřská plemena zahrnuje přírůstek od narození do konce testu (podíl 28 %), LMP (2 %), počet živě narozených selat ve vrhu (25 %), počet všech selat ve vrhu (15 %), počet dochovaných selat na prasnici a rok (10 %), mezidobí (12 %) a počet struků (8 %). Index otcovských plemen oproti tomu sestává z přírůstku (46 %), LMP (46 %), celkového počtu spermii (4 %) a počtu funkčních spermii (4 %; E. Žáková, osobní komunikace). V tabulce 2.2-A jsou uvedeny vybrané šlechtitelské cíle do roku 2030 (Žáková & Stibal 2021) a skutečný stav populací v roce 2023 (na základě dat od SCHP).

Tabulka 2.2-A: Vybrané šlechtitelské cíle pro jednotlivá plemena do roku 2030.

Plemeno	Živě narozená selata		Přírůstek [g/den]		Podíl libového masa [%]	
	<i>cíl</i>	<i>skutečnost</i>	<i>cíl</i>	<i>skutečnost</i>	<i>cíl</i>	<i>skutečnost</i>
CLW	15,5	14,5	750	698	58,0	60,1
CLA	14,8	14,4	770	723	58,0	60,5
DU	-	-	780	715	60,5	59,9
LWS	-	-	790	728	61,5	60,7
PE	-	-	740	750	63,0	63,1

Zkratky: CLA = česká landrase; CLW = české bílé ušlechtilé;

DU = duroc; LWS = bílé otcovské; PE = pietrain

2.3. Matematicko-statistické základy šlechtění

2.3.1. Rozložení proměnlivosti

Šlechtitelský cíl vyjadřuje představu o tom, jaká zvířata jsou v daných podmínkách považována za nejlepší, aby byla využita pro tvorbu následujících generací (Bourdon, 2014). Základním předpokladem pro srovnání a výběr jedinců je existence populační proměnlivosti (ve statistickém pojetí přesněji rozptylu neboli variance, σ^2). V genetice kvantitativních vlastností se celkový rozptyl užitkovosti (fenotypový, σ_p^2) rozkládá na část založenou geneticky (genotypovou, σ_G^2) a část ovlivněnou prostředím (σ_E^2). Genotypový rozptyl sestává z aditivity (σ_a^2), dominance (σ_d^2) a interakce (epistáze, σ_i^2). Rozptyl prostředí zahrnuje vlivy trvalé (systematické, permanentní; σ_{Ep}^2) a dočasné (nesystematické, temporární, σ_{Et}^2). Trvalé prostředí se podrobněji dělí na vnitřní (věk, pohlaví, vrh...) a vnější (stádo, rok, období, typ krmení...). Dočasné prostředí většinou není možné zaznamenat, a proto spadá do náhodné chyby (Jakubec et al. 2003; Bourdon 2014).

V závislosti na studovaných vlastnostech a podmínkách je dále potřeba vzít v úvahu zvláštní složky celkové variability. Velký význam může mít rozptyl maternálních efektů, jimiž matka v prenatalním či postnatalním stádiu vývoje ovlivňuje (budoucí) sledovanou užitkovost potomka. Jejich zanedbání potenciálně vede k nadhodnocení odhadu aditivního genetického rozptylu (Kruuk 2004). Mezi maternální efekty patří: maternální genetický efekt (σ_m^2), který je založen na aditivně působících genech spojených s reprodukcí matky (např. geny pro mléčnost); a efekt trvalého prostředí matky (σ_{Epm}^2), který je zakotven ve složkách prostředí (např. úroveň odchovu; Kruuk & Hadfield 2007). Jelikož trvalé prostředí matky působí na skupinu potomků systematicky, je v níže uvedeném zobecnění zahrnuto do σ_{Ep}^2 .

S přihlédnutím k tomuto členění se v uvedené práci aditivním genetickým rozptylem (σ_a^2) rozumí rozptyl přímého genetického efektu jedince pro zkoumanou vlastnost. Maternální genetický efekt je sice také založen aditivně, ale danou vlastnost ovlivňuje nepřímo – prostřednictvím genů

matky pro jiné vlastnosti. Zároveň mezi těmito složkami může existovat závislost, přesněji kovariance ($\sigma_{a \times m}$), která vyjadřuje podobnost mezi matkou a potomkem (Willham 1972).

Zohlednit lze také další kovariance jako interakci mezi genotypem a prostředím. Avšak ta je obvykle zanedbávána bez vážnějších důsledků (Falconer 1989), a proto nebude sledována ani v této práci. Zároveň jsou σ_a^2 , σ_i^2 a σ_{Et}^2 souhrnně považovány za složky chybového rozptylu (reziduálního; σ_e^2). Fenotypový rozptyl je tedy v této práci rozložen následovně:

$$\sigma_p^2 = \sigma_G^2 + \sigma_E^2 = \sigma_a^2 + \sigma_m^2 + \sigma_{a \times m} + \sigma_{Ep}^2 + \sigma_e^2. \quad [1]$$

Z pohledu čistokrevné plemenitby je nejdůležitější složkou aditivní genetický účinek. Jedná se o sumu nezávislých efektů všech genů jedince, jež ovlivňují určitou vlastnost. Odchylka této sumy od průměru populace se nazývá aditivní genetická hodnota nebo nejčastěji plemenná hodnota (Bourdon 2014). Aditivita popisuje většinu podobnosti mezi příbuznými jedinci a dá se využít pro tvorbu genetického pokroku, neboť se předává ve sledu generací. Aditivní genetický rozptyl obvykle představuje přes 50 % a někdy téměř 100 % genetického rozptylu (Hill et al. 2008; Su et al. 2012).

2.3.2. Populační genetické parametry

Rozlišení složek proměnlivosti vede k odhadu populačních genetických parametrů: koeficientu dědivosti, koeficientu opakovatelnosti a genetické korelace. **Koeficient dědivosti** (heritability, h^2) obecně vyjadřuje poměr genotypové proměnlivosti ku fenotypové čili determinační koeficient mezi genotypovými hodnotami a užitkovostí (Bourdon 2014). Ale v této práci bude koeficient dědivosti dále chápán pouze v užším a praktičtějším pohledu aditivního působení genů, tedy:

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}. \quad [2]$$

Někteří autoři (Miar et al. 2014; Lukač 2016) pak používají upravený výpočet, kdy je proměnlivost v užitkovosti očištěna o systematické efekty. V tomto případě je výpočet následující:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_m^2 + \sigma_{a \times m} + \sigma_e^2}. \quad [3]$$

Whittemore (2006) uvádí, že typické reprodukční ukazatele jsou nízce dědivé (0,10–0,20), zatímco ukazatele výkrmnosti a jatečné hodnoty nabývají středních až středně vysokých hodnot (0,30–0,70).

Pokud jedinci mají pro danou vlastnost více měření, využívá se **koeficient opakovanosti** (w^2) neboli korelace mezi opakovanými měřeními. Ten je definován jako poměr proměnlivosti v produkční schopnosti ($\sigma_{PS}^2 = \sigma_a^2 + \sigma_{Ep}^2$) ku fenotypové proměnlivosti:

$$w^2 = \frac{\sigma_{PS}^2}{\sigma_P^2} = \frac{\sigma_a^2 + \sigma_{Ep}^2}{\sigma_P^2}. \quad [4]$$

Ukazatele jako velikost vrhu nebo hmotnost selat dosahují hodnot opakovatelnosti 0,15–0,30 (Bourdon 2014).

Při analýze více vlastností současně jsou hodnoceny také (aditivní) **genetické korelace** (r_a), které udávají průměrnou sílu vztahu mezi aditivními genetickými efekty:

$$r_a = \frac{\sigma_{a(XY)}}{\sigma_{a(X)}\sigma_{a(Y)}}, \quad [5]$$

kde $\sigma_{a(XY)}$ značí kovarianci mezi vlastnostmi X a Y , $\sigma_{a(X)}$ směrodatnou odchylku vlastnosti X a $\sigma_{a(Y)}$ směrodatnou odchylku vlastnosti Y . Existence tohoto vztahu je vysvětlována pleiotropními efekty genů pro užitkové vlastnosti a genovými vazbami. Jelikož se síla genové vazby může měnit, je i genetická korelace pro danou dvojici vlastností populačně specifická (Xu 2022).

2.3.3. Smíšený model

Odhad složek rozptylu, a přeneseně také populačních genetických parametrů, je prvním krokem genetického hodnocení ve šlechtění. Nejjednodušší metody odhadu využívají obecné lineární modely (GLM) jako regresi a analýzu rozptylu (ANOVA), přičemž často staví na podobnosti blízkce příbuzných jedinců nebo na selekčních pokusech (Xu 2022). Ovšem i velmi komplexní GLM, který je zpravidla chápán jako model s pevnými efekty, mimo jiné předpokládá nezávislost pozorování, nebo přesněji nezávislost reziduálních chyb odhadů (Gelman & Hill 2007; Lazic 2010). V kontextu šlechtění zvířat inherentně dochází k porušení tohoto předpokladu, neboť jedinci jsou geneticky příbuzní a zároveň se v reálné populaci často tvoří hierarchické struktury (Xu 2022). Porušení nezávislosti pozorování (replikátů) vede k pseudoreplikaci, která značí umělé zvětšení výběrového souboru (Lazic 2010). Z toho důvodu v dnešní šlechtitelské praxi prosazují sofistikovanější přístupy, tedy především smíšené modely řešené spojenou metodologií maximální věrohodnosti a nejlepší lineární nevychýlené předpovědi, případně bayesovské hierarchické modely (Gianola & Rosa 2015).

Zobecněný lineární smíšený model (GLMM) se často označuje také jako víceúrovňový či hierarchický model (pro skupinová neboli klastrovaná data). Stejně tak se v kontextu opakovaných měření (panelových dat) můžeme setkat s longitudinálním modelem (Gelman & Hill 2007). V každém případě je jeho principem současné zahrnutí pevných i náhodných efektů (někdy též pevných a náhodných faktorů). Maticově lze smíšený model zapsat jako:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}, \quad [6]$$

kde \mathbf{y} je vektor pozorování pro danou vlastnost, \mathbf{b} je vektor neznámých pevných efektů, \mathbf{u} je vektor neznámých náhodných efektů, \mathbf{e} je vektor náhodných chyb a \mathbf{X} , \mathbf{Z} jsou matice plánu pokusu (incidenční, designové), které přiřazují prvky v \mathbf{b} a \mathbf{u} k prvkům v \mathbf{y} . Pro \mathbf{y} , \mathbf{u} a \mathbf{e} platí očekávané hodnoty (E) a variačně-kovarianční struktura (V):

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Xb} \\ 0 \\ 0 \end{bmatrix}, V \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{ZGZ}^T + \mathbf{R} & \mathbf{ZG} & \mathbf{R} \\ \mathbf{GZ}^T & \mathbf{G} & 0 \\ \mathbf{R} & 0 & \mathbf{R} \end{bmatrix}, \quad [7]$$

kde \mathbf{G} je kovarianční matice pro náhodné efekty a \mathbf{R} je kovarianční matice náhodných chyb neboli reziduí (Kaps & Lamberson 2004). Klíčová je přítomnost matice \mathbf{G} , z níž vyplývá, že na rozdíl od modelu jen s pevnými efekty nelze všechny efekty shrnout do jednoho vektoru \mathbf{b} pomocí souhrnné matice \mathbf{X} .

Pojmenování efektů jako pevných a náhodných souvisí s tím, co je matematicky odhadováno. Model s **pevnými efekty** lze vyjádřit jako vícenásobnou lineární regresi (s využitím umělých neboli dummy proměnných pro kategorické proměnné), kde odhadovanými parametry jsou jednotlivé regresní koeficienty (\mathbf{b}). Tyto efekty jsou proto napříč jedinci, případně skupinami, pevné ve smyslu neměnné a veškerá proměnlivost je připsána reziduálním chybám (\mathbf{e} ; Kreft & de Leeuw 1998; Gelman 2005). Odhady jsou získány metodou nejmenších čtverců (LSE) nebo obecněji metodou maximální věrohodnosti (MLE; Robinson 1991). Pevné efekty jsou typicky odhadovány pro faktor, který má malý a předem známý počet úrovní, jež se při opakování pokusu objeví znovu. Příklady takových faktorů v kontextu šlechtění jsou plemeno, pohlaví, technika krmení nebo způsob ošetření (Urban 2006; Schaeffer 2010; Agresti 2019).

Oproti tomu **náhodný efekt** značí, že daná úroveň faktoru byla náhodně vybrána ze souboru (populace) s velkým až nekonečným množstvím úrovní. Náhodné efekty se mezi jedinci či skupinami mění, a proto lze odhadnout pouze parametry popisující jejich rozdělení (rozptyly), případně parametry popisující jejich závislost (kovariance). Tyto odhady popisují část proměnlivosti, která není považována za reziduální chybu, a tudíž mohou být zobecněny pro celý základní soubor (Gelman 2005; Xu 2022). Zároveň spotřebují méně stupňů volnosti (Gomes 2022). Konkrétní hodnoty lze náhodným efektům předpovědět až na základě parametrů rozdělení, obvykle pomocí nejlepší lineární nevychýlené předpovědi (Robinson 1991). Nejdůležitější náhodný efekt ve šlechtění představuje jednotlivé zvíře. Dalšími příklady jsou rodiny, skupiny vrstevníků nebo obecně skupiny se stejnými podmínkami prostředí až celé chovy (Schaeffer 2018; Xu 2022). Avšak jak je popsáno v kapitole 2.4.2, v posledních dvou případech závisí statistická definice efektu na dané situaci.

Modely pouze s pevnými efekty nacházejí hojně využití v jednodušších analýzách, neboť jsou výpočetně snadné (LSE) a často také efektivní. Nicméně smíšené modely jsou obvykle správnější, protože jsou univerzální. Tuto výhodu přibližují Gelman & Hill (2007) nebo Kreft & de Leeuw (1998) propojením s bayesovskou statistikou, v jejímž rámci se pevné efekty dají považovat za speciální (extrémní) případy náhodných efektů. Pevné efekty vyplývají buďto z kompletního pooling (nulový rozptyl efektů → úplně sdružená data bez podskupin → jeden regresní model), nebo neexistujícího pooling (nekonečný rozptyl efektů → n podskupin analyzováno nezávisle → n regresních modelů). Prostor mezi kompletním a neexistujícím poolingem vyplňuje parciální pooling, ve kterém jsou sdíleny informace mezi skupinami. Parciální pooling tak poskytuje stabilnější odhady (předpovědi) než neexistující pooling a zajímavější parametry než kompletní pooling. Jinými slovy, hodnota řešení efektu jako pevného bude vyšší než jeho řešení jako náhodného, zvláště pokud některé úrovně faktoru zahrnují nízký počet pozorování nebo se v nich vyskytují extrémní hodnoty. Zato předpovědi náhodných efektů jsou i v takových případech smrštěny k průměru (shrinkage). Zároveň náhodné efekty umožňují prostřednictvím kovariancí zohlednit vztah mezi pozorováními.

Tabulka 2.3-A: Srovnání pevných a náhodných efektů.

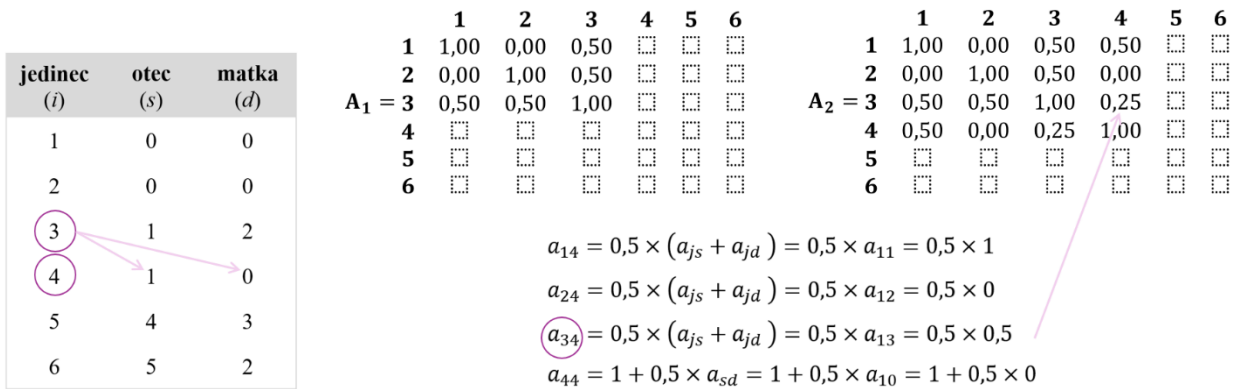
Rozdíl	Pevné efekty	Náhodné efekty
koeficienty	konstantní	proměnlivé (z normálního rozdělení)
závislost pozorování (reziduí)	porušuje předpoklad	lze využít
charakter tříd	malý počet, nemění se s opakováním pokusu	větší počet, náhodně vybrané z populace
inference	pro velikost efektu v daném experimentu	pro celou populaci (v případě variance)
metoda řešení	LSE, MLE	MLE (variance), BLUP (efekty)
výhody	jednoduchý test velikosti konkrétního efektu, kontrola časově neměnných faktorů (které by mohly maskovat vztah mezi hlavními proměnnými)	obecnější, obvykle přesnější, zohlednění nepozorované heterogenity, méně stupňů volnosti, umožňují menší skupiny
příklady	pohlaví, počet laktací, věk, typ krmiva, plemeno	jedinec, matka, vrh, stádo-rok-období

Kovariance mohou být modelovány jak v matici **G**, tak v matici **R**. Použit lze řadu různých struktur, které jsou zpravidla prezentovány na případu modelů s opakovanými měřeními nebo na víceznakových modelech, přičemž v obou případech jsou měření korelována přes jedince. *Nejjednodušší kovarianční struktura* má podobu $\mathbf{I}\sigma^2$, kde σ^2 značí homogenní rozptyl a **I** značí jednotkovou matici, takže kovariance mezi pozorováními jsou nulové. Rozměry **I** v případě matice **G** odpovídají počtu úrovní faktoru, kdežto v případě matice **R** počtu pozorování. *Komponovaná symetrická matice* obsahuje nejen homogenní rozptyly, ale také nenulové homogenní kovariance, takže korelace mezi dvěma měřeními zůstává stejná, nehledě na jejich vzdálenosti. Z pohledu běžných opakovaných měření je velmi praktickou strukturou *autoregresní matice*, pro kterou rovněž stačí odhadnout pouze dva parametry: homogenní rozptyl (σ^2) a autoregresní koeficient ($\rho^t < 1$). Exponent t se rovná vzdálenosti (počtu časových bodů) mezi dvojicí měření, takže se korelace snižuje s přibývajícím rozstupem mezi měřeními. *Toeplitzova matice* se od předchozí liší tím, že je pro každou vzdálenost mezi měřeními odhadnuta vlastní kovariance. Nejvíce různých parametrů obsahuje *nestrukturovaná matice*, ve které jsou všechny rozptyly i kovariance heterogenní. Tato matice zároveň nachází uplatnění v přístupu náhodné regrese, kdy jsou modelovány variabilní průsečíky či sklony (Littell et al. 2000; Kaps & Lamberson 2004).

Ve šlechtění zvířat se pak obzvláště uplatňuje **matice aditivní genetické příbuznosti** neboli rodokmenová matice (**A**), která je základem takzvaného animal modelu. V nejjednodušším případě tak platí: $\mathbf{G} = \mathbf{A}\sigma_a^2$. Animal model v konečném důsledku umožňuje předpovědět efekt každého jedince, který je zahrnut v matici **A**. Zároveň nezáleží na tom, zda daný jedinec disponuje pozorováním pro vlastní užitek (např. potenciální mléčná užitek býka na základě užitečnosti jeho dcer). Díky kvantifikovaným příbuzenským vztahům jsou pozorování propojena napříč jedinci, takže vzniká obdoba modelu pro opakovaná „nepřímá“ měření (Mrode 2005; Xu 2022).

Tvorba matice **A** vychází ze sloupcového zápisu rodokmenu (jedinec, otec, matka) s kódováním $i \in \langle 1; n \rangle$ tak, aby rodiče předcházeli potomky. Výpočet prvků této matice probíhá rekurzivní metodou a je založen na teorii alel identických původem (IBD), tedy zděděných od společného předka. Diagonální prvky (a_{ii}) obsahují koeficient příbuzenské plemenitby (inbreedingu; F_i), který pro jedince i vyjadřuje pravděpodobný podíl IBD alel na stejném lokusu: $a_{ii} = 1 + F_i = 1 + 0,5 \times a_{sd}$, kde a_{sd} značí příbuznost mezi otcem a matkou daného jedince. Mimodiagonální prvky (a_{ij}) vyjadřují příbuznost mezi jedinci i a j se zohledněním inbreedingu: $a_{ij} = a_{ji} = 0,5 \times (a_{js} + a_{jd})$, kde a_{js} je příbuznost i -tého jedince (v řádku) s otcem j -tého jedince (ve sloupci) a a_{jd} je příbuznost i -tého jedince s matkou j -tého jedince (Bourdon 2014; Mrode 2005). Příklad výpočtu je uveden na obrázku 2.3-A.

Obrázek 2.3-A: Doplnění matice A na příkladu výpočtu příbuzností pro jedince 4 s jedinci 1, 2 a 3. Zleva rodokmen, matice na začátku kroku (A_1), použité výpočty a matice na konci kroku (A_2).



2.3.4. REML

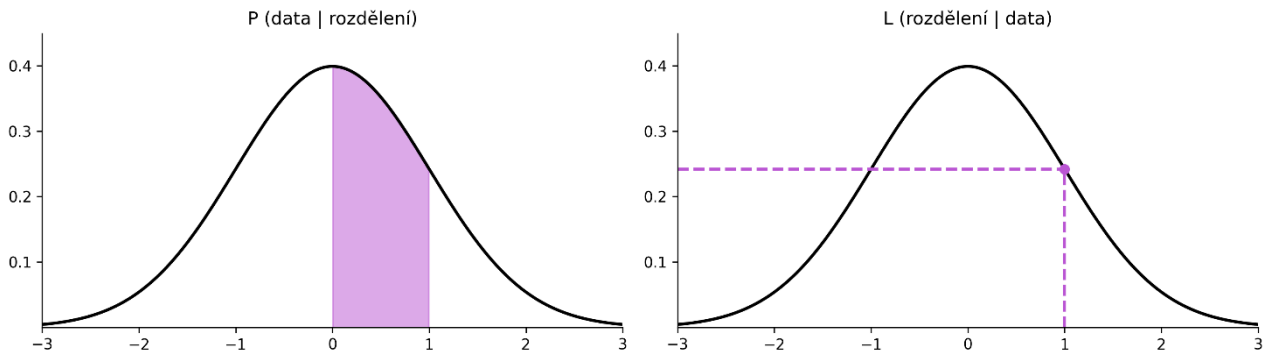
Jak bylo nastíněno v předchozí podkapitole, navržený smíšený model v první řadě slouží k odhadu pevných efektů a (ko)variancí náhodných efektů. Pro takzvané fitování modelu, tedy nalezení co nejlepších parametrů modelu ve vztahu k pozorovaným datům, se nejčastěji používají metody maximální věrohodnosti. Věrohodnost $L(\theta|y)$ vyjadřuje „správnost“ zvolených parametrů θ (popisujících určité rozdělení) při pevně daných hodnotách náhodné proměnné y (obrázek 2.3-B). Tím se liší od pravděpodobnosti $P(y|\theta)$, která udává, do jaké míry lze očekávat hodnoty y za předpokladu pevně daného rozdělení. Základní věrohodnostní funkce pro n nezávislých pozorování proměnné y v normálním rozdělení má tvar (Kaps & Lamberson 2004; Xu 2022):

$$L(\mu, \sigma^2 | y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu)^2}{2\sigma^2}}. \quad [8]$$

Uvážíme-li nejjednodušší animal model $y = 1\mu + a + e$, kde $y \sim N(\mathbf{Xb}, \mathbf{ZGZ}^T + \mathbf{R})$; $\mathbf{Xb} = 1\mu$ (μ vyjadřuje obecný průměr); $\mathbf{G} = \mathbf{A}\sigma_a^2$; a $\mathbf{R} = \mathbf{I}\sigma_e^2$, potom za předpokladu vícerozměrného normálního rozdělení (MVN) platí:

$$\begin{aligned}
 L(\mu, \sigma_a^2, \sigma_e^2 | y_1, \dots, y_n) & \quad [9] \\
 &= \frac{1}{\sqrt{(2\pi)^n |\mathbf{ZGZ}^T + \mathbf{R}|}} \exp \left\{ -\frac{1}{2} (y - \mathbf{Xb})^T (\mathbf{ZGZ}^T + \mathbf{R})^{-1} (y - \mathbf{Xb}) \right\} \\
 &= \frac{1}{\sqrt{(2\pi)^n |\mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2|}} \exp \left\{ -\frac{1}{2} (y - 1\mu)^T (\mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2)^{-1} (y - 1\mu) \right\}
 \end{aligned}$$

Obrázek 2.3-B: Pravděpodobnost (plocha pod křivkou daného normálního rozdělení) vs. věrohodnost (hodnota hustoty pravděpodobnosti odpovídající dané hodnotě náhodné proměnné).



Maximalizace věrohodnosti tedy znamená nalezení takových parametrů θ , s nimiž dojde k maximalizaci hodnoty funkce hustoty pravděpodobnosti. Pro tento výpočet se využívá logaritmus věrohodnostní funkce (LL), díky němuž lze pracovat se sumou prvků namísto jejich produktu, a tak je odstraněn problém extrémně nízké hodnoty hustoty pravděpodobnosti při vysokém počtu pozorování. V předchozím příkladu je tedy potřeba maximalizovat funkci:

$$LL(\mu, \sigma_a^2, \sigma_e^2 | y_1, \dots, y_n) = -\frac{1}{2} \ln |\mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2| - \frac{1}{2} (\mathbf{y} - \mathbf{1}\mu)^T (\mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2)^{-1} (\mathbf{y} - \mathbf{1}\mu) \quad [10]$$

Metoda maximální věrohodnosti (ML) předpokládá, že pevné efekty jsou odhadnuté bez chyby. Tento předpoklad ovšem většinou neplatí, tudíž se odhady složek rozptylu pro náhodné efekty stávají vychýlenými (podhodnocenými). Z toho důvodu byla zavedena metoda restringované maximální věrohodnosti (**REML**), která maximalizuje pouze tu část věrohodnosti, která není závislá na pevných efektech. Tím oproti ML vzniká nevýhoda, že nelze srovnávat věrohodnost modelů, které se liší pevnými efekty. K „odstranění“ pevných efektů slouží transformační matice \mathbf{K}^T tak, aby $\mathbf{K}^T \mathbf{X} = 0$. Odhad rozptylu tak ztrácí jeden stupeň volnosti, čímž je potlačeno podhodnocení. Pro obecný smíšený model platí:

$$\mathbf{K}^T \mathbf{y} = \mathbf{K}^T (\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}) = \mathbf{K}^T \mathbf{Z}\mathbf{u} + \mathbf{K}^T \mathbf{e}, \quad [11]$$

takže zlogaritmovaná věrohodnost pro REML nabývá tvaru:

$$LL_R(\mathbf{K}^T \mathbf{y}) = -\frac{1}{2} \ln |\mathbf{K}(\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R})\mathbf{K}^T| - \frac{1}{2} \mathbf{y}\mathbf{K}^T [\mathbf{K}(\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R})\mathbf{K}^T]^{-1} \mathbf{K}\mathbf{y}, \quad [12]$$

kde první část penalizuje složitost modelu a druhá část představuje odchylku pozorovaných hodnot od předpovědí modelu čili vhodnost modelu (Lynch & Walsh 1998; Xu 2022).

Pro REML (podobně jako pro ML) existuje mnoho výpočetních metod. Zpravidla se jedná o iterativní metody s využitím derivací, neboť hodnota funkce v bodě, kde se první derivace rovná nule, značí její maximum. Misztal (2008) v kontextu šlechtění zvířat charakterizoval tři základní

algoritmy, které se primárně liší využitím (řádu) derivace. Algoritmus DF (derivative-free čili bez derivace) je nejjednodušší, ale v dnešní době se používá zřídka kvůli složitosti modelů. Naopak velmi populárním algoritmem je EM (expectation-maximization čili očekávání-maximalizace), který využívá derivaci prvního řádu. Jeho výhodou je spolehlivost a robustnost, nevýhodou pak vysoký počet potřebných iterací. Z algoritmů založených na derivaci druhého řádu je nejvíce oblíbený AI (average information čili průměrná informace), který je oproti EM obvykle mnohem rychlejší.

V této práci je využita metoda **EM REML**, již přiblížili třeba Taylor & Everett (1985). V prvním kroku jsou zvoleny počáteční hodnoty rozptylů (např. σ_a^2 , σ_{Ep}^2 a σ_e^2 odpovídající náhodným efektům aditivně genetickým, trvalého prostředí, respektive reziduálních chyb). Samotné iterace pak probíhají cyklickým opakováním kroků E a M. Krok E spočívá ve výpočtu očekávaných hodnot náhodných efektů na základě naměřených hodnot a aktuálního nastavení parametrů (hodnot rozptylů). V kroku M pomocí získaných očekávaných hodnot dochází k úpravě parametrů tak, aby byla maximalizována věrohodnost. Tyto kroky se opakují, dokud není dosaženo konvergence, tedy nejvyššího přípustného rozdílu v hodnotách parametrů mezi dvěma iteracemi.

Je vhodné dodat, že REML často bývá nahrazován Gibbsovým vzorkováním, které patří do rodiny Markovových řetězců Monte Carlo čili do bayesovské metodologie. Zatímco REML náleží mezi deterministické metody, tedy při opakování výpočtu se stejnými vstupními hodnotami dojde stejnou cestou ke stejnému výsledku, Gibbsovo vzorkování je stochastickou metodou na principu náhodné procházky parametrickým prostorem. Každý krok tedy závisí pouze na současné pozici, nikoliv na sledu předchozích pozic. V každém kroku je vybrán jeden z parametrů θ_i , pro který je vygenerována nová (náhodná) hodnota z podmíněného rozdělení pravděpodobnosti $P(\theta_i | \theta_j, \dots, \theta_n, D)$, tedy v závislosti na fixních hodnotách ostatních parametrů a naměřených datech. Tento nově vzniklý vzorek parametrů (nová hodnota parametru θ_i v kombinaci s nezměněnými hodnotami parametrů θ_j až θ_n) pak představuje výchozí pozici pro další krok. Po dostatečném počtu iterací a vyřazení vzorků z počáteční nestabilní fáze (zahoření, burn-in) mohou být požadované složky rozptylu odhadnuty například jako průměr zbývajících vzorků (Kruschke 2015). Gibbsovo vzorkování se nejen ve šlechtění zvířat těší rostoucí oblibě pro snadno uchopitelnou teorii, a to i v případě složitých problémů jako vysoce hierarchických modelů nebo chybějících údajů. Oproti EM algoritmu je také možno přímo vypočítat střední chyby odhadů. Na druhou stranu kvalita odhadů záleží na nastavení procesu, jehož optimalizace není přesně dána (Misztal 2008).

2.3.5. BLUP

Závěrečným krokem genetického hodnocení za účelem výběru nejvhodnějších zvířat je předpověď plemenných hodnot, což ze statistického hlediska znamená získat řešení pro náhodné (aditivní genetické) efekty. V roce 1949 za tímto účelem Charles Roy Henderson poprvé představil rovnice smíšeného modelu (MME). Pro výpočet MME se následně vžil pojem nejlepší lineární nevychýlená předpověď (**BLUP**; Schaeffer 1991). BLUP byla použita pro zobecnění selekčních indexů, jejichž cílem je odhadnout genetickou hodnotu zvířete jako sumu vážených informací o vlastní užitkovosti a užitkovosti příbuzných jedinců (Xu 2022). Metoda BLUP je označována jako *nejlepší*, neboť maximalizuje korelace mezi skutečnými a předpovězenými hodnotami efektů. K předpovědi využívá *lineární* funkce pozorování. Předpovědi jsou *nevychýlené*, protože průměrná očekávaná hodnota efektu je rovna jeho skutečné hodnotě. A termín *předpověď* označuje řešení náhodných efektů proto, že odhadnout náhodné efekty podobně jako pevné efekty nelze. Přesto se kvůli historickému vývoji v anglicky psané literatuře zažila zkratka EBV pro estimated breeding value (Mrode 2005), která přešla do češtiny jako OPH pro odhadnutou plemennou hodnotu (Jakubec et al. 1999).

Předpovědi náhodných efektů se vyskytly už v předchozí kapitole jako součást REML, jelikož iterativní metody jako EM REML k odhadu složek rozptylu potřebují právě BLUP. Výstup z REML tedy obsahuje plemenné hodnoty jako vedlejší produkt. Ovšem ve šlechtitelské praxi jsou rozlišovány programy s prvotním účelem odhadu (ko)variancí (REML, Gibbsovo vzorkování), na něž navazují programy pro předpověď plemenných hodnot (BLUP) v tom smyslu, že pro výpočet MME je potřeba zadat předem známé (ko)variance. Výhoda tohoto postupu spočívá ve skutečnosti, že REML bývá výpočetně mnohem náročnější než BLUP a zároveň populační genetické parametry není potřeba aktualizovat zdaleka tak často jako plemenné hodnoty. Nicméně také samotná BLUP je náročná na výpočetní kapacitu, zejména kvůli potřebě invertovat takzvanou levou stranu MME. Specializované programy pro BLUP proto nabízejí různé algoritmy k řešení tohoto problému. Softwarové balíčky pro šlechtitele zvířat vyvíjejí například vědci na Georgijské univerzitě (BLUPF90; Misztal et al. 2022), Finském institutu přírodních zdrojů (MiX99; Lidauer et al. 2019) nebo Wageningenské univerzitě (MixBLUP; Vandenplas et al. 2022).

V současné šlechtitelské praxi je BLUP využívána pro různé varianty animal modelu čili pro modely s maticí příbuznosti (Bourdon 2014). Níže je přiblíženo řešení čtyř animal modelů podle Mrodeho (2005), jejichž principy se uplatňují v této práci.

1) *Nejjednodušší animal model po vzoru [6]:* $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}, \quad [13]$$

$$\mathbf{V} \begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 \\ 0 & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

kde $\hat{\mathbf{b}}$ je vektor odhadů pro pevné efekty, $\hat{\mathbf{a}}$ je vektor předpovědí pro náhodné (přímé) aditivní genetické efekty, \mathbf{X} a \mathbf{Z} jsou designové matice těchto efektů, \mathbf{R} je kovarianční matice náhodných chyb, \mathbf{e} je vektor náhodných chyb, \mathbf{G} je obecná kovarianční matice pro náhodné efekty (mimo chyb), \mathbf{y} je vektor pozorování, \mathbf{V} je kovarianční struktura náhodných efektů včetně chyb, \mathbf{A} je matice rodokmenové příbuznosti, \mathbf{I} je jednotková matice, σ_a^2 je (přímý) aditivní genetický rozptyl a σ_e^2 je reziduální rozptyl. Jelikož \mathbf{R} je v tomto případě jednotková, lze [13] dále zjednodušit na:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \mathbf{A}^{-1} \sigma_e^2 / \sigma_a^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \end{bmatrix}.$$

2) *Animal model se společným (trvalým) prostředím jedinců:* $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{e}$:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} & \mathbf{X}^T \mathbf{W} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \mathbf{A}^{-1} \sigma_e^2 / \sigma_a^2 & \mathbf{Z}^T \mathbf{W} \\ \mathbf{W}^T \mathbf{X} & \mathbf{W}^T \mathbf{Z} & \mathbf{W}^T \mathbf{W} + \mathbf{I} \sigma_e^2 / \sigma_c^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \\ \mathbf{W}^T \mathbf{y} \end{bmatrix}, \quad [14]$$

$$\mathbf{V} \begin{bmatrix} \mathbf{a} \\ \mathbf{c} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 & 0 \\ 0 & \mathbf{I}\sigma_c^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

kde nejmenovaná společná označení odpovídají případu 1), $\hat{\mathbf{c}}$ je vektor předpovědí pro náhodné efekty společného prostředí, \mathbf{W} je incidenční matice těchto efektů a σ_c^2 je odpovídající rozptyl.

3) *Animal model s maternálním genetickým efektem:* $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wm} + \mathbf{e}$:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} & \mathbf{X}^T \mathbf{W} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \mathbf{A}^{-1} \sigma_e^2 / \sigma_a^2 & \mathbf{Z}^T \mathbf{W} + \mathbf{A}^{-1} \sigma_e^2 / \sigma_{am} \\ \mathbf{W}^T \mathbf{X} & \mathbf{W}^T \mathbf{Z} + \mathbf{A}^{-1} \sigma_e^2 / \sigma_{am} & \mathbf{W}^T \mathbf{W} + \mathbf{A}^{-1} \sigma_e^2 / \sigma_m^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \\ \mathbf{W}^T \mathbf{y} \end{bmatrix}; \quad [15]$$

$$\mathbf{V} \begin{bmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & 0 \\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

kde nejmenovaná společná označení odpovídají případu 1), $\hat{\mathbf{m}}$ je vektor předpovědí pro náhodné maternální genetické efekty, \mathbf{W} je incidenční matice těchto efektů, σ_m^2 je odpovídající rozptyl a σ_{am} je aditivní genetická kovariance mezi přímými a maternálními efekty.

4) *Víceznakový animal model*: $\mathbf{y}_i = \mathbf{X}_i \mathbf{b}_i + \mathbf{Z}_i \mathbf{a}_i + \mathbf{e}_i$:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}; \quad [16]$$

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \vdots \\ \hat{\mathbf{b}}_i \end{bmatrix}, \hat{\mathbf{a}} = \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \vdots \\ \hat{\mathbf{a}}_i \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & & \\ & \ddots & \\ & & \mathbf{X}_i \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}_i \end{bmatrix}, \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_i \end{bmatrix};$$

$$\mathbf{V} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_i \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_{a_{11}}^2 & \cdots & \mathbf{A}\sigma_{a_{1i}} \\ \vdots & \ddots & \vdots \\ \mathbf{A}\sigma_{a_{i1}} & \cdots & \mathbf{A}\sigma_{a_{ii}}^2 & & & \\ & & & \sigma_{e_{11}}^2 & \cdots & \sigma_{e_{1i}} \\ & & & \vdots & \ddots & \vdots \\ & & & \sigma_{e_{i1}} & \cdots & \sigma_{e_{ii}}^2 \end{bmatrix},$$

kde společná označení odpovídají případu 1) a dolní index 1 až i značí jednotlivé znaky (vysvětlované proměnné).

Pro selekci zvířat je vedle výsledné plemenné hodnoty důležitá také přesnost (nebo spolehlivost) její předpovědi. **Přesnost** se značí r , jelikož jde teoreticky o korelaci mezi skutečnými (TBV) a předpovězenými plemennými hodnotami (EBV). Zaveden byl také termín **spolehlivost**, který je druhou mocninou přesnosti (determinačním koeficientem předpovědi), tudíž se značí r^2 (Jakubec et al. 1999). Spolehlivost EBV pro jedince i Henderson vyjádřil jako:

$$r_i^2 = 1 - \frac{d_i \sigma_e^2}{\sigma_a^2} = 1 - \frac{PEV_i}{\sigma_a^2}, \quad [17]$$

kde d_i je diagonálním prvkem invertované levé strany MME, přesněji submatice $\mathbf{Z}^T \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_e^2}{\sigma_a^2}$, a PEV_i je zkratkou pro rozptyl chyby předpovědi jedince i . PEV tedy vyjadřuje složku aditivního genetického rozptylu, která není vysvětlena předpovědí (Mrode 2005).

Hodnota d_i , potažmo spolehlivost, závisí na vhodnosti modelu, počtu pozorování (jak od samotného jedince, tak od jeho příbuzných), dědivosti (popř. opakovatelnosti) a blízkosti příbuzenských vztahů. Například při dědivosti 0,30 a jednom pozorování vlastní užitkovosti očekáváme spolehlivost 0,30. Při stejné dědivosti a pozorováních od 20 potomků vzroste spolehlivost na 0,62 (Bourdon, 2014). Pro odhad spolehlivosti jsou vyvíjeny různé aproximační metody na

iterativní nebo vzorkovací bázi, protože ve velkých populacích je přímá inverze levé strany MME nepraktická až neproveditelná (Misztal & Wiggans 1988; Meyer 1989; Hickey et al. 2009).

2.4. Výběr vhodného modelu

Pro šlechtitele je nejlepším statistickým modelem ten, který dokáže co nejlépe předpovědět budoucí užítkovost na základě známého genetického pozadí (Legarra & Reverter 2017). V této kapitole proto budou představeny nejběžnější statistické metody, které se používají za účelem porovnání modelů pro předpověď plemenných hodnot – ať se jedná o modely s odlišnou skladbou faktorů, nebo odlišně sestavenými maticemi příbuznosti (rodokmenová vs. genomická). Dále budou shrnuty používané modely pro selekci na výšku hřbetního tuku a svalu.

2.4.1. Metody pro srovnání animal modelů

Jelikož navržené smíšené modely bývají v první řadě fitovány pomocí (RE)ML, lze srovnat už výslednou logaritmickou věrohodnost (López-Romero & Carabaño 2003; Lukač 2016). Logaritmická věrohodnost (ve formě $-2LL$) měří nevysvětlenou proměnlivost. Z této hodnoty se pak počítají informační kritéria jako Akaikovo (AIC) nebo Bayesovo (Schwarzovo; BIC), která vyjadřují **vhodnost modelu** korigovanou o jeho komplexnost (počet parametrů):

$$AIC = -2LL + 2k; BIC = -2LL + k \times \log n, \quad [18]$$

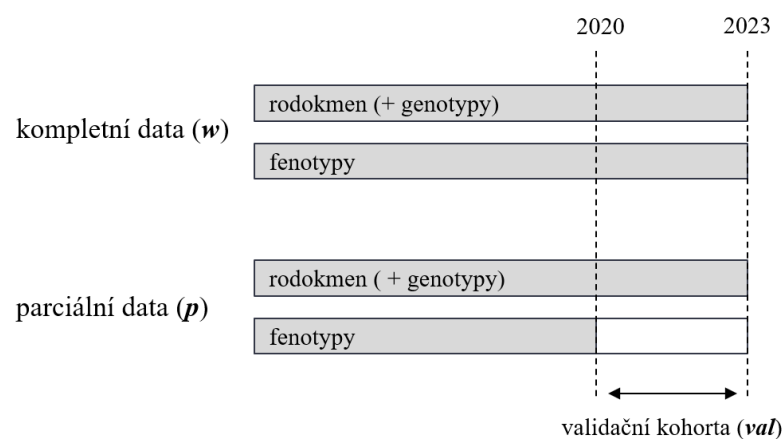
kde LL je logaritmická věrohodnost z ML nebo REML, k je počet parametrů modelu a n je počet pozorování (Field & Miles 2010). Pokud se modely liší v pevných efektech, nelze přímo srovnávat věrohodnosti z REML. V takovém případě se používá dvojí fitování: nejprve pomocí ML pro porovnání modelů a poté pomocí REML pro odhad (ko)variancí vybraného modelu (Verbyla 2019). Tento autor ovšem také navrhl alternativní metodu, díky níž je možné hodnoty z REML rozložit a tyto složky následně srovnat.

Avšak vhodnost modelu popisuje pouze známá čili minulá data. Kdežto šlechtitele obvykle zajímají nová neproověřená zvířata, tedy schopnost modelu předpovědět budoucí stav (Legarra & Reverter 2017). Za tímto účelem se využívají **křížové validace**. K jejich oblíbenosti přispěl příchod genomické selekce, jelikož dosavadní genetické hodnocení založené na rodokmenu se opíralo hlavně o testování potomků (Legarra & Reverter 2018). Křížové validace se zakládají na principu rozdělení datového souboru na trénovací neboli referenční data a testovací neboli validační data. Na trénovacích datech je model „naučen“ (fitován) a následně jsou jeho parametry využity pro předpověď plemenných hodnot (ať tradičních = EBV, nebo genomických = GEBV) v testovacích datech, z nichž mohou být zároveň odebrány některé údaje. Srovnáním skutečných hodnot testovacích dat (např. korigovaných fenotypů) s předpovězenými hodnotami je tak hodnocena **prediktivní výkonnost** (Daetwyler et al. 2013; Runcie & Cheng 2019).

Základní univerzální křížovou validaci představuje **k-fold**. V této metodě je soubor rozdělen do k přibližně stejně velkých podskupin, například pěti. Čtyři podskupiny utvoří referenční soubor (se všemi fenotypy i genotypy), na jehož základě jsou předpovídány (G)EBV zbylé, validační podskupiny (se skrytými fenotypy). Tento proces se opakuje pro každou podskupinu, případně dochází k nové definici podskupin pro zajištění větší náhodnosti výběru (Pérez-Cabal et al. 2012; Daetwyler et al. 2013). Zvláštním případem k -fold je metoda leave-one-out, kdy se k rovná počtu pozorování. Její předností oproti klasické k -fold je nezávislost na náhodném rozdělení do podskupin. Na druhou stranu je výpočetně smysluplná pouze na menších souborech (Cheng et al. 2021).

Nicméně v kontextu šlechtění se tyto metody ve své původní formě příliš nehodí, neboť rozdělení dat ignoruje populační strukturu podmíněnou rodinami nebo sledem generací. Z tohoto důvodu se jako praktičtější jeví **dopředná validace**, která definuje pouze dva datové soubory podle určitého datumu (obrázek 2.4-A). Prvním je kompletní datový soubor (w), jenž slouží jako referenční. Druhý datový soubor se nazývá parciální (p), protože od zvoleného datumu jsou v něm skryty fenotypy (ovšem počet zvířat zůstává stejný). Ověření se provádí na takzvané validační kohortě (val), tedy na vybraných zvířatech se skrytými fenotypy (Legarra & Reverter 2017; Runcie & Cheng 2019). Validační kohortou mohou být všechna zvířata s odebranými užitkovostmi, podskupina (genotypovaných) kandidátů pro selekci nebo podskupina s podobnými spolehlivostmi předpovědi (Song et al. 2019; Zhou et al. 2019; Melnikova et al. 2021; Salek Ardestani et al. 2021).

Obrázek 2.4-A: Příklad tvorby dat pro dopřednou validaci.



Dat rozdělených podle principu dopředné (nebo i křížové) validace využívá metoda prediktivní schopnosti neboli **prediktivita** (PrA), která pro validační kohortu koreluje fenotypy očištěné o pevné efekty (y_w^*) s (G)EBV předpovězenými na parciálních datech (EBV_p ; Lourenco 2018), tedy:

$$PrA_{val} = r(y_w^*, EBV_p). \quad [19]$$

Při vydělení prediktivity odmocninou z koeficientu dědivosti lze získat odhad **predikční přesnosti** (Serão et al. 2016), která je zároveň aproximací níže uvedené populační přesnosti (Legarra & Reverter 2018): $acc_{PrA} = \frac{PrA_{val}}{h} \sim acc_{LR}$. Obdobou prediktivity je střední kvadratická chyba predikce (Daetwyler et al. 2013).

Avšak Legarra & Reverter (2018) upozorňují na patrnou nevýhodu prediktivity: zvířata ve validační kohortě musejí mít buď zjištěné fenotypy, nebo jejich přesné odhady na základě potomstva. Zejména u mladých býků tohle není vždy proveditelné (např. u mléčné užitkovosti nebo u maternálních efektů). Chyby do odhadu může vnést také překorekce fenotypů. Ještě náchylnější na chyby je pak odhad predikční přesnosti, protože počítá s dalším parametrem (Lourenco 2018).

Legarra & Reverter (2018) tak navrhli rámcovou metodiku, kterou pojmenovali LR (jelikož vychází z lineární regrese). **Metoda LR** cílí zvláště na BLUP, neboť její metriky ověřují, nakolik jsou předpovědi „nejlepší“ a „nevychýlené“. Níže jsou uvedeny čtyři nejdůležitější metriky a v závěru kapitoly je na jednoduchých datech demonstrováno, jak se hodnoty těchto metrik mění v různých scénářích (obrázek 2.4-B až 2.4-E). Výpočty platí pro validační kohortu a EBV lze zaměnit za GEBV. Metriky jsou aproximační, takže zjednodušeně $EBV_w \sim TBV$. Avšak autoři při jejich odvození poukazují například na skutečnost, že $r(EBV_w, EBV_p) \neq r(TBV, EBV)$.

1) Vychýlení:

$$\mu_{p,w} = \overline{EBV}_p - \overline{EBV}_w; E(\mu_{p,w}) = 0, \quad [20]$$

kde $\overline{EBV}_{p/w}$ je průměr EBV parciálního/kompletního souboru a $E(\mu_{p,w})$ je očekávaná hodnota vychýlení. Tento parametr se v klasické lineární regresi podobá průsečíku a byl zohledňován už dříve (Legarra & Reverter 2017). Macedo et al. (2020a) doporučili jeho standardizaci genetickou směrodatnou odchylkou (σ_a), aby bylo možné porovnat vychýlení různých vlastností nebo datových souborů. Dále proto bude vychýlení chápáno v této standardizované formě:

$$\mu_{p,w} = \frac{\overline{EBV}_p - \overline{EBV}_w}{\sigma_a}. \quad [21]$$

2) Disperze (rozptýlení):

$$1 - b_{w,p} = 1 - \frac{cov(EBV_w, EBV_p)}{var(EBV_p)}; E(1 - b_{w,p}) = 0, \quad [22]$$

kde $b_{w,p}$ je regresní koeficient, $cov(EBV_w, EBV_p)$ je kovariance mezi EBV kompletního a parciálního souboru, $var(EBV_p)$ rozptyl EBV parciálního souboru a $E(1 - b_{w,p})$ je očekávaná hodnota disperze.

Kladná hodnota disperze, která vyplývá z $b_{w,p} < 1$ a je často pozorována u genomických předpovědí, tedy může značit přílišný rozptyl EBV_p , nebo naopak nedostatečný rozptyl EBV_w .

3) Inverzní korelace:

$$\rho^{-1} = \frac{sd(EBV_w) \times sd(EBV_p)}{cov(EBV_w, EBV_p)}; E(\rho^{-1}) = \frac{acc_w}{acc_p}, \quad [23]$$

kde $sd(EBV_{w/p})$ je směrodatná odchylka EBV parciálního/kompletního souboru a $E(\rho^{-1})$ je očekávaná hodnota inverzní korelace a $acc_{w/p}$ je populační přesnost EBV kompletního/parciálního souboru. Tento parametr tak ukazuje nárůst v přesnosti po přidání více fenotypů (Pravia et al. 2023). Je tedy vhodné vyzdvihnout, že korelace $\rho(EBV_w, EBV_p)$ vyjadřuje poměr přesností, nikoliv přímý výpočet přesnosti jako v teoretickém případě $r(TBV, EBV)$.

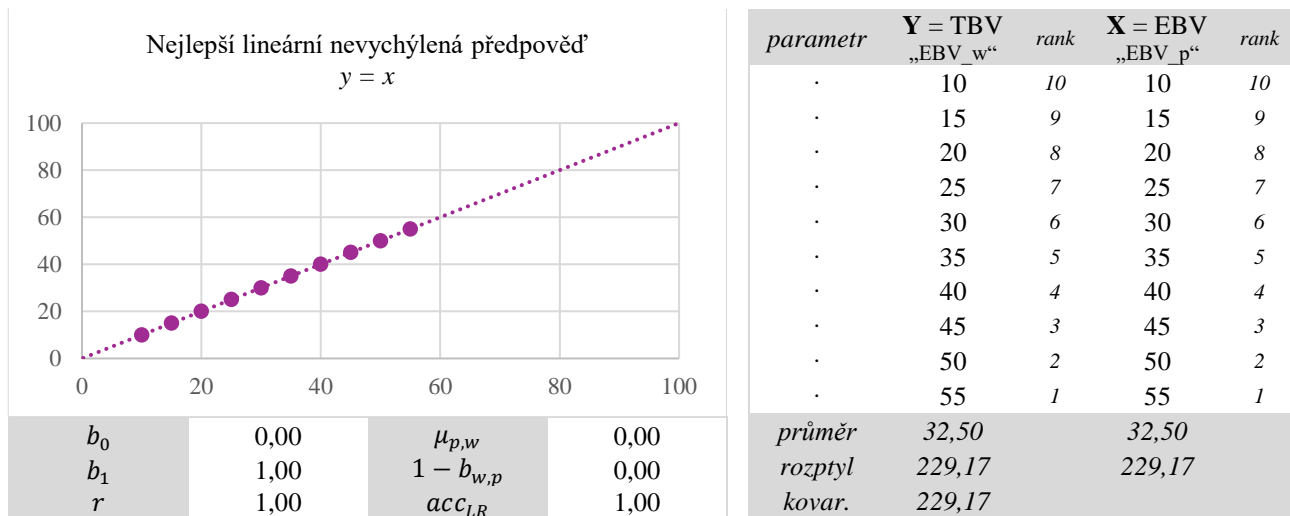
4) Populační přesnost:

$$acc_{LR} = \sqrt{\frac{cov(EBV_w, EBV_p)}{(1 - \bar{F}) \times \sigma_a^2}}; E(acc_{LR}) = acc_p \quad [24]$$

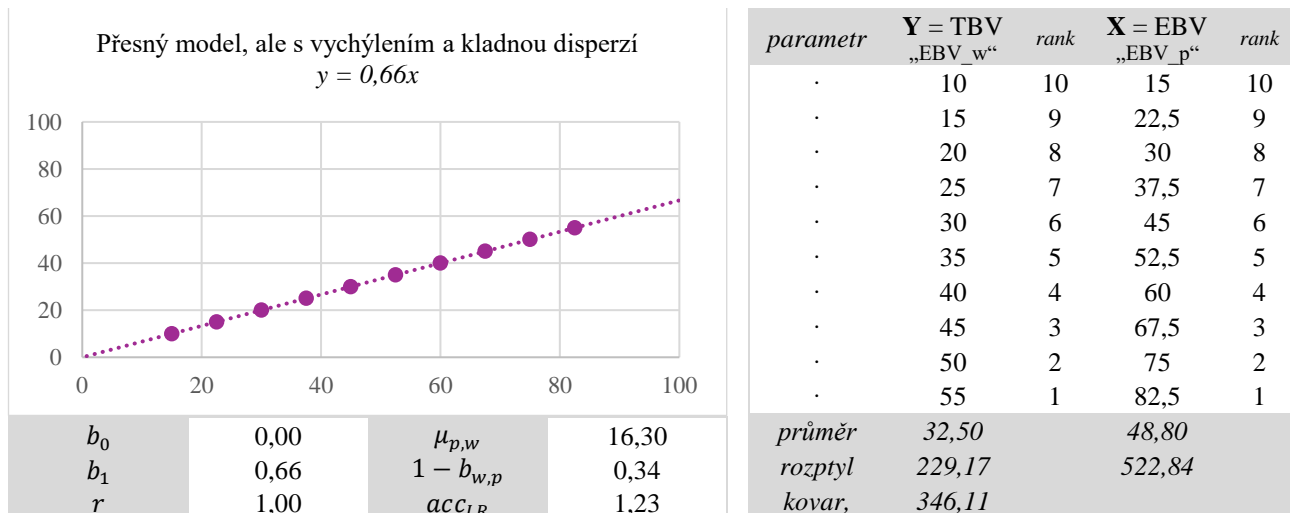
kde \bar{F} je průměrný koeficient inbreedingu validační kohorty, σ_a^2 aditivní genetický rozptyl odhadnutý v daném modelu a $E(acc_{LR})$ je očekávaná hodnota populační přesnosti. Tento vzorec představuje zjednodušení původního vzorce (Jang et al. 2022). V původní verzi platí, že $acc_p = r(TBV, EBV)$. Populační přesnost podle odvození Legarry & Revertera (2018) tedy vychází z teoretické definice přesnosti jakožto korelace, která hodnotí jedince dohromady. Populační přesnost vyjadřuje, jak dobře dokáže daný model stanovit genetické založení všech jedinců, tedy jak vysoký je potenciální genetický zisk (při konstantní hodnotě dalších parametrů). Je tedy nutné ji oddělit od individuální přesnosti, jež stanovuje riziko při výběru daného zvířete a odhaduje se pomocí příslušného prvku inverze levé strany MME (jak bylo popsáno v kapitole 2.3.3.). Populační přesnost se nerovná průměru individuálních přesností, neboť zohledňuje také závislosti mezi individuálními přesnostmi. Například pokud je hodnocena skupina vlastních sourozenců na základě rodičovského průměru a oba rodiče mají EBV s přesností 1,00, individuální přesnost každého sourozence činí 0,71. Avšak populační přesnost je 0,00, protože všichni sourozenci mají úplně stejný rodičovský průměr.

Přesto se průměry individuálních přesností používají pro porovnání modelů také, jak je uvedeno například v kapitole 2.5.4. Mimo to stojí za zmínku deterministické metody, které jsou založené na takzvaném efektivním počtu nezávislých chromozomálních segmentů a představují alternativu k výše popsaným empirickým validacím (Erbe et al. 2013; Dekkers et al. 2021). Jejich popis by ovšem přesahoval rámec této práce.

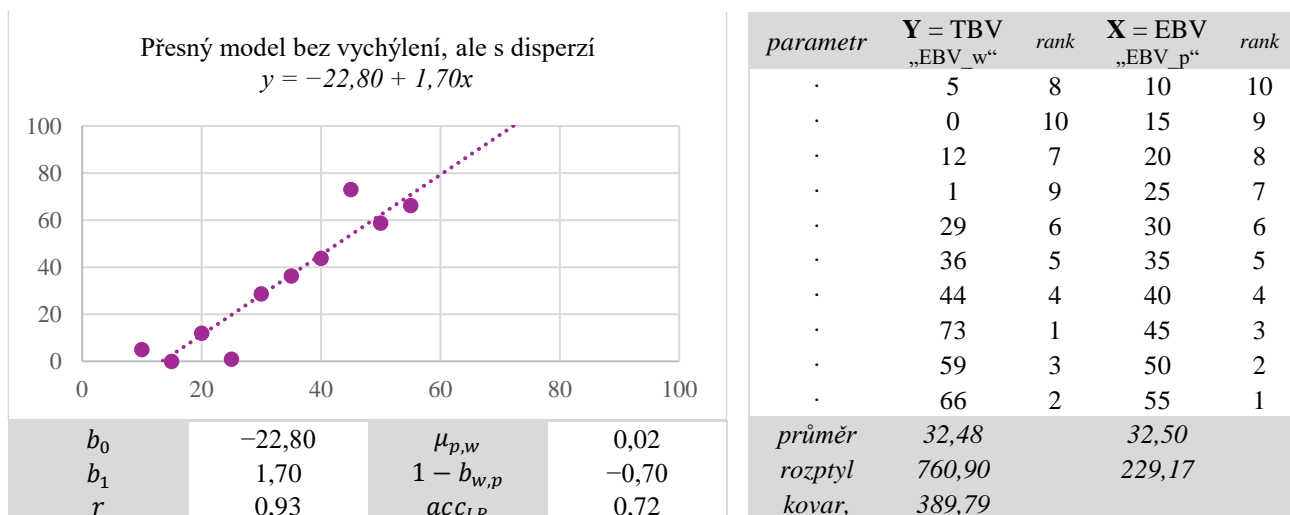
Obrázek 2.4-C: Znárodnění správného BLUP modelu v rámci jednoduché regrese a metody LR.



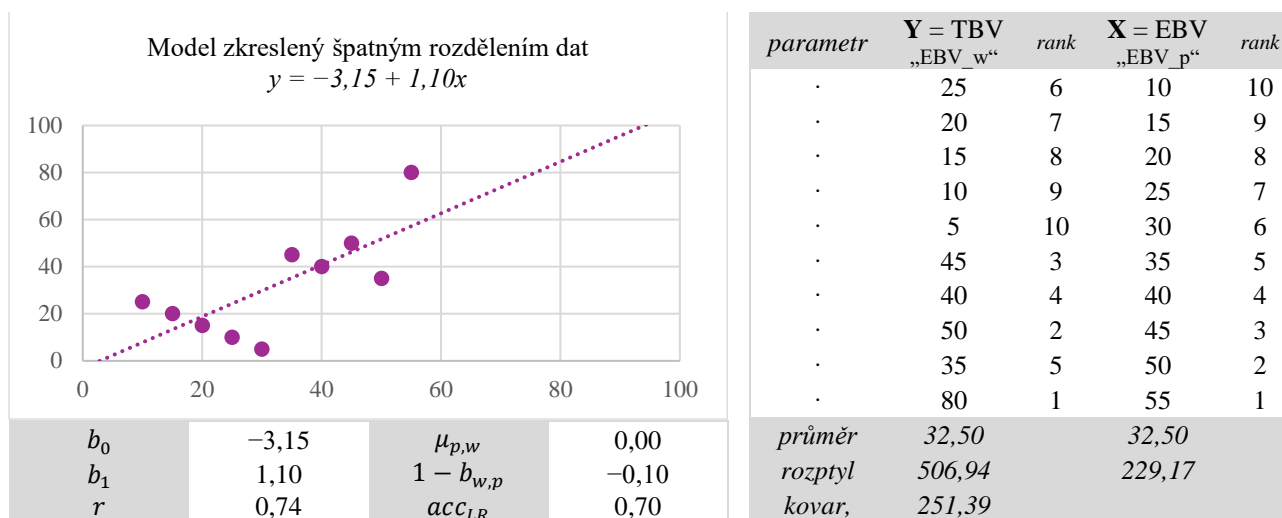
Obrázek 2.4-B: Častý případ – nadhodnocení GEBV (EBV_p) u mladých zvířat.



Obrázek 2.4-D: Model s vysokou zápornou disperzí, přestože je velmi přesný a bez vychýlení.



Obrázek 2.4-E: Nedostatek LR metody při zanedbání systematických trendů – metriky nabývají přijatelných hodnot, i když se výrazně mění pořadí zvířat.



Poznámky: EBV_p lze zaměnit za GEBV; pro výpočet acc_{LR} je za σ_a^2 dosazen rozptyl TBV (acc_{LR} může z toho důvodu přesahovat hodnotu 1,00)

Zkratky: $1 - b_{w,p}$ = disperse podle metody LR; acc_{LR} = populační přesnost podle metody LR; b_0 = absolutní člen; b_1 = regresní koeficient; EBV = předpovězená plemenná hodnota (_w = z kompletního souboru; _p = z parciálního souboru); $\mu_{p,w}$ = vychýlení podle metody LR; r = korelační koeficient; rank = pořadí jedinců podle EBV v daném souboru; TBV = skutečná plemenná hodnota; X = nezávislá proměnná; Y = závislá proměnná

2.4.2. Modely pro výšku tuku a svalů

Hlavním výstupem této dizertační práce je návrh víceznakového modelu za účelem genetického hodnocení výšky hřbetního tuku (BF) a výšky svalu *longissimus thoracis et lumborum* (MD). V českém šlechtitelském programu, podobně jako třeba v Německu (Schulz et al. 2019), jsou zatím tyto vlastnosti hodnoceny nepřímo prostřednictvím regresní rovnice pro odhad LMP. Současná rovnice pro LMP (k přístroji Mindray) však zahrnuje dva poměrně vysoké záporné regresní koeficienty pro BF, a naopak jeden nízký kladný koeficient pro MD:

$$LMP = 69,5000 - 0,6380 \times BF_A - 1,3130 \times BF_B + 0,0576 \times MD. \quad [25]$$

Po dosazení vstupních hodnot pro obě vlastnosti (např. BF = 6 mm a MD = 48 mm) tak vyplývá, že BF ovlivňuje výsledný LMP zhruba čtyřikrát více než MD. Jinými slovy, selekce na LMP znamená převážně selekci proti BF.

Zejména u mateřských plemen se další pokles BF stal nežádoucím vzhledem ke korelaci s reprodukční užitkovostí (Houde et al. 2010; Roongsitthichai & Tummaruk 2014), a tak váhový koeficient pro LMP v CPH klesl až na nynější 2 % (kapitola 2.2.). Nicméně zároveň se tím omezil selekční tlak na MD, který by byl malý sám o sobě kvůli nízkému regresnímu koeficientu. Součinnost těchto dvou faktorů tak možná zapříčinila pokles MD, jenž je pozorován v posledních letech

u jatečných prasat (a také u čistokrevných, jak znázorňuje obrázek 2.4-E). Jako řešení se proto nabízí přímé hodnocení BF a MD víceznakovým modelem po vzoru Francie (G. Dumas a C. García-Baccino, osobní komunikace), které v takovém případě může být praktičtější než hodnocení složené vlastnosti jako LMP (Snowder & Fogarty 2009).

V tabulce 2.4-B jsou shrnuty vybrané studie, které se zabývaly sestavením víceznakového modelu pro BF, MD a případně také průměrný denní přírůstek (ADG), živou hmotnost (BW) nebo jatečnou hmotnost (CW). Kromě samozřejmého aditivního genetického efektu zvířete jsou obvykle zahrnuty prediktory pohlaví, skupiny vrstevníků se stejnými podmínkami prostředí (CG) a vrhu. Častá je také regrese na hmotnost a věk, ačkoliv se zároveň hmotnost či přírůstek vyskytují mezi korelovanými vlastnostmi (např. Hermesch et al. 2000; Dube et al. 2013; Lukač 2016). Avšak víceznakový model, v němž se vlastnost A používá jako vysvětlovaná proměnná i jako prediktor vlastnosti B, ze statistického hlediska není správný (I. Misztal, D. Lourenco a S. Gezan, osobní komunikace). Někteří autoři dále zohledňují maternální efekty (Dube et al. 2013; Dong et al. 2020) nebo pořadí vrhu (parita; Hermesch et al. 2000; Salces et al. 2006; Lopez et al. 2018).

Zvláštní pozornost zasluhuje definice faktoru CG. Jeho typickým příkladem je kombinace stádo-rok-období (SRO), ovšem nabývá i dalších forem jako třeba pohlaví-stádo-rok a stádo-rok-týden (Lopez et al. 2019; Khanal et al. 2019), případně může být zastoupen samostatnými faktory vyšší úrovně (Csató et al. 2002; Salces et al. 2006; Lukač 2016). Specifický případ CG představuje vrh, který ovšem bývá zohledňován jako samostatný faktor. Toto rozlišení má svůj význam proto, že vrh je poměrně malou a geneticky homogenní skupinou do období předvýkrmu, zatímco CG zpravidla slučuje větší množství vrhů v závislosti na organizaci chovů. Z toho pak vyplývá odlišné statistické pojetí, kdy CG bývá definován spíše jako pevný efekt a vrh bez výjimky jako náhodný.

Volba mezi pevným a náhodným efektem CG představuje balancování mezi přesností a vychýlením předpovědí (Van Vleck 1987; Silva et al. 2019). Tato volba záleží především na populační struktuře, takže ji lze hodnotit například ve vztahu k efektivnímu počtu potomků na plemeníka v rámci CG (van Bebber et al. 1997) nebo k propojenosti CG (Škorput et al. 2018). Pevná CG přináší výhodu v případě existence velkých systematických rozdílů, které mohou být způsobeny podmínkami prostředí nebo nedostatkem propojenosti populace. Častým problémem bývá nenáhodné rozdělení plemeníků napříč stády, zejména využívání nejlepších plemeníků v nejlepších chovech (Visscher & Goddard 1993; Frey et al. 1997; Babot et al. 2003; Škorput et al. 2018). Právě z důvodu existence systematických genetických rozdílů mezi chovy bylo pro česká mateřská plemena doporučeno modelovat CG jako pevný efekt (Wolf J & Wolfová M 2012).

Nicméně Schaeffer (2009) argumentuje, že faktor CG (obzvláště ve formě SRO) ze své podstaty zahrnuje vysoký počet neopakovatelných kategorií, které odpovídají spíše náhodnému výběru z normálního rozdělení. Považovat SRO čili trojitou interakci za pevný efekt by teoreticky mělo znamenat současné zohlednění všech efektů vyššího řádu, což ale nebývá zvykem (patrně kvůli přílišné složitosti modelu). Schaeffer (2009) proto doporučuje nastavit CG jako náhodný efekt a genetické rozdíly mezi stády nebo celkový genetický pokrok populace v čase podchytit pevným efektem stáda, respektive roku. Ugarte et al. (1992) za podobným účelem navrhuje použít náhodný efekt rok-období vnořený do pevného efektu stáda. V souladu s těmito doporučeními Nagy et al. (2004) pozorovali mírné zpřesnění modelu po zavedení náhodné CG u maďarského bílého ušlechtilého. Silva et al. (2019) předpověděli, že náhodná CG zvýší roční genetický zisk u brazilského holštýnského skotu. Vyšší prediktivní výkonnost modelu s náhodnou CG popsali také Babot et al. (2003), kteří ale zároveň upozornili, že předpokládaná genetická odezva může být nadhodnocena kvůli výše popsaným systematickým trendům.

Co se týče populačních genetických parametrů pro zkoumané vlastnosti, příložená tabulka 2.4-A ukazuje, že v literatuře průměrná hodnota koeficientu dědivosti činí 0,44 pro BF a 0,32 pro MD. Průměrná genetická korelace mezi nimi je $-0,23$, avšak jednotlivé odhady se mohou výrazně lišit. Rozdíly způsobuje především genetické pozadí populací (země, rok) a konstrukce modelu. Například BF očištěná o efekt BW je jinou statistickou veličinou nežli původní BF, což se projeví hlavně na genetických korelacích. K tomu někteří autoři ve vzorci pro výpočet koeficientu dědivosti chápou fenotypovou proměnlivost včetně systematických efektů (kapitola 2.3.2.).

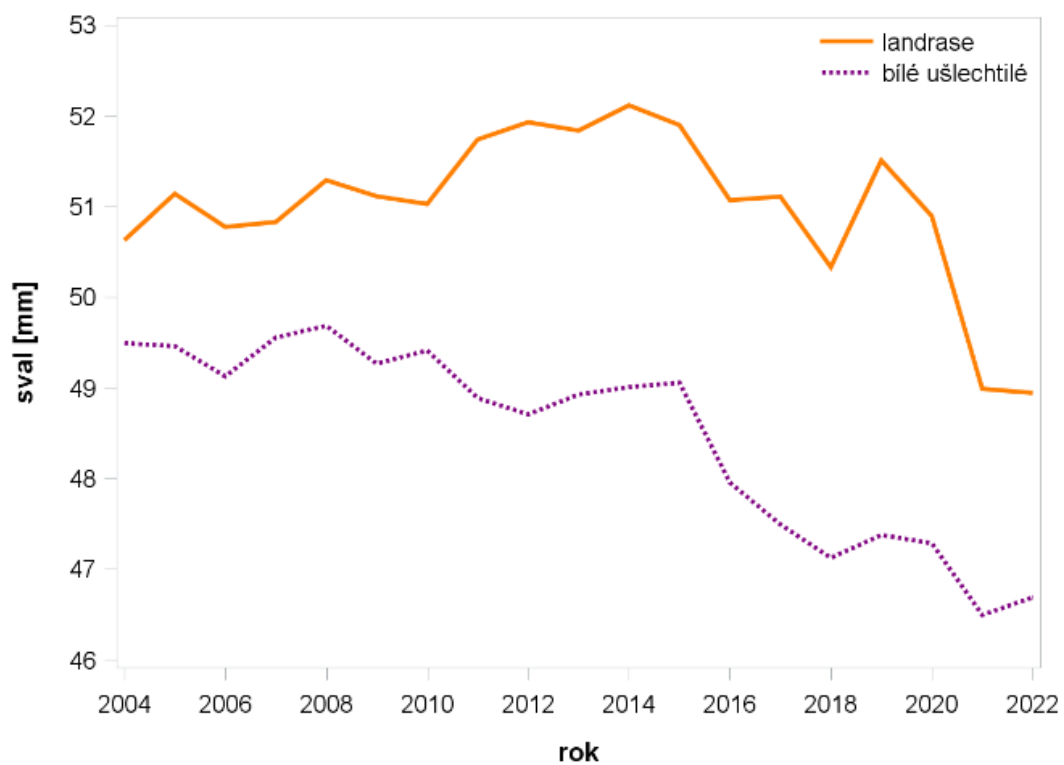
Tabulka 2.4-A: Průměrné, minimální a maximální hodnoty populačních genetických parametrů pro přírůstek (či hmotnost), výšku hřbetního tuku a výšku svalu ve sledovaných studiích.

Parametr	Koeficient dědivosti			Koeficient genetické korelace		
	ADG (BW)	BF	MD	ADG–BF (BW–BF)	ADG–MD (BW–MD)	BF–MD
průměr	0,34 (0,34)	0,44	0,32	0,03 (0,24)	$-0,05$ (0,23)	$-0,23$
minimum	0,10 (0,28)	0,27	0,20	$-0,26$ ($-0,31$)	$-0,40$ ($-0,31$)	$-0,72$
maximum	0,80 (0,44)	0,61	0,55	0,37 (0,63)	0,36 (0,77)	0,39

Zkratky: ADG = průměrný denní přírůstek; BW = hmotnost; BF = výška špeku; MD = výška svalu

Použité zdroje: Johnson et al. (1999); Bizelis et al. (2000); Hermeschet al. (2000); Chen et al. (2002); Csató et al. (2002); Salces et al. (2006); Dube et al. (2013); Ros-Freixedes et al. (2013); Miar et al. (2014); Lukač (2016); Lopez et al. (2019); Davoli et al. (2019); Szyndler-Nędza et al. (2019); Khanal et al. (2019); Dong et al. (2020)

Obrázek 2.4-E: Roční vývoj BF a MD u mateřských plemen (vlastnosti očištěny o efekty ultrazvukového přístroje, chovu, pohlaví, hmotnosti a věku).



Tabulka 2.4-B: Navržené modely pro genetické hodnocení výšky hřbetního tuku či svalu.

Zdroj	Plem.	N	Vlastnost	Efekty		h ²	Genetické korelace	
				pevné	náhodné			
Bizelis et al. (2000)	LW	1 332	BF	pohl., rok, období	zvíře	0,40	BF–MD	–0,36
	LA	1 244	MD			0,55		
Chen et al. (2002) ¹	LW	361 300	BF	pohl., CG	vrh, zvíře	0,48	BF–MD	–0,25
	LA	71 097	MD			0,40		
Csató et al. (2002)	LW	67 422	BF	pohl., stádo, rok×měsíc, stanice, BW (R)	zvíře	0,49	BF–MA	–0,35
	LA	33 079	MA			0,34		
Davoli et al. (2019)	LW	≥ 838	BF	pohl., datum porážky, věk (R)	vrh, zvíře	0,48	BF–MA	–0,35
			MA			0,30		
Dong et al. (2020) ²	LW	≥ 45 053	BF	pohl., rok–měsíc	vrh, matka, zvíře	0,36	BF–MA	0,39
			MA			0,31		
Dube et al. (2013)	LW	≥ 4 128	ADG	pohl., CG, příjem krmiva (R), počáteční BW (R)	matka, zvíře	0,38	BF–MA	0,21
			BF			0,44		
Hermesch et al. (2000) ³	LW	≥ 1383	ADG	plem., týden, parita, BW (R), věk (R)	vrh, zvíře	0,36	CW–BF	–0,31
	LA		BF			0,25		
Johnson et al. (1999) ⁴	LW	26 706	MD	stádo–měsíc, věk (R)	vrh, zvíře	0,21	BF–MD	–0,72
			ADG			0,21		
Khanal et al. (2019) ⁵	hybrid	≥ 40 951	ADG	pohl., CG, parita	turnus, zvíře	0,27	ADG–BF	–0,26
			BF			0,61		
Lopez et al. (2018)	LW	≥ 174 013	MA	pohl., parita, CG	zvíře	0,21	BF–MA	–0,41
	LA	≥ 59 093	ADG			0,27		
Lukač (2016) ⁶	LW	14 746	ADG	stádo, rok, období, BW (R), věk (R)	zvíře	0,29	ADG–BF	0,37
	LA	20 979	BF			0,42		
Miar et al. (2014)	hybrid	2 048	MD	pohl., farma, turnus, věk (R)	vrh, zvíře	0,30	BF–MA	–0,27
			ADG			0,30		
Ros-Freixedes et al. (2013)	DU	≥ 80 687	ADG	pohl., turnus, věk (R)	vrh, zvíře	0,42	ADG–BF	–0,02
			BF			0,45		
Salces et al. (2006) ⁷	DU	6 804	MD	pohl., farma, rok, období, parita	zvíře	0,21	BF–MA	–0,29
			ADG			0,36		
Szyndler-Nędza et al. (2019) ⁸	PUL	6 042	ADG	stádo, rok	otec, matka, zvíře	0,25	ADG–MA	–0,40
			BF			0,21		

Zkratky: ADG = průměrný denní přírůstek; BF = výška hřbetního tuku; BW = živá hmotnost; CG = skupina zvířat se stejnými podmínkami; CW = jatečná hmotnost; DU = duroc; pohl. = pohlaví; MA = plocha svalu; MD = výška svalu; LA = landrase; LW = bílé ušlechtilé; PUL = Pulawska; R = pevná regrese; NA = nedostupný údaj

Poznámky (horní index): 1 = konstantní hm.; 2 = konstantní věk; 3 = pouze kanci, zprůměrované odhady; 4 = pouze kanci; 5 = výsledky pouze z jedné hybridní linie; 6 = pouze prasničky, zprůměrované odhady; 7 = zprůměrované odhady; 8 = pouze prasničky, zprůměrované odhady

2.5. Využití SNP ve šlechtění

V moderním šlechtění hrají stále větší úlohu molekulární markery (Khlestkina 2014), neboť kladně působí na největší limitující faktory šlechtitelského programu: rychlost zařazování vybraných zvířat do plemnitby a přesnost předpovědí plemenných hodnot. Od využití markerů se tak obecně očekává zvýšení selekčního efektu, což vyplývá ze vzorce pro roční genetický zisk (ΔG):

$$\Delta G = \frac{ir\sigma_A}{L}, \quad [26]$$

kde i značí intenzitu selekce, r přesnost EBV, σ_A aditivní genetickou směrodatnou odchylku a L generační interval v letech (Van Eenennaam et al. 2014)

Markery pomáhají odhalovat původ genetické proměnlivosti, která je způsobená mutacemi. Mutace mohou působit na úrovni chromozomálního segmentu, repetitivních sekvencí nebo jednotlivých nukleotidů. Ve třetím případě pak hovoříme o jednonukleotidových polymorfizmech (SNP), jež se nejčastěji vyskytují ve dvou alelách neboli variantách (Teama 2018). Jednobodová mutace se obvykle nazývá SNP, pokud frekvence minoritní alely překročí 1 % (Wright 2005).

Z molekulárních markerů mají pro šlechtění největší význam právě SNP, protože jsou snadno detekovatelné, vyskytují se napříč celým genomem ve velkém množství a jsou geneticky stabilní (Yang et al. 2013; Khlestkina 2014). V genomu prasete (*Sus scrofa f. domestica* L.) o velikosti 2,7 Gb (Raudsepp & Chowdhary 2011) bylo detekováno 64,4 milionu SNP (Zhou et al. 2017), což odpovídá přibližně 2,4 % genomu.

2.5.1. Genotypovací metody

Biotechnologických metod, které zjišťují přítomnost konkrétního SNP, existuje celá řada. Metody používané ve šlechtění zvířat rozdělili (Koopae & Koshkoiyeh 2014) do dvou hlavních kategorií: tradiční postupy s využitím gelové elektroforézy a vysoce výkonné postupy. Obě kategorie jsou metodicky zastoupeny i v praktické části této dizertace.

Do první kategorie spadá kombinace polymerázové řetězové reakce a délkového polymorfizmu restrikčních fragmentů (**PCR-RFLP**). Tato genotypovací technika je velmi jednoduchá, ale vyžaduje znalost cílového regionu včetně SNP a zároveň přítomnost sekvence pro restrikční enzym, která vzniká, nebo naopak zaniká díky mutaci (Koopae & Koshkoiyeh 2014). Samotná PCR slouží k namnožení (amplifikaci) požadovaného úseku izolované DNA a svým principem se podobá *in vivo* replikaci. Amplifikace probíhá po 20 až 40 cyklů v termocykleru, jehož účelem je rychlá změna teploty. Cyklus obsahuje tři hlavní kroky: 1) rozpletení dvoušroubovice (denuraci; 92–96 °C); 2) nasednutí primerů (hybridizaci, annealing; 45–65 °C), které ohraničují cílový úsek; a 3) prodloužení řetězců (elongaci, syntézu; 72 °C) pomocí termorezistentní DNA

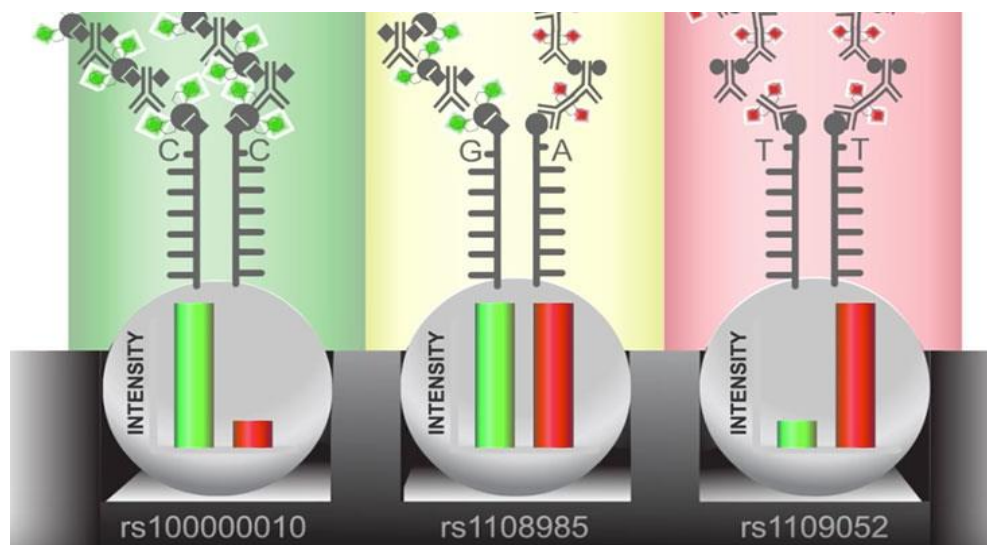
polymerázy. Jelikož se počet segmentů s každým cyklem zdvojnásobí, po n cyklech takto teoreticky vznikne 2^n kopií. Samotné genotypování přichází po přidání restriční endonukleázy, která štěpí palindromickou sekvenci. Následuje separace restričních fragmentů na základě velikosti pomocí gelové elektroforézy (typicky s agarózovým nebo akrylamidovým gelem) a vizualizace výsledků s využitím interkalačních fluorescenčních barviv (Kočárek 2008; Hashim & Al-Shuhaib 2019).

Ve šlechtitelské praxi však už řadu let převládají vysoce výkonné postupy, protože se ve statistických modelech uplatňují desítky až stovky tisíc SNP napříč celým genomem. K jejich zjištění slouží zejména technologie DNA čipů (microarrays). Tento čip je skleněná nebo silikonová destička, na níž jsou přichyceny známé úseky DNA, typicky oligonukleotidy, které mají úlohu sond. Sondy jsou komplementární k požadovaným úsekům vzorku DNA, nazývaným jako cíle (target). Po navázání sondy na cílovou sekvenci dochází k emisi fluorescenčního signálu, jenž je detekován specializovanou kamerou a počítačovým systémem jako konkrétní alela. Z technických řešení DNA čipů se prosadily především alelická diskriminace pomocí hybridizace a alelicky specifické prodlužování primeru. Zatímco v prvním případě je označena cílová DNA, ve druhém případě je značka (tag) zabudována až během prodloužení (Missoum 2018).

Genotypová data, která jsou využita v této práci, vznikla během tvorby referenční populace pro národní šlechtitelský program pomocí technologie Illumina Infinium® BeadArray. Pro získání kvalitních vzorků bylo doporučeno odebrat štetinové cibulky a izolovat DNA magnetickou separací (Krupa et al. 2019b). Proces Illumina Infinium následně vyžaduje namnožení (20–24 h při 37 °C) a enzymatickou fragmentaci DNA. Poté může být až 24 vzorků najednou nanášeno na čip do oddělených polí. Samotná hybridizace cílové DNA k oligonukletidovým sondám trvá 16–24 h při 48 °C. Jako nosič sond slouží kuličky (beads) o průměru 3 μm, z nichž je odvozen název BeadArray. Každá kulička má známou pozici na destičce a na svém povrchu nese okolo milionu kopií sondy pro jeden lokus. Sondy plní úlohu primerů a jsou navrženy tak, aby jejich sekvence končila jednu bázi před cílovým lokusem. Rozlišení alel je pak zjištěno prodloužením vlákna sondy o jednu bázi se specifickým hapténem, ke kterému se váže fluorescenčně značený protein. V závěrečné části postupu jsou genotypy detekovány ve skeneru po excitaci značek laserem (Missoum 2018; Illumina 2019). Tento krok je zachycen na obrázku 2.5-A.

DNA čipy s vysokou hustotou využívané ve šlechtění prasat vycházejí z práce Ramose et al. (2009), kteří navrhli čip s pokrytím přes 60 tisíc SNP. Novější komerčně dostupné čipy do velké míry kopírují původní návrh, jehož průměrná vazebná nerovnováha 0,36–0,46 zdaleka převyšuje minimum 0,20, které je potřebné pro dosažení přesnosti předpovědi GEBV 0,85 (Samorè & Fontanesi 2016).

Obrázek 2.5-A: Signály pro trojici SNP na čipu Illumina. Každý cílový lokus může být značen dvěma barvami (zelenou a červenou), jejichž relativní intenzita určí genotyp (Illumina 2024).



2.5.2. Markery asistovaná selekce

Základy pro využití molekulárních markerů ve šlechtění se datují ke konci 80. let a do 90. let 20. století, kdy se pomocí přístupu kandidátních genů nebo detailního mapování podařilo nalézt jak přímé markery (příčinné mutace), tak zejména markery ve vazbě s lokusy kvantitativních vlastností (QTL). Tyto regiony mají vysvětlovat pozorovatelnou část proměnlivosti, a tudíž jsou ekonomicky důležité (Dekkers 2004; Ibáñez-Escriche et al. 2014). Využití přímých i QTL markerů pro účely šlechtění se zpočátku jevílo velmi slibně a získalo název markery asistovaná selekce (MAS; Dekkers 2004). MAS přinesla výhody při šlechtění na vlastnosti, jejichž fenotyp je obtížně měřitelný nebo měřitelný pouze u jednoho pohlaví, případně na vlastnosti s nízkou dědivostí. Genotypování je možné provádět už u mláďat, čímž se zkracuje generační interval. Zároveň se MAS mohla uplatit v programech využívajících křížení (Wakchaure & Ganguly 2015).

Komerční šlechtění prasat nejprve těžilo ze SNP g.1843C>T, který objevili Fujii et al. (1991) v genu *RYR1* (*HAL*). Recessivní alela se totiž ukázala být příčinou pro dlouhodobě pozorovaný prasečí stresový syndrom, který souvisí se zhoršenou kvalitou masa. Do té doby se šlechtitelé snažili tuto poruchu odstranit pomocí pracovně náročného testu s halotanovým plynem a nákladné bylo rovněž hledání skrytých přenašečů. Z toho důvodu genotypování na polymorfismus g.1843C>T nabídlo jednoduchou, levnou a stoprocentně efektivní alternativu (Knol et al. 2016). Dalším úspěchem byla identifikace alely *Rendement Napole* v genu *PRKAG3*, jež také ovlivňuje kvalitu masa, a to zejména u plemene hampshire (Milan et al. 2000). Mutace c.307G>A v genu *FUT1* souvisí s rezistencí vůči bakterii *E. coli*, která způsobuje průjem po odstavu a edémovou chorobu (Bao et al. 2011). Významné QTL markery pak byly objeveny poblíž genů *IGF2* a *ESR1*. Paternálně exprimovaná alela ve 3. intronu genu *IGF2* vysvětlovala až 30 % fenotypové proměnlivosti v osvalení a až 20 %

proměnlivosti výšky hřbetního tuku, takže v populacích šlechtěných na zmasilost byla nepřímou selekcí prakticky zafixována (Van Laere et al. 2003). Rothschild et al. (1996) ve své asociační analýze pro *ESRI* zjistili rozdíl 1,5 selete ve vrhu mezi homozygotními genotypy.

Do praktické části této práce je zahrnut polymorfismus Asn298Asp (c.1426G>A; rs81219178) v genu *MC4R*, který objevili Kim et al. (2000b). Gen *MC4R* kóduje melanokortin 4 receptor na buněčné membráně a missense mutace v kodonu 298 (alela A) způsobuje poškození tohoto receptoru, neboť nedochází k aktivaci adenylátcyklatázy. Tím je narušena signální dráha regulující energetickou homeostázi, takže alela A vede k vyšší žravosti, vyššímu tučnění a celkově k rychlejšímu růstu zvířat. Vysokou frekvenci alely A je proto možné najít v populacích plemen jako duroc nebo hampshire (Kim et al. 2004). Přítomnost významné asociace mezi tímto polymorfizmem a produkčními ukazateli podpořila řada autorů (např. Houston et al. 2004; Van Den Maagdenberg et al. 2007; Piórkowska et al. 2010; Van Den Broeke et al. 2015). Mutace sice není označena za příčinnou, protože konečný efekt není biologicky tak výrazný jako na příkladu *RYRI* a zároveň se zdá být také populačně specifický (Stachowiak et al. 2006; Fan et al. 2009; Wang et al. 2019). Avšak o její možné významnosti svědčí i skutečnost, že v jejím regionu se vyskytují QTL signály (PigQTLdb 2024), jejichž pojmenování lze nalézt i v mapových souborech komerčních DNA čipů (např. ALGA0007081).

Snaha o nalezení významných markerů stále pokračuje, čemuž nahrává rozmach celogenomového sekvenování a celogenomových asociačních studií (Tang et al. 2019; Li et al. 2020). Rostoucí veřejná databáze pro prasata nyní zahrnuje téměř 49 tisíc QTL pro více než 670 vlastností (PigQTLdb 2023). Nicméně původní předpovědi ohledně potenciálu MAS se ukázaly být příliš optimistické. Relevantních markerů bylo ve vztahu k nákladům aplikováno poměrně málo, neboť se brzy potvrdil infinitesimální model kvantitativních vlastností. Většina QTL má nevýrazný efekt, anebo je jejich efekt nekonzistentní. Příčina spočívá v odlišných genetických vazbách mezi populacemi a generacemi či v nedostatečně velkých vzorcích některých asociačních studií. MAS se proto stala spíše doplňkovým nástrojem ke standardnímu hodnocení smíšenými modely, hlavně prostřednictvím přímých markerů (Dekkers 2004; Van Eenennaam et al. 2014; Ibáñez-Escriche et al. 2014; Knol et al. 2016).

Spojenou metodiku smíšených modelů a MAS, tedy markery asistovanou BLUP, představili Fernando & Grossman (1989). V ní je každý marker považován za pevný efekt se třemi úrovněmi podle genotypu (0, 1, 2). Tímto přístupem teoreticky dojde k lepší korekci efektů QLT nežli v obvyklé asociační analýze a zároveň se očekává zpřesnění předpovědi EBV. Ovšem vzrůstající počet markerů zahrnutých do modelu (až 30) vyústil v jejich přechod do kategorie „polonáhodných“ efektů, které využívají regularizační metody založené na dědivosti markerů (Knol et al. 2016).

2.5.3. První metody genomické selekce

Počátek 21. století se stal také počátkem éry genomické selekce (Misztal et al. 2020). Tento pojem se v literatuře objevoval už dříve, a to i v jiných kontextech (Genest & Daniel 1966; Hillel et al. 1990; Haley & Visscher 1998). Avšak obecně se rozšířil po použití v záhlaví článku od Meuwissena et al. (2001), kteří předvíдали využití hustých celogenomových markerových map pro zvýšení genetického zisku. Na simulovaných datech s 50 tisíci haplotypovými markery zároveň prezentovali různé metodické postupy pro předpověď efektů těchto markerů a EBV, které zásadně ovlivnily další vývoj.

Genomická selekce je nyní chápána jako využití velkého počtu SNP napříč genomem pro předpověď genetického založení jedince, tedy nejčastěji pro předpověď genomických plemenných hodnot (GEBV; Van Eenennaam et al. 2014). Tento pojem tak zahrnuje celou řadu různých metod, jejichž pojmenování, výklad a klasifikace se mohou lišit. Následující přehled je proto primárně založen na review Misztala & Legarry (2017), kteří v souladu s chronologickým vývojem metodiky rozlišují: 1) SNP-BLUP a související bayesovské postupy; 2) vícekrokovou genomickou BLUP; a 3) jednokrokovou genomickou BLUP.

SNP-BLUP vychází z původního návrhu Meuwissena et al. (2001). Na příkladu využití 50 tisíc SNP by klasický BLUP animal model zahrnoval předpovědi pro dalších 50 tisíc náhodných efektů, takže výsledná GEBV bude sumou těchto předpovědí vynásobených 0, 1, nebo 2 (podle genotypu). Předpokladem SNP-BLUP je, že efekty SNP opisují normální rozdělení a mají stejné rozptyly. Z tohoto pohledu představují nelineární metody jako BayesA/B/C/R nebo LASSO rozšíření SNP-BLUP, jelikož dávají efektům markerů i jiné než normální rozdělení a různé rozptyly. Dá se totiž očekávat, že některé markery, jež jsou blízko u příčinných mutací či majorgenů, mají poměrně velký efekt, zatímco jiné mají efekt nulový (Meuwissen et al. 2016).

Termín genomická BLUP (**GBLUP**) se objevil v návaznosti na práci VanRadena (2008), jenž představil tři metody sestavení matice genomické příbuznosti (**G**). Ta obecně zpřesní tradiční očekávanou (tj. rodokmenovou) příbuznost mezi jedincem i a j díky vypočtení korelace mezi genotypy napříč všemi SNP (Meuwissen et al. 2016). Ze způsobů pro sestavení matice **G** se nejvíce rozšířila VanRadena metoda 1 (Masuda 2019). V této metodě je nejprve třeba sestavit matice **M** a **P**. Matice **M** má rozměry $n \times m$, kde n je počet jedinců (řádků) a m je počet lokusů (sloupců). Prvky **M** jsou hodnoty -1, 0 a 1 pro homozygota A, heterozygota, respektive homozygota B. Matice **P** má znovu rozměry $n \times m$ se shodnými hodnotami pro jednotlivé sloupce. Hodnoty sloupce j jsou vypočteny jako: $P_j = 2(p_j - 0,5)$, kde p_j je frekvence druhé alely na lokusu j . Zadruhé je sestavena matice **Z** jako: $Z = M - P$. Zatřetí je sestavena matice **G** s rozměry $n \times n$ jako:

$$\mathbf{G} = \frac{\mathbf{z}\mathbf{z}^T}{2 \sum p_j(1 - p_j)}, \quad [27]$$

kde jmenovatel zajišťuje škálování \mathbf{G} tak, aby byla analogická k matici rodokmenové příbuznosti \mathbf{A} (VanRaden 2008). [Poznámka: Index i , který se vyskytuje ve VanRadenově článku, byl nahrazen indexem j , který je v maticovém zápisu běžnější pro označení sloupců.]

Vzhledem k vývoji genomické selekce se pro GBLUP zpětně začalo používat také podrobnější označení víceřadová/dvouřadová GBLUP (msGBLUP/tsGBLUP). Prvním krokem msGBLUP je předpověď EBV pro všechna zvířata pomocí klasické BLUP. Nato jsou pro genotypované jedince vypočteny pseudofenotypy (či odregresované hodnoty), které jsou zbaveny informací od předků a vrstevníků, aby ve výsledku nebyly započítány dvakrát. Například pseudofenotyp pro plemenného býka vznikne z průměru mléčné užitkovosti jeho dcer. Ve druhém kroku přichází na řadu samotná GBLUP, která zahrnuje pseudofenotypy jako závisle proměnnou a také matici genomické příbuznosti. Výsledkem GBLUP jsou takzvané přímé genomické hodnoty. Tyto předpovědi pak spolu s odchylkou v rodičovském průměru tvoří GEBV (VanRaden et al. 2009; Misztal & Legarra 2017; Piccoli et al. 2018). Při stejném nastavení parametrů by měly být GEBV z GBLUP identické jako předpovědi z SNP-BLUP (Meuwissen et al. 2016).

Hayes et al. (2009) navrhl jiný postup, který ovšem také předpovídá GEBV ve více krocích. Nejprve je genom rozdělen do malých segmentů podle SNP. V referenční populaci, která obsahuje pouze zvířata s fenotypy i genotypy, jsou předpovězeny efekty těchto segmentů. V následujících generacích pak lze každému genotypovanému zvířeti přiřadit odpovídající segmenty, a tudíž předpovědět GEBV. V další víceřadové metodě jsou genomické předpovědi (nazvané molekulární plemenné hodnoty) využity jako korelovaná vlastnost (MacNeil et al. 2010; Kachman et al. 2013).

Víceřadová GBLUP přinesla zpřesnění předpovědí, ale zároveň se projeví některé nedostatky (Aguilar et al. 2010). GEBV je možné stanovit pouze u genotypovaných zvířat, tedy typicky u býků, jimž chybí vlastní mléčná užitkovost. Pro spojení celého procesu se proto využívají pseudofenotypy, které jsou samy zatíženy chybou. Ve výsledku tak dochází ke ztrátě a zkrácení informací (Legarra et al. 2014). Patry & Ducrocq (2011) nadto upozornili na problém předvýběru kandidátů pro plemenitbu. Pokud jsou GEBV stanoveny krátce po narození a na jejich základě je proveden předvýběr býčků, následné záznamy o užitkovosti pocházejí ze vzorku zvířat, u nějž je porušen předpoklad náhodného mendelistického výběru.

2.5.4. Jednokroková GBLUP

Východisko z předešlých omezení navrhli Legarra et al. (2009), Misztal et al. (2009) a Christensen & Lund (2010), když představili sjednocení rodokmenové a genomické matice do matice realizované příbuznosti (\mathbf{H}). Výpočty se vešly do jednoho kroku, a proto tato metoda získala označení jednokroková GBLUP (ssGBLUP; Misztal & Legarra 2017).

Nejjednodušší animal model ($\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$) má pomocí ssGBLUP řešení:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{H}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}, \quad [28]$$

kde $\hat{\mathbf{b}}$ je vektor odhadů pro pevné efekty, $\hat{\mathbf{a}}$ je vektor předpovědí pro náhodný genetický efekt zvířete, \mathbf{X} a \mathbf{Z} jsou incidenční matice těchto efektů, \mathbf{H} je matice realizované příbuznosti, \mathbf{y} je vektor pozorování, σ_e^2 je chybový rozptyl a σ_a^2 je aditivní genetický rozptyl. Inverze matice \mathbf{H} je získána ze vztahu:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad [29]$$

kde \mathbf{A} je kompletní matice rodokmenové příbuznosti, kterou lze rozdělit do čtyř bloků (\mathbf{A}_{mn}^{-1}), a \mathbf{G} je matice genomické příbuznosti, která zahrnuje pouze genotypovaná zvířata. Od matice \mathbf{G} tak musí být odečteny příbuznosti v \mathbf{A}_{22} , aby se předešlo dvojímu započítání rodokmenových informací. Zároveň platí, že v případě genotypování všech zvířat je ssGBLUP ekvivalentní k GBLUP podle VanRadena (2008), naopak v případě absence genotypování je ssGBLUP ekvivalentní k BLUP (Lourenco et al. 2020).

Zjednodušení procesu poskytlo několik výhod v porovnání s víceřadovými metodami. GEBV jsou předpovídány všem jedincům (tj. i negenotypovaným). Došlo k eliminaci některých předpokladů a parametrů nezbytných pro msGBLUP (Aguilar et al. 2010), jako například nutnosti získávat odregresované hodnoty a složitě řešit dvojí započítávání některých zdrojů informací (Misztal et al. 2020). Je potlačeno vychýlení způsobené předvýběrem (Patry & Ducrocq 2011). Matice \mathbf{H} teoreticky poskytuje zpřesnění příbuzností mezi všemi jedinci, takže jsou automaticky nastaveny váhy pro kombinaci různých druhů informací do výsledné předpovědi. Rovněž může být odhalen skrytý inbreeding a v konečném důsledku je dosaženo přesnějších GEBV (Amaya Martínez et al. 2019).

Christensen & Lund (2010) na simulované populaci prasat demonstrovali průměrnou přesnost EBV ve výši 0,35 pro BLUP, 0,59 pro msGBLUP a 0,66 pro ssGBLUP. Piccoli et al. (2018) na simulované populaci masného skotu obdrželi podobné rozdíly mezi tradiční BLUP ($r = 0,34$)

a genomickými metodami ($r = 0,53-0,70$), ale mezi víceřadovými a jednokřadovými genomickými postupy byly rozdíly pouze okolo 1 % ve prospěch ssGBLUP. Zvýšení přesnosti v rámci 1–3 % pozorovali Koivula et al. (2012) a Afrazandeh et al. (2022) na reálných datech skotu, přičemž podpořili hypotézu o nižším vychýlení předpovědí v případě ssGBLUP. Guarini et al. (2019) analyzovali rychlost dojivosti u kanadského holštýnského skotu a na základě domácích dat uvedli průměrnou přesnost 0,47 pro BLUP, 0,59 pro msGBLUP a 0,64 pro ssGBLUP. Příbyl et al. (2015) doporučili ssGBLUP také kvůli nejvyšší korelaci domácích předpovědí s předpověďmi mezinárodního hodnocení. Převaha ssGBLUP nad jinými metodami se projevuje obzvláště při analýze vlastností s nízkým koeficientem dědivosti (Van Eenennaam et al. 2014; Guarini et al. 2018).

Hlavní nevýhody ssGBLUP oproti víceřadovými či bayesovským metodám se dají shrnout do dvou bodů. První nevýhodou jsou vysoké nároky na výpočetní techniku a na přizpůsobení programů, což vyplývá zejména z teoretické potřeby invertovat matici \mathbf{G} . Druhou nevýhodou vnáší předpoklad uniformního příspěvku všech markerů, jenž koresponduje s infinitesimálním modelem (Van Eenennaam et al. 2014; Amaya Martínez et al. 2019). Liu et al. (2014) tak navrhli další postup, jednokřadový SNP model, který později získal označení ssSNP-BLUP. V něm jsou na rozdíl od ssGBLUP přímo odhadovány efekty jednotlivých SNP, a dokonce je možné jim přiřadit různé podíly vysvětleného genetického rozptylu. Jelikož matice \mathbf{B} , která obsahuje markery, má oproti matici \mathbf{G} v ssGBLUP jednoduchou strukturu, mohou být vyvinuty efektivnější výpočetní strategie. A zároveň ssSNP-BLUP odstraňuje nevýhody víceřadových metod, neboť analyzuje všechny fenotypy i genotypy najednou.

Přesto ssGBLUP zůstává stěžejní metodou u většiny druhů hospodářských zvířat – zvláště v populacích, kde nejsou genotypována všechna hodnocená zvířata (Bermann et al. 2022). Výpočetní náročnost se daří zvládat díky různým algoritmům, které pracují s řídkými maticemi, rozkládáním hustých informačních bloků nebo skupinami nejdůležitějších zvířat (Aguilar et al. 2018; Misztal et al. 2020). Také se podařilo sjednotit metodiku s redukovaným animal modelem, který omezuje počet rovnic díky využití nejdůležitějších jedinců a poté provádí zpětné výpočty pro zbytek souboru (Nilforooshan 2023). Zároveň je možné dopočítat efekty markerů drobnou úpravou vzorce:

$$\hat{\mathbf{m}} = \frac{\sigma_m^2}{\sigma_a^2} \mathbf{DZ}^T \mathbf{G}^{-1} \hat{\mathbf{a}}, \quad [30]$$

kde $\hat{\mathbf{m}}$ je vektor předpovědí pro efekty markerů, σ_m^2 je rozptyl efektů markerů, σ_a^2 je aditivní genetický rozptyl, \mathbf{D} je kovarianční matice efektů markerů, \mathbf{Z} je popsána v rovnici [27], \mathbf{G} je matice genomické příbuznosti a $\hat{\mathbf{a}}$ je vektor EBV. Tohoto kroku využívá celogenomová asociační studie založená na ssGBLUP (ssGWAS), která má ve srovnání s klasickou GWAS větší sílu testu, protože

nejsou využity pouze informace od genotypovaných zvířat (Bermann et al. 2022). Zároveň Lourenco et al. (2017) předvedli, že efekty markerů lze znovu doplnit do modelu pomocí váhových koeficientů, čímž vzniká vážená ssGBLUP (WssGBLUP). Ovšem v této simulační studii nastalo zlepšení oproti základnímu modelu jen v případě malého počtu genotypovaných zvířat nebo malého počtu QTL. Ani Brzáková et al. (2022) na reálných datech nepozorovali výrazný přínos WssGBLUP. Nicméně tyto výsledky nejsou příliš překvapivé, jelikož už dřívější porovnání ssGBLUP s ostatními metodami ukázala, že zohlednění různých efektů QTL zvýšilo přesnost předpovědi poměrně málo (Van Eenennaam et al. 2014).

Za nevýhodu ssGBLUP by se dala považovat i skutečnost, že sestavení matice \mathbf{H} nemusí být jednoznačné kvůli singularitě matice \mathbf{G} nebo její omezené slučitelnosti s příslušným blokem matice \mathbf{A} . Nevhodně sestavená matice \mathbf{H} se pak promítá do věrohodnosti a výkonnosti modelu. Nicméně tuto skupinu nedostatků omezuje nebo úplně odstraňuje nastavení níže uvedených parametrů (Christensen et al. 2012; Misztal et al. 2020; Nilforooshan 2022):

- 1) Opravení matice \mathbf{G} pomocí spojení s příslušným blokem rodokmenové matice (angl. **blending**): $\mathbf{G} = \alpha \mathbf{G}_0 + (1 - \alpha) \mathbf{A}_{22}$, kde \mathbf{G}_0 značí původní genomickou matici a váhový koeficient α nabývá hodnot 0–1. Takto jsou do matice \mathbf{G} zahrnuty polygenní efekty, tedy část proměnlivosti, kterou nelze vysvětlit markery. Zároveň tento krok může sloužit k odstranění singularity \mathbf{G}_0 , kvůli níž by nešla invertovat. Tento problém se běžně objevuje při nevhodném poměru počtu genotypovaných zvířat ku počtu SNP nebo při vysoké podobnosti mezi některými zvířaty (Lourenco et al. 2020).
- 2) Škálování (angl. **scaling**) inverzních submatic: $\tau \mathbf{G}^{-1} - \omega \mathbf{A}_{22}^{-1}$, kde koeficienty τ a ω zajišťují kompatibilitu s \mathbf{A}^{-1} . Genotypování zvířat neprobíhá náhodně, a proto vzniká rozdíl mezi genetickou bází rodokmenové a genomické matice. Navíc nastává nesoulad kvůli výpočetním algoritmům, pokud \mathbf{A}^{-1} nezohledňuje inbreeding. Škálování tedy potlačuje potenciální nadhodnocení GEBV (Lourenco et al. 2020; Bermann et al. 2022).
- 3) Regresní ladění (angl. **tuning**) inverzních submatic. Matice \mathbf{A}_{22} často zahrnuje mnoho generací a předpovědi jsou tak regresovány směrem ke generacím zakladatelů, u nichž chybí informace o alelových četnostech. Naopak matice \mathbf{G} regresuje předpovědi k nejnovějším generacím, ze kterých pocházejí genotypovaná zvířata. Za tímto účelem vznikly matematické postupy, které přibližují průměry obou matic (Masuda 2019; Nilforooshan 2022).

2.5.5. Genomická selekce u prasat

Důležitým milníkem genomické selekce u prasat bylo uvedení dostupných DNA čipů zejména ve spolupráci se společností Illumina (Ramos et al. 2009) a osekvenování celého genomu (Groenen

et al. 2012). Preferovanou metodou pro předpověď GEPH se stala ssGBLUP, neboť komerčním společnostem poskytla důležité praktické výhody (Legarra et al. 2014; Knol et al. 2016). Ovšem kvůli biologickým a zootechnickým odlišnostem v porovnání s dojeným skotem čelí šlechtění prasat při zavádění genomické selekce dalším výzvám. V literatuře jsou zmiňovány tři hlavní a úzce propojené aspekty: 1) vliv na genetický zisk; 2) návratnost investice; a 3) přenos selekčního efektu do finálních hybridů.

Jak genomická selekce ovlivní genetický zisk? Genomická selekce má největší dopad u mladých zvířat. Zatímco pro šlechtitele skotu představuje výhodu zejména zkrácení generačního intervalu při zanedbatelné ztrátě přesnosti, u prasat má genomická selekce oporu ve zvýšení přesnosti EBV při zachování generačního intervalu (Knol et al. 2016; Mote & Rothschild 2020). Generační interval u prasat je nižší než dva roky a jeho další snížení se jeví z provozního hlediska jako nepraktické (Samorè & Fontanesi 2016). Přesto i u prasat lze zvýšit genetický zisk prostřednictvím generačního intervalu, pokud se vyplatí genotypování většího množství selat (z jednoho vrhu). V takovém případě je možné vlastním sourozencům velmi brzy přiřadit individuální hodnoty, a nikoliv čekat na projev vlastní užitkovosti nebo reprodukci dcer (Mote & Rothschild 2020).

Hlavní potenciál genomické selekce ovšem spočívá ve zpřesnění předpovědi, a to zejména u vlastností, které mají nízkou dědivost (reprodukce); mají omezený počet záznamů (pouze u jednoho pohlaví či odolnost vůči chorobám); případě jsou zaznamenatelné až po více letech (dlouhověkost) nebo po porážce příbuzných jedinců (kvalita masa; Samorè & Fontanesi 2016). Z některých výsledků vyplývá, že se podařilo přiblížit nárůstu v přesnosti, jaký předvedli Christensen & Lund (2010) ve své simulaci. Mote & Rothschild (2020) udávají, že průměrná přesnost GEBV pro vlastnosti popisující velikost vrhu je o 52–68 % vyšší než v případě EBV. Christensen et al. (2012) dosáhli zvýšení o 18 % pro přírůstek a o 57 % pro konverzi. Ibáñez-Escriche et al. (2014) dokonce uvádějí zvýšení o 85 % pro přírůstek a o 72 % pro výšku svalu. Největší rozdíly byly potvrzeny u kandidátů pro selekci bez vlastní užitkovosti nebo potomstva, tedy při porovnání EBV na základě rodičovského průměru a GEBV. Oh et al. (2016) nebo Salek Ardestani et al. (2021) odhadli až trojnásobné zpřesnění předpovědi pro počet selat, respektive pro přírůstek.

Nicméně výsledky jiných autorů takto pozitivní závěry neměly. Song et al. (2019) u sedmi vlastností tělesného rámce pozorovali průměrně pouze 1% zvýšení individuálních přesností oproti tradiční BLUP. Fangmann et al. (2017) zpřesnili konvenční předpovědi u mladých zvířat o 3–4 %. V obou těchto případech ovšem bylo k dispozici poměrně málo genotypovaných zvířat.

Vedle přesnosti a generačního intervalu působí genomická selekce také na aditivní genetickou proměnlivost. Hidalgo et al. (2020) na intenzivně selektované populaci prasat ukázali, že genomická

selekce přispěla ke ztrátě genetické proměnlivosti, a tudíž k poklesu dědivosti. Navíc se zvýraznila negativní genetická korelace mezi růstovou schopností a fitness. Klesající proměnlivost v kombinaci s minimálním nárůstem v přesnosti, který lze očekávat po ogenotypování velké části populace, se negativně podepíší na výši genetického zisku. Wientjes et al. (2023) v této souvislosti varují, že genomická selekce oproti klasické urychluje fixaci alel včetně těch nechtěných. Jednou z příčin je větší riziko genetického draftu (angl. genetic hitchhiking).

Jak zajistit návratnost investice? V porovnání zejména s býky dojeného skotu je u prasat více kandidátů pro selekci, přičemž jejich ekonomická hodnota je výrazně nižší. Náklady na genotypování se tak stávají větším problémem (Ibáñez-Escriche et al. 2014; Knol et al. 2016). SCHP v současnosti počítá s cenou asi 1000 Kč na zvíře (E. Žáková, osobní komunikace). To je sice výrazně méně než v počátcích zavádění ssGBLUP u prasat (Samorè & Fontanesi 2016), ale v kontextu ceny zhruba 5000 Kč za průměrné JUT třídy S (TIS SZIF 2024) se stále jedná o podstatnou položku.

Proto je zapotřebí uskutečnit předvýběr například po projevu vlastní užítkovosti nebo alespoň pomocí rodičovského průměru. V simulaci Henryona et al. (2012) stačilo ogenotypovat pouze 5 % selekčních kandidátů vybraných na základě fenotypu, aby bylo dosaženo 86 % potenciálního genetického zisku. Další strategií, jež se uplatňuje i v komerčních společnostech, je využití levnějších čipů s menším počtem SNP u části populace a následná imputace podle jedinců genotypovaných čipy s vysokou hustotou (Samorè & Fontanesi 2016). Huang et al. (2012) zredukovali projektované náklady na šestinu, když použili 60tisícový čip jen pro otce a zbytek důležité populace (matky, prarodiče, kandidáty) genotypovali čipem s pouze 384 SNP. Přitom byla zachována 93,5% spolehlivost imputace. Čipy u prasat obecně nemusí být tak husté jako u skotu, protože vazebná nerovnováha je mnohem vyšší (Veroneze et al. 2013).

Jak přenést genetický zisk do užitkových chovů? I když selekce probíhá u čistokrevných zvířat, konečným záměrem je přenesení genetického zisku k finálním hybridům. Kvůli dominanci, imprintingu, interakci genotypu s prostředním nebo měnícím se genovým vazbám se korelace mezi výsledky čistokrevné populace a populace kříženců obvykle pohybují od 0,45 do 0,87 (Stock et al. 2020). Proto byly navrženy komplexní modely, které obsahují fenotypy kříženců nebo neaditivní efekty. Genomická selekce pak přináší možnost přesnější předpovědi těchto efektů, a k tomu není nutné zaznamenávat křížencům rodokmen (Ibanez-Escriche & Simianer 2016; Samorè & Fontanesi 2016). Přesto je taková metodika velmi náročná provozně i výpočetně (Lopes 2016). Esfandyari et al. (2020) zjistili, že odhady genetických korelací mezi čistokrevnými zvířaty a kříženci pro různé vlastnosti byly při použití ssGBLUP zatíženy mnohem menší chybou než při použití BLUP. Samotné hodnoty korelačních koeficientů však nezvýhodnily ani jeden model. Navíc v této studii bylo ogenotypováno okolo 14 % čistokrevných zvířat a u některých vlastností až 100 % kříženců.

3. Cíl práce a vědecká hypotéza

3.1. Hlavní a vedlejší cíl práce

Hlavním cílem této práce je předpověď genetického založení prasat evidovaných v kontrole užitečnosti a navržení metodického postupu pro předpověď genomických plemenných hodnot v rámci programu CzePig. **Vedlejším cílem** je zhodnocení vybraných kandidátních genů pro možnost využití v markery asistované selekci.

3.2. Vědecká hypotéza

Molekulárněgenetické informace přispějí ke zpřesnění předpovědi genetického založení v domácích populacích prasat.

4. Metodika

4.1. Asociační analýza polymorfizmu v *MC4R*

4.1.1. Fenotypová data

První studie (Calta et al. 2023), která byla zahrnuta do této dizertační práce, testovala možný efekt interakce mezi polymorfizmem c.1426G>A (p.Asn298Asp) v genu *MC4R* a krmnou dávkou. Pro tento účel byla využita data od 558 prasat z pokusů provedených na testační stanici ČZU (CZ21038206). Zvířata byla ustájena v klimatizované stáji ve dvoučtých kotcích podle pohlaví. Průměrná živá hmotnost na začátku testu činila 25,6 kg ($s = 3,7$ kg) a na konci testu 110,4 kg ($s = 9,6$ kg). Délka testu byla závislá na dosažení porážkové hmotnosti, a tak se pohybovala od 10 do 15 týdnů (průměrný věk 157,4 dnů, $s = 17,9$ dnů). Zvířata byla krmena kompletní krmnou směsí. V první fázi, od 25 do 65 kg živé hmotnosti, byla krmná dávka stejná pro obě krmné kategorie (13,2 MJ ME/kg). Ve druhé fázi (65 až 110 kg) zvířata v kategorii *ad libitum* (ADL) dostávala neomezenou krmnou dávku, která empiricky odpovídá 13,4 MJ ME/kg. Oproti tomu restringovaná kategorie (RES) měla přísun 12,9 MJ ME/kg. Tabulka 4.1-A zachycuje rozložení zvířat napříč sedmi testovanými skupinami, které byly využity pro vytvoření finálního datového souboru.

Tabulka 4.1-A: Struktura testovaných zvířat.

Test	Křížení/plemeno	N	N dle pohlaví	N dle krmení
1	(DU×LWS) × (CLA×CLW)	72	36♀ + 36♂	72 ADL
2	(PE×HA) × (CL×CLW)	72	36♀ + 36♂	72 ADL
3	PE × (CLA×CLW)	65	65♂	33 ADL + 32 RES
4	Topigs Duroc	68	22♀ + 23♂ + 23♂	68 ADL
5	DanBred	69	18♀ + 18♂ + 33♂	69 ADL
6	(PE×LWS) × (CLA×CLW)	139	70♀ + 69♂	52 ADL + 87 RES
7	LWS × (CLA×CLW)	73	24♀ + 49♂	73 ADL
<i>celkem</i>		558	271♀ + 182♂ + 105♂	439 ADL + 119 RES

Symbole a zkratky: ♀ = prasnička; ♂ = vepřík; ♂ = kaneček; ADL = *ad libitum* krmná dávka; CLA = česká landrase; CLW = české bílé ušlechtilé; DU = duroc; HA = hampshire; N = počet zvířat; PE = pietrain, RES = restringovaná krmná dávka

Během výkrmu byly sledovány obvyklé užitkové vlastnosti jako živá hmotnost na konci testu (BW), průměrný denní příjem krmiva (ADFI), poměr konverze krmiva (FCR) a přírůstek (ADG). Po porážce byla zjištěna hmotnost jatečně upraveného těla (CW), jatečná výtěžnost (JV) a výška hřbetního tuku (BF) spolu s výškou svalu *longissimus thoracis* (MD) dvoubodovou metodou (ZP)

pro dopočet podílu libové svaloviny (LMP). Následně byly jatečné kusy podrobeny disekci (Walstra & Merkus 1996) pro zjištění hmotnosti hlavních masitých částí (LC) a jejich složení (tukové krytí vs. libové maso). Vlastnosti kvality masa, tedy pH a teplota 45 min *post mortem*, byly měřeny ve svalech *longissimus thoracis* a *semimembranosus*. Dále byla 24 h *post mortem* měřena ztráta odkapem (DL; van Laack & Smulders 1992).

4.1.2. Genotypování

Po porážce byly odebrány vzorky krve, které byly uchovány ve zkumavkách s K₂EDTA při teplotě -20 °C. Pro izolaci DNA byl použit DNA Mini Kit for Blood/Cultured Cells (Geneaid Biotech). Požadované úseky DNA byly namnoženy polymerázovou řetězovou reakcí (PCR) s primery ohraničujícími zkoumaný polymorfismus Asn298Asp v *MC4R* (Kim et al. 2000a): A = 5'-TACCCTGACCATCTTGATTG-3'; B = 5'-ATAGCAACAGATGATCTCTTTG-3'. Průběh PCR byl nastaven na 2 min při 95 °C a následně 30 cyklů s kroky 30 s při 95 °C, 30 s při 55 °C a 30 s při 68 °C. Pro zjištění genotypů pak byla využita metoda polymorfismu délky restrikčních fragmentů (RFLP) s polymerázou *TaqI*, přičemž fragmenty byly vizualizovány na 1,5% agarovém gelu.

4.1.3. Statistické vyhodnocení

Vzhledem k omezenému množství fenotypových údajů a nedostupnosti rodokmenu byl zvolen jednoduchý lineární model pouze s pevnými efekty, který byl vyhodnocen pomocí PROC GLM v SAS/STAT[®], verze 9.4:

$$Y_{ijkmno} = \mu + MC4R_i + KD_j + MC4R \times KD_k + PLEM_m + POHL_n + VYK_o + \beta CW + \varepsilon_{ijkmno},$$

kde μ je celkový průměr; $MC4R_i$ je efekt genotypu v *MC4R* ($i \in \langle 1; 3 \rangle$: GG, GA, AA); KD_j je efekt kategorie podle krmné dávky ($j \in \langle 1; 2 \rangle$: ADL, RES); $MC4R \times KD_k$ je efekt interakce genotypu a krmné dávky ($k \in \langle 1; 6 \rangle$: odpovídá možným kombinacím $i \times j$); $PLEM_m$ je efekt kombinace křížení či plemene ($m \in \langle 1; 7 \rangle$: odpovídá pořadí testu tabulce 4.1-A); $POHL_n$ je efekt pohlaví ($n \in \langle 1; 3 \rangle$: prasnička, vepřík, kaneček); VYK_o je efekt délky výkrmu ($o \in \langle 1; 6 \rangle$: odpovídá porážce po 10.–15. týdnu); βCW je regresní koeficient spolu s proměnnou pro jatečnou hmotnost; a ε_{ijkmno} je chyba odhadu. Úplný model byl využit pro JV, LMP, BF a LC. Pro zbylé vlastnosti byl využit model bez CW.

Proměnlivost vysvětlená modelem byla vypočtena jako koeficient determinace ($R^2 = 1 - \frac{SSE}{SST}$, kde SSE je suma čtverců reziduálních chyb a SST je celková suma čtverců). Významnost prediktorů byla testována pomocí F-testu a vyjádřena p -hodnotami. Velikost efektu pro hlavní prediktor, interakci $MC4R \times KD$, byla vyjádřena pomocí parciální čtvercové éty ($\eta^2 = \frac{SSP}{SSP+SSE}$),

kde SSP je suma čtverců daného prediktoru a SSR je suma čtverců reziduálních chyb). Průměry jednotlivých tříd sledované interakce byly zachyceny jako průměry nejmenších čtverců spolu se střední chybou a graficky vyjádřeny jako 95% intervaly spolehlivosti pomocí programovacího jazyka Python, verze 3.11 (knihovny Pandas, Matplotlib a NumPy). Rozdíly mezi průměry byly rovněž popsány jak z hlediska významnosti (p -hodnotami), tak z hlediska standardizované velikosti efektů (Cohenovým $d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{MSE}}$, kde \bar{x}_1 a \bar{x}_2 jsou průměrné odhady dvojice tříd a MSE je průměrný čtverec reziduálních chyb).

4.2. Genetické hodnocení výšky tuku a svalu

4.2.1. Výběr BLUP-AM: data

Hlavní analýza této dizertační práce byla rozdělena na dvě části, přičemž první část se týkala návrhu vhodného konvenčního animal modelu (Calta et al. 2024). Pro tento účel byla od Svazu chovatelů prasat, z.s. poskytnuta užitková data 81 819 jedinců českého bílého ušlechtilého (CLW) a 37 366 jedinců české landrase (CLA), kteří ukončili polní test mezi lednem 2013 až zářím 2022. Polní test odpovídal metodice pro kontrolu užitkovosti (SCHP 2010). Zvířata byla zařazena do testu ve věku 80–88 dnů. Kotce obsahovaly 6–15 zvířat stejného pohlaví a plemene s předepsanou plochou 0,8–1,0 m² na zvíře. Technika ustájení musela splňovat standardy pro čistokrevná plemena s *ad libitum* přístupem k vodě a standardizované krmné směsi. Délka testu byla pro prasničky průměrně o týden kratší (56 dnů) než pro kanečky (63 dnů). Test byl ukončen ve věku 128–159 dnů.

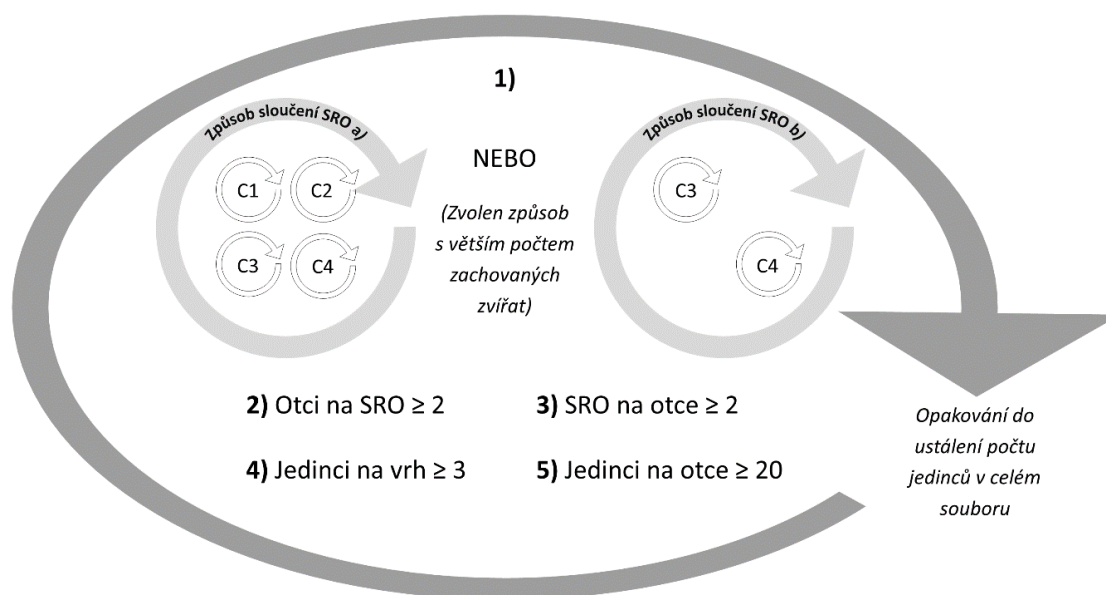
Do analýzy byly jako vysvětlující proměnné zahrnuty: identifikační čísla jedince, otce a matky, plemeno, pohlaví, datum narození, datum začátku testu, datum konce testu, číslo stáda, číslo vrhu, ultrazvukový přístroj (Sonomark SM 100 M a Mindray DP-20), počet všech narozených selat, počet živě narozených selat, pořadí vrhu, hmotnost na začátku testu a věk na konci testu. Z dostupného rodokmenu bylo využito až 21 generací předků kvůli zahrnutí genetických skupin. Genetické skupiny zastupovaly neznámé rodiče a byly vytvořeny na základě informací o plemeni, zemi původu a období narození.

Vysvětlovanými proměnnými byly hmotnost na konci testu (BW), průměrný denní přírůstek (ADG) a *in vivo* ultrazvuková měření výšky hřbetního tuku (BF) a výšky svalu (MD). Hodnota BF byla vypočtena jako průměr měření v bodech *A* a *B*, zatímco MD byla měřena pouze v bodě *B*. Bod *A* se nachází v bederní oblasti a bod *B* v hrudní oblasti, v obou případech 70 mm ventrálně od středové osy hřbetu. Přesná definice bodů je uvedena v metodice Stibala (2016).

Data pro statistické vyhodnocení byla upravena pomocí programu SAS/STAT®, verze 9.4. Extrémní pozorování byla definována pomocí percentilů (méně než 0,25, nebo více než 99,75) a smazána. Následně byly vytvořeny skupiny zvířat se stejnými podmínkami prostředí jako sdružený

efekt stáda, roku a období (SRO). Období bylo definováno měsícem ukončení testu, ovšem za účelem zachování co největšího počtu pozorování mohla být každá skupina spojena až se dvěma přilehlými skupinami (tj. utvořit období až tří měsíců). Kvůli zajištění propojenosti mezi skupinami musely být splněny minimální požadavky: 20 zvířat v jednom SRO, dva otci v jednom SRO, dvě SRO na otce, 20 zvířat na otce a tři testovaná zvířata v jednom vrhu (obrázek 4.2-A). Tyto optimalizované datové soubory byly použity pro odhad populačních genetických parametrů. Změny v počtu pozorování spolu s popisnými statistikami hlavních vlastností zachycuje tabulka 4.2-A.

Obrázek 4.2-A: Schéma vnořených cyklů (DO loops) v programu SAS pro optimalizaci SRO.



Poznámky: C1 = cyklus pro každé SRO: když $n < 20 \rightarrow$ sluč s předchozím SRO (pokud bude dodržena maximální délka období); C2 = jako C1, ale slučování s následujícím SRO; C3 = jako C1, ale v každém cyklu navíc možnost sloučení až se dvěma předchozími SRO; C4 = jako C3, ale pro slučování s následujícími SRO

4.2.2. Výběr BLUP-AM: vyhodnocení

Nynější selekční index pro mateřská plemena zahrnuje dvě produkční vlastnosti (přírůstek a zmasilost), a proto byly modely v této práci navrženy obdobným způsobem. Víceznakový model tedy jako vysvětlované proměnné kromě BF a MD obsahoval také ADG, nebo BW (se zahrnutím regrese na věk). Potenciálně významné prediktory byly určeny principem postupné eliminace v programu SAS/STAT® (PROC GLMSELECT). Věk a BW byly dále posouzeny pomocí parciální korelace a faktoru zvětšení rozptylu (VIF) kvůli možné kolinearitě.

Následně byly navrženy tříznakové animal modely s různými kombinacemi faktorů (MME po vzoru [13] až [16] v kapitole 2.3.5.). Nejvíce informativní varianty jsou shrnuty v Tabulce 4.2-B. Modely jsou rozděleny do dvou hlavních kategorií podle doprovodné vlastnosti (A = ADG, B = BW). Obě kategorie sestávají ze čtyř variant modelu s rostoucí komplexností.

Odhad (ko)variancí byl založen na optimalizovaných datových souborech. Pro řešení MME jednotlivých modelů sloužila metoda restringované maximální věrohodnosti (REML) v programu REMLF90 (Misztal et al. 2018) s konvergenčním kritériem 10^{-12} . Pro výpočet koeficientů dědivosti a genetických korelací byly použity vzorce [3] a [5] uvedené v kapitole 2.3.2.

Porovnání modelů bylo primárně založeno na metrikách lineární regrese (LR; Legarra & Reverter 2018), konkrétně na (standardizovaném) vychýlení, disperzi a populační přesnosti předpovědi EBV. Zohledněna byla také prediktivita (PrA; Lourenco 2018) a Akaikovo informační kritérium (AIC; Field & Miles 2010) zjištěné z REMLF90 jako difference od základního modelu (D_AIC). Výkladu těchto metrik se věnuje kapitola 2.4.1 (rovnice [17] až [23]).

Validační soubory potřebné pro LR a PrA byly vytvořeny v souladu s principem dopředné validace (obrázek 2.4-A). Kompletní soubor (*w*) představoval základní data (81 829 CLW a 37 366 CLA). V tomto případě byly skupiny SRO tvořeny pouze s požadavkem alespoň 10 zvířat a období mohlo sestávat z více jak tří měsíců, aby byla zachována všechna pozorování. Parciální datový soubor (*p*) vznikl na základě kompletního souboru tak, že byly skryty fenotypy zvířat v posledních třech letech (2020–2022). Do validační kohorty (*val*) byla zahrnuta ta zvířata se skrytými fenotypy, jejichž oba rodiče měli dostupné fenotypy v obou souborech (7 932 CLW a 4 573 CLA). Popisné statistiky validačních kohort lze vidět v tabulce 4.2-A. Pro předpověď EBV a prekorekci fenotypů byl použit program BLUPF90, respektive PREDICTF90 (Misztal et al. 2018).

Tabulka 4.2-A: *Popisné statistiky vybraných vlastností podle plemene v základním, optimalizovaném a validačním souboru.*

Vlastnost	České bílé ušlechtilé			Česká landrase		
	<i>základní</i> \bar{x} (s)	<i>optimal.</i> \bar{x} (s)	<i>validace</i> \bar{x} (s)	<i>základní</i> \bar{x} (s)	<i>optimal.</i> \bar{x} (s)	<i>validace</i> \bar{x} (s)
počet	81 819	67 940	7 932	37 366	31 936	4 573
věk [dny]	99,17 (11,99)	99,80 (11,66)	100,06 (13,37)	104,40 (9,84)	104,97 (9,08)	105,78 (9,57)
BW [kg]	143,95 (5,93)	143,94 (5,91)	143,52 (5,46)	145,35 (5,57)	145,38 (5,54)	145,06 (5,67)
ADG [g]	689,10 (79,42)	693,46 (76,81)	697,01 (87,10)	718,22 (61,87)	722,04 (56,34)	729,11 (58,58)
BF [mm]	6,83 (1,31)	6,80 (1,25)	6,20 (0,97)	6,81 (1,32)	6,77 (1,26)	5,98 (0,92)
MD [mm]	50,00 (5,94)	50,04 (5,69)	46,42 (6,14)	51,09 (5,45)	51,12 (5,21)	47,90 (5,36)

Zkratky: ADG = průměrný denní přírůstek; BW = živá hmotnost; BF = výška špeku; MD = výška svalu;
s = směrodatná odchylka; \bar{x} = průměr

Tabulka 4.2-B: Podoba nejvíce informativních tříznakových modelů.

Model	Pevné efekty	Náhodné efekty
A1	pohlaví + SRO + UZ	vrh + zvíře (A)
A2	pohlaví + rok + stádo + UZ	SRO + vrh + zvíře (A)
B1	pohlaví + SRO + UZ + věk (R)	vrh + zvíře (A)
B2	pohlaví + rok + stádo + UZ + věk (R)	SRO + vrh + zvíře (A)
_VS	<i>jako základní model</i>	... + vrh×SRO
_OS	<i>jako základní model</i>	... + otec×stádo
_MAT	<i>jako základní model</i>	... + EPM + matka (A)

Modely: A1 a A2 = základní modely pro ADG, BF a MD; B1 a B2 = základní modely pro BW, BF a MD; _VS, _OS a _MAT = rozšíření základních modelů o další náhodné efekty

Zkratky: A = náhodný efekt s maticí příbuznosti; ADG = průměrný denní přírůstek, BF = výška hřbetního tuku, BW = živá hmotnost; EPM = trvalé prostředí matky; MAT = maternální efekty; MD = výška svalu; OS = otec×stádo; SRO = stádo-rok-období; R = pevná regrese; UZ = ultrazvukový přístroj (pouze pro BF a MD); VS = vrh×SRO

4.2.3. Porovnání BLUP a ssGBLUP: data

4.2.3.1. Fenotypová data

Druhá část hlavní analýzy přímo navazovala na výběr vhodného modelu, ovšem fenotypová data od SCHP byla aktualizována, aby bylo možné zahrnout nově získané údaje z DNA čipů. Proto byla využita data z kontroly užitkovosti od ledna 2013 do prosince 2023. V tomto případě byla zahrnuta také otcovská plemena, která v souladu se současnou praxí byla hodnocena dohromady.

Oproti druhé části byly nejen populační genetické parametry, nýbrž také parametry validace modelů odhadnuty na základě optimalizovaných datových souborů (které v tomto případě budou chápány jako „kompletní“). Tato změna vyplynula z diskuze s oponenty článku v *Livestock Science* a s prof. Příbylem. Optimalizované soubory byly získány stejným způsobem, jaký je popsán v kapitole 4.2.2. **Error! Reference source not found.**, a zahrnovaly 72 436 pozorování pro české bílé u šlechtilé (CLW), 33 901 pozorování pro českou landrase (CLA) a 17 447 pozorování pro otcovská plemena (SIR). Poslední jmenovaný soubor zahrnoval plemena duroc (DU; 4 825 pozorování; 28 % ze SIR), bílé otcovské (LWS; 6 378; 37 %), pietrain (PE; 2 711; 16 %) a jejich křížence (3 533; 20 %). Dostupný rodokmen byl využit celý kvůli zahrnutí genetických skupin. Popisné statistiky hlavních vlastností ve využitých datových souborech jsou zachyceny v tabulce 4.2-C a rozdělení zvířat v tabulce 4.2-D.

Tabulka 4.2-C: Popisné statistiky podle plemenných skupin v kompletním a validačním souboru.

Vlastnost	České bílé ušlechtilé		Česká landrase		Otcovské linie		
	<i>kompletní</i>	<i>validace</i>	<i>kompletní</i>	<i>validace</i>	<i>kompletní</i>	<i>validace</i>	<i>validace PE</i>
	\bar{x} (s)	\bar{x} (s)	\bar{x} (s)	\bar{x} (s)	\bar{x} (s)	\bar{x} (s)	\bar{x} (s)
počet	72 436	4 267	33 901	3 151	17 447	1 473	324
věk [dny]	144,05 (5,93)	144,20 (6,20)	145,42 (5,61)	146,47 (5,30)	146,03 (5,06)	142,97 (2,04)	154,10 (2,92)
BW [kg]	100,05 (11,93)	102,97 (14,15)	105,18 (9,28)	107,43 (8,19)	106,18 (11,94)	107,23 (11,59)	121,69 (8,08)
ADG [g]	694,60 (77,99)	713,38 (86,46)	723,28 (57,26)	733,43 (48,93)	726,92 (75,90)	749,97 (80,65)	789,71 (51,03)
BF [mm]	6,75 (1,26)	6,18 (0,87)	6,70 (1,26)	5,94 (0,86)	6,39 (1,33)	6,42 (0,96)	4,84 (0,56)
MD [mm]	49,72 (5,76)	46,34 (5,17)	50,77 (5,30)	47,19 (4,33)	59,45 (6,29)	58,39 (6,18)	61,70 (4,00)

Zkratky: ADG = průměrný denní přírůstek; BW = živá hmotnost; BF = výška špeku; MD = výška svalu; PE = pietrain;
s = směrodatná odchylka; \bar{x} = průměr

Tabulka 4.2-D: Počty a struktura zvířat s vlastní užitkovostí, využitých ve validaci nebo se známým genotypem podle plemenných skupin.

Plemeno	Rodokmen	Fenotyp	Validace	Genotyp	$G \cap F$	$G \cap V$	G / F
CLW	81 046	72 436	4 267	1 053	876	52	1,5 %
CLA	37 323	33 901	3 151	397	288	35	1,2 %
SIR	22 064	17 447	1 473	360	234	30	2,1 %
DU		4 825	612	109	67	12	
LWS		6 378	69	82	55	2	
PE		2 711	289	169	112	6	
PE*		2711	324	169	112	14	
DU×LWS		1 389	131				
DU×PE		670	60				
LWS×PE		1 474	312				

Poznámky: kurzívou plemena a počty jako podskupiny SIR; PE* = počty zvířat pro PE v samostatné validaci

Zkratky: CLA = české landrase; CLW = české bílé ušlechtilé; DU = duroc; F = fenotyp; G = genotyp;

$G \cap F$ a $G \cap V$ = průnik množin G a F, respektive G a V; G / F = podíl genotypů vůči fenotypům;

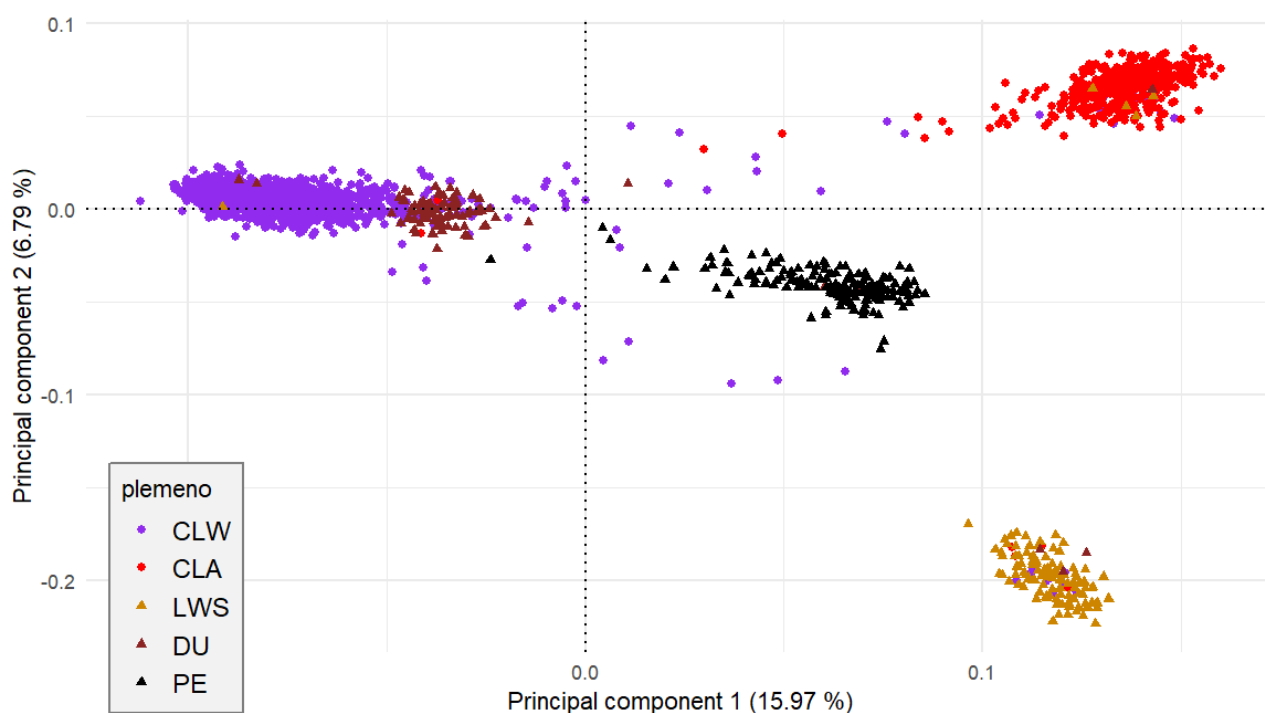
LWS = bílé otcovské; PE = pietrain; SIR = otcovská plemena; V = validace

4.2.3.1. Genotypová data

Genotypy byly dostupné pro 1 947 zvířat na třech různých DNA čípech: Illumina Porcine SNP60v2 (61 565 SNP), Illumina PorcineSNP60 BeadChip (59 319 SNP) a GeneSeek GGP Porcine 50K (50 697 SNP). Po vzoru Žákové et al. (2022) tak byly uvažovány polymorfizmy, které byly společné pro všechny tři čipy – celkem 32 896 SNP. Genotypová data byla v první řadě zhodnocena analýzou hlavních komponent (PCA; obrázek 4.2-B), která byla provedena v programu RStudio, verze 4.0 podle upraveného skriptu od Mészárose (2021). Pomocí PCA byl ověřen předpoklad shlukování jedinců podle plemen a zároveň byli odstraněni jedinci, jejichž plemenná příslušnost podle rodokmenu neodpovídala očekávané plemenné příslušnosti podle DNA čipů (v uvedeném grafu jsou to body s jinou barvou, než jakou má nejbližší shluk). Další konflikty v rodokmenu byly identifikovány pomocí SEEKPARENTF90 (Misztal et al. 2022).

Zbylé kroky kontroly kvality byly provedeny zvláště pro každou plemennou skupinu (CLW, CLA, SIR) pomocí programu PLINK, verze 1.90 (Purcell & Chang 2023). Takto byla odstraněna zvířata s podílem chybějících genotypů vyšším než 10 %, SNP s podílem chybějících vzorků vyšším než 10 % a SNP s frekvencí minoritní alely nižší než 5 %. Kontrolou kvality prošlo 1 053 vzorků (29 586 SNP) u českého bílého ušlechtilého, 397 vzorků (30 177 SNP) u české landrase a 360 vzorků (31 958 SNP) u otcovských plemen. Podrobnější rozdělení genotypovaných zvířat zachycuje tabulka 4.2-D.

Obrázek 4.2-B: Analýza hlavních komponent (PCA) pro 1947 dostupných vzorků.



Zkratky: CLA = české landrase; CLW = české bílé ušlechtilé; DU = duroc; LWS = bílé otcovské; PE = pietrain

4.2.4. Porovnání BLUP a ssGBLUP: vyhodnocení

Porovnání BLUP animal modelů probíhalo podle stejné metodiky jako v kapitole 4.2.2., ovšem vhodnost modelů podle AIC tentokrát srovnána nebyla kvůli odlišným počtům zvířat načtených z rodokmenu (Jang et al. 2022). Větší důraz byl kladen na normalitu rozdělení (plemenných hodnot, popř. spolehlivostí), neboť jde o jeden z předpokladů metody LR jakožto regresní analýzy. Proto byly pro validační kohortu v každém modelu vytvořeny kvantilové diagramy (Q-Q ploty) pomocí PROC UNIVARIATE v programu SAS/STAT®, verze 9.4.

Validační kohorty pro mateřská plemena sestávaly ze zvířat narozených v roce 2021–2023 se známým fenotypem, jejichž rodiče měli vlastní užitek v kompletním i parciálním souboru (tedy obdobně jako v předešlé analýze). V případě sdružené analýzy otcovských plemen jsou kvůli různorodosti a nevhodné struktuře skupiny prezentovány dva přístupy výběru zvířat do validační kohorty: 1) všechna mladá zvířata (2021–2023) se známým fenotypem pocházející z nejlépe propojeného chovu; 2) pouze zvířata plemene pietrain se stejnými podmínkami jako pro mateřská plemena. Dále byl testován například způsob, kdy byl efekt plemenné příslušnosti korigován nadvakrát, tedy nejprve pomocí animal modelu a poté odečtením průměrů nejmenších čtverců z GLM. Nicméně ani tento způsob normalitu dat pro validaci nezlepšil.

Design nejdůležitějších modelů pro tuto část práce je shrnut v tabulce 4.2-E. Kvůli využití optimalizovaných datových souborů také pro validaci a kvůli analýze otcovských plemen byl znovu porovnán efekt SRO jako pevný a náhodný. Zároveň bylo nutné podrobněji otestovat volitelná nastavení parametrických souborů, a to zejména s ohledem na genetické skupiny (skupiny neznámých rodičů, dále UPG), metodu řešení soustavy rovnic a sestavení matice \mathbf{H} pro ssGBLUP. Příklad parametrického souboru zachycuje obrázek 4.2-C.

Složky rozptylu v programu REMLF90 byly odhadnuty pro modely bez UPG. Jak ukázali například Wolf J & Wolfová M (2012), (ko)variance konvenčních animal modelů se po přidání UPG téměř nemění, a tudíž UPG nejsou nutné. Při kombinaci REML a ssGBLUP sami vývojáři programu doporučují UPG vynechat, jelikož v opačném případě dochází k nestabilitě výpočtů až k nemožnosti konvergence (D. Lourenco, osobní komunikace).

Pro předpovědi EBV pomocí BLUPF90 byly použity jak základní verze modelů, tak verze s náhodnými UPG (*OPTION random_upg*). Náhodné UPG jsou považovány za vhodnější než pevné UPG jak z teoretického hlediska, tak kvůli výpočetní stabilitě, zvláště při vysokém počtu skupin s malým počtem potomků (Bermann et al. 2021; Masuda et al. 2022). To byl i náš případ, neboť sdružený rodokmen obsahoval 175 různých UPG, přičemž 92 z nich zahrnovalo méně než 10 přímých

potomků. Při použití ssGBLUP byla matice \mathbf{H} rozšířena o matici \mathbf{Q}_2 (*OPTION exact_upg*), která přiřazuje genotypovaná zvířata příslušným skupinám (Misztal et al. 2013).

BLUPF90 dále nabízí různé algoritmy pro řešení MME. Výchozím algoritmem je iterativní metoda konjugovaných gradientů (PCG), která nahrazuje výpočetně náročnou přímou inverzi. Nicméně rozsah dat umožnil použití přesnější přímé metody založené na Choleského dekompozici, tedy balíček YAMS (zkratka pro Yet Another Mixed-Model Solver), jenž představuje vylepšení balíčku FSPAK (Misztal et al. 2022). Řešení pomocí YAMS se dokonce ukázalo být rychlejší než PCG a rovněž přináší možnost výpočtu individuálních přesností. Spearmanův koeficient pořadové korelace mezi EBV získanými pomocí PCG a YAMS činil okolo 0.99.

Vybraný genomický model by navíc analyzován s různým nastavením parametrů pro blending (*OPTION AlphaBeta a b*) a tuning (*OPTION tunedG x*), které jsou přiblíženy v kapitole 2.5.4. Škálování (*OPTION TauOmega t w*) využito nebylo, jelikož novější verze BLUPF90 automaticky zohledňují inbreeding při tvorbě \mathbf{A}^{-1} , a tudíž se tato možnost nedoporučuje (Misztal et al. 2022). Pro blending je výchozím nastavením *AlphaBeta 0.95 0.05*, kdy dochází k opravení původní genomické matice \mathbf{G}_0 s 95% váhou pomocí matice \mathbf{A}_{22} s 5% váhou. Dále byl váhový koeficient \mathbf{G}_0 stanoven na 80 %, 65 % a 50 %. V případě tuningu je výchozím nastavením *tunedG 2*, které upravuje průměrnou hodnotu (mimo)diagonálních prvků matice \mathbf{G} na průměr (mimo)diagonálních prvků matice \mathbf{A}_{22} . Možnost *tunedG 0* značí žádnou úpravu; *1* značí úpravu průměru diagonály \mathbf{G} na 1,00 a průměru mimodiagonály na 0,00; *3* značí úpravu průměru celé \mathbf{G} na průměr \mathbf{A}_{22} ; a *4* značí úpravu, kterou navrhli Vitezica et al. (2011) a která se podobá výchozí možnosti 2 (Masuda 2019).

Analyzovány byly také efekty jednotlivých SNP v programu PostGSF90, který pro tento účel využívá ssGWAS. Pro všechny tři soubory byl použit model G1 (tabulka 4.2-E) a pouze ty genotypy, pro něž byla v mapovém souboru dostupná pozice, tedy 20 790, 21 221 a 22 474 SNP pro české bílé ušlechtilé, českou landrase, respektive otcovská plemena. Hlavním předmětem této analýzy byla vysvětlená proměnlivost jednotlivými SNP, ovšem v souladu s metodikou GWAS byla posouzena také významnost SNP. Pro tento účel byla použita jednotná hodnota záporného dekadického logaritmu pro hladinu významnosti $\alpha = 0,05$ s Bonferroniho korekcí vzhledem k souboru s nejvíce SNP, tedy: $-\log_{10} \frac{0,05}{22474} = 5,65$.

Tabulka 4.2-E: Závěrečné porovnání tradičních a genomických BLUP animal modelů.

Model	Pevné efekty	Náhodné efekty
A1	plemeno + pohlaví + SRO + UZ	vrh + zvíře (A)
A2	plemeno + pohlaví + rok + stádo + UZ	SRO + vrh + zvíře (A)
G1	jako A1	vrh + zvíře (H)
G2	jako A2	SRO + vrh + zvíře (H)
_UPG	jako základní model	... + UPG

Poznámky: A1 a A2 = BLUP-AM (bez UPG); G1 a G2 = ssGBLUP-AM (bez UPG); efekt plemene pouze pro soubor otcovských plemen; _UPG = přidání náhodných UPG; vše tříznakové modely pro ADG, BF a MD

Zkratky: A = matice rodokmenové příbuznosti; ADG = průměrný denní přírůstek; BF = výška hřbetního tuku; H = kombinovaná matice rodokmenové a genomické příbuznosti; MD = výška svalu; SRO = stádo-rok-období; UPG = skupina neznámého rodiče (genetická skupina); UZ = ultrazvukový přístroj (pouze pro BF a MD)

Obrázek 4.2-C: Příklad parametrického souboru pro RENUMF90.

```

DATAFILE
opt_CLW_w.txt
TRAITS _
9 10 11  ## 9 = ADG; 10 = BF; 11 = MD
FIELDS_PASSED TO OUTPUT

WEIGHT(S)

RESIDUAL_VARIANCE
1609.    7.654    36.72
 7.654    0.6996    0.4442
36.72    0.4442    14.73
EFFECT
3 3 3 cross alpha ## FE - pohlaví
EFFECT
8 8 8 cross alpha ## FE - SRO
EFFECT
0 7 7 cross alpha ## FE - UZ přístroj
EFFECT
5 5 5 cross alpha ## RE - vrh
RANDOM
diagonal
(CO)VARIANCES  ## složky rozptylu RE
 626.7    0.000    0.000
  0.000    0.1156    0.000
  0.000    0.000    2.243
EFFECT
1 1 1 cross alpha ## RE - jedinec
RANDOM
animal
FILE          ## rodokmenový soubor
rod_ng.txt
FILE_POS      ## jedinec, otec/upg, matka/upg a dvě nuly
1 4 5 0 0
SNP_FILE
snp_CLW.txt
UPG_TYPE
in_pedigrees
(CO)VARIANCES  ## složky rozptylu RE
 419.2    3.512    19.06
  3.512    0.06184    0.2451
 19.06    0.2451    1.757
OPTION use_yams  ## alternativa solv_method PCG/FSPAK/SOR
OPTION sol se   ## kromě solutions uloží take střední chyby
OPTION random_upg
OPTION exact_upg
OPTION AlphaBeta 0.95 0.05  ## blending
OPTION tunedG 2  ## tuning

```

5. Výsledky a diskuze

5.1. Asociační analýza polymorfizmu v MC4R

5.1.1. Významnost modelu

V obou kategoriích podle krmné dávky (KD) se nacházely všechny tři genotypy podle p.Asp298Asn v MC4R. V *ad libitum* kategorii (ADL) byly zjištěny alelické frekvence 0,56 pro G (298Asp) a 0,44 pro A (298Asn), v restringované kategorii (RES) 0,62 pro G a 0,38 pro A. Díky tomu bylo možné otestovat významnosti hlavních efektů (Tabulka 5.1-A). Determinační koeficient modelu průměrně dosáhl 48 % s minimem u vlastnosti pH masa (16 %) a maximem u jatečné výtěžnosti (74 %). Zatímco doprovodné efekty – křížení/plemeno, pohlaví, délka výkrmu, jatečná hmotnost (CW) – byly významné u většiny vlastností ($p < 0,05$), v případě efektů genotypu MC4R, KD a interakce MC4R×KD bylo možné nulovou hypotézu zamítnout jen v případě některých ukazatelů výkrmnosti a kvantitativní stránky jatečné hodnoty. Chybějící nebo malou průkaznost vybraného polymorfizmu ve vztahu k produkčním ukazatelům zaznamenali také Park et al. (2002), Stachowiak et al. (2006), Lyadskiy et al. (2011), Muñoz et al. (2011) nebo Wang et al. (2019). Neprůkazné rozdíly jsou obvykle připisovány genetické různorodosti studovaných populací.

Efekt MC4R se v signifikantní míře ($p < 0,01$) projevil pouze u vlastnosti konverze krmiva (FCR) a u zbylých produkčních ukazatelů se pohyboval okolo hranice 0,10. Efekt KD byl vysoce významný pro průměrný denní přírůstek (ADG) a CW ($p < 0,001$). Zvířata ADL převyšovala RES také v tučnosti, což se projevilo na podílu libové svaloviny (LMP), výšce hřbetního tuku (BF), nebo podílu tukového krytí na hlavních masitých částech (LC_F). Z pozorovaných efektů MC4R a KD pak vyplynuly významné rozdíly v rámci jejich interakce v případě FCR ($p < 0,01$) a jatečné výtěžnosti (JV; $p < 0,001$). U vlastností ADG a CW se rozdíly blížily hladině významnosti 0,05. Ovšem velikost efektu interakce MC4R×KD se u těchto vlastností dá považovat za malou, neboť parciální čtvercová éta dosáhla maximálně 0,023 (Lakens 2013).

5.1.2. Významnost genotypu

Naše studie byla zřejmě první, která analyzovala efekt MC4R p.Asp298Asn ve dvou krmných kategoriích. Předšlé autorské kolektivy se zabývaly především zvířaty krmenými *ad libitum* (Meidtner et al. 2006; Stachowiak et al. 2006; Van Den Maagdenberg et al. 2007; Salajpal et al. 2009; Piórkowska et al. 2010; Muñoz et al. 2011; Fontanesi et al. 2013; Van Den Broeke et al. 2015), případně KD nespécifikovaly nebo nezahrnuly do analýzy (Kim et al. 2000a; Kwon et al. 2015; Getmantseva et al. 2017). Obě úrovně KD se objevily ve studii Dvořákové et al. (2011), ovšem pouze jako sdružený efekt s kombinací křížení. Houston et al. (2004) a Jokubka et al. (2006) pak studovali efekt p.Asp298Asn u prasat krmených restringovaně. Množství krmiva přitom výrazně ovlivňuje

klíčové produkční ukazatele (Lebret 2008), s nimiž je zároveň asociován gen *MC4R*. Proto věříme, že by oba efekty měly být zohledněny společně, jako tomu bylo v případě výzkumu lidského *MC4R*, který provedli Khodarahmi et al. (2020).

Tabulka 5.1-A: Souhrn významnosti prediktorů v lineárním modelu a síla efektu hlavního sledovaného prediktoru (interakce genotypu *MC4R* a krmné dávky).

Vlastnost	MC4R	KD	MC4R×KD	Plem.	Pohlaví	Výkrm	CW	R ²	η ²
ADG [g]	0,101	***	0,065	***	***	***		0,688	0,023
FCR [g]	**	ns	**	***	***	***		0,448	0,023
CW [kg]	0,073	***	0,069	***	ns	***		0,316	0,010
JV [%]	0,118	ns	***	***	***	0,056	***	0,739	0,008
LMP [%]	0,098	*	ns	***	***	***	***	0,347	0,000
BF [mm]	ns	0,101	ns	***	***	***	***	0,673	0,005
LC _R [%]	ns	ns	ns	***	***	***	*	0,479	0,009
LC _F [%]	ns	0,056	ns	***	***	***	***	0,499	0,000
LC _M [%]	ns	ns	ns	***	***	***	*	0,481	0,005
panenka [g]	ns	ns	ns	***	***	ns	*	0,450	0,006
pH _{MLT}	ns	ns	ns	*	ns	***		0,178	0,000
pH _{MS}	ns	ns	ns	0,059	ns	***		0,156	0,000
t _{MLT} [°C]	ns	**	ns	***	ns	***		0,496	0,004
t _{MS} [°C]	ns	ns	ns	***	*	***		0,542	0,001
DL [%]	ns	ns	ns	***	ns	ns		0,633	0,005

Záhlaví: KD = krmná dávka; MC4R = genotyp podle p.Asp298Asn; MC4R×KD = interakce; plem. = kombinace křížení či plemeno; R² = determinační koeficient GLM; výkrm = délka výkrmu; CW = jatečná hmotnost; η² = parciální čtvercová éta pro MC4R×KD;

Vlastnosti: ADG = průměrný denní přírůstek; BF = výška hřbetního tuku; CW = jatečná hmotnost; DL = ztráta odkapem 24 h *post mortem*; FCR = konverze krmiva; JV = jatečná výtěžnost; LC_F = podíl depozitního tuku v masitých částech; LC_M = podíl masa v masitých částech; LC_R = podíl masitých částí v CW; LMP = podíl libového masa; MLT = *m. longissimus thoracis*; MS = *m. semimembranosus*; pH = vodíkový exponent 45 min; *post mortem*; t = teplota 45 min *post mortem*

Významnosti (p-hodnoty): * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; ns = $p > 0.05$

Podrobnější analýza interakce MC4R×KD (tabulka 5.1-C) naznačila, že p.Asp298Asn by se mohl projevat jiným směrem nebo jinou intenzitou právě v návaznosti na podmínky krmení. V rámci kategorie ADL byl pozorován významný rozdíl ($p < 0,05$) v CW, kdy zvířata genotypu *GG* byla lehčí než *AA*, což korespondovalo s celkovým trendem v ADG ($GG < GA < AA$). Tato zjištění odpovídají závěrům, které publikovali předešní autoři (kapitola 2.5.2.). Zároveň byla alela *G* asociována s nižší JV a vyšším LMP, byť u těchto vlastností nebylo proložení seřazených průměrů genotypů lineární.

Také v kategorii RES byl genotyp *GG* asociován s vyšší zmasilostí (podle LMP a BF). Avšak patrný kontrast oproti kategorii ADL nastal v případě genotypu *AA*, jehož nositelé vykázali nejnižší

ADG v kombinaci s nejvyšším FCR, tedy i nejnižší CW. Ve všech třech případech se jednalo o průkazný rozdíl ($p < 0,05$, nebo dokonce $p < 0,001$) ve srovnání s heterozygoty GA.

Je vhodné podotknout, že uvedené p -hodnoty neprošly korekcí, jelikož naší snahou bylo především zasadit rozdíly do patřičného kontextu intervalů spolehlivosti či velikosti efektů, jak doporučují Perneger (1998) nebo Blume et al. (2019). Zřejmým omezením této studie jsou počty zvířat, zejména v kategorii RES, které se promítly do poměrně velkých středních chyb (tabulka 5.1-C a obrázek 5.1-A). Míra rozdílů mezi genotypy pro vlastnost CW je dále přiblížena v tabulce 5.1-B, která ukazuje standardizovanou a peněžitou velikost efektu. Podle Cohena d lze ze všech párových porovnání vyvodit slabý účinek interakce MC4R×KD (Lakens 2013) s výjimkou rozdílu mezi genotypy GA a AA v kategorii RES ($d = 0,62$). Ovšem jako více ilustrativní se jeví rozdíly v očekávané ceně JUT podle systému SEUROP (TIS SZIF 2024). V kategorii ADL by měly rozdíly mezi porážkovými hmotnostmi podle genotypu hodnotu 47 až 61 Kč za jatečný kus (1,1–2,5 % průměrné ceny) a v kategorii RES 116 až 236 Kč (2,7–5,5 %).

Tabulka 5.1-B: Vliv p .Asp298Asn na jatečnou hmotnost vyjádřený Cohenovým d (nad diagonálou) a cenovým rozdílem (pod diagonálou) mezi genotypy a krmnými kategoriemi.

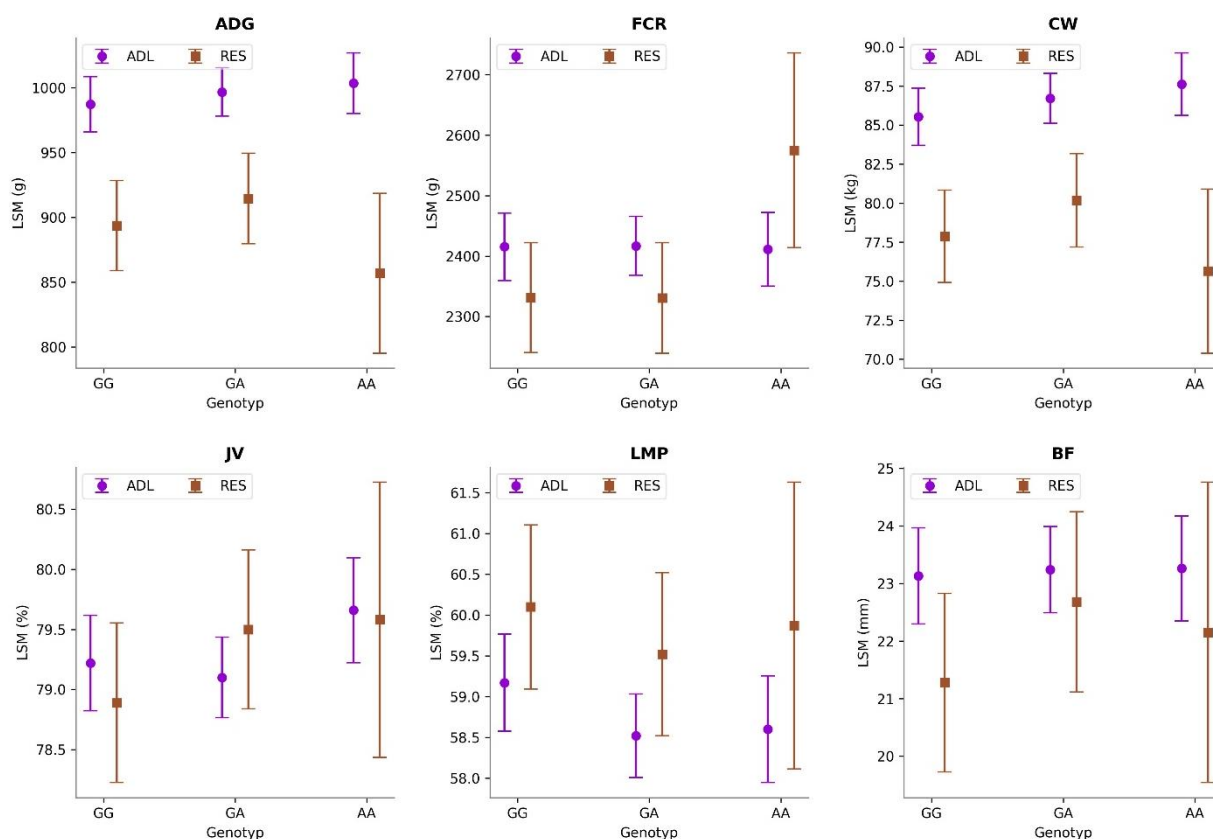
ADL	GG	GA	AA	RES	GG	GA	AA
GG		0,16	0,29	GG		0,31	0,32
GA	61 Kč		0,13	GA	120 Kč		0,62
AA	109 Kč	47 Kč		AA	116 Kč	236 Kč	

Pozorované rozdíly mohou být vysvětleny náhodnou chybou, ale také odlišným zapojením alel v závislosti na podmínkách pro růst jedince. Během většiny výkrmové fáze převažují v těle prasete bílkoviny nad tuky a krmná restrikce dominanci bílkovin prodlužuje, přičemž záchovná krmná dávka tvoří větší podíl z celkové krmné dávky (Whittemore & Kyriazakis 2007). Proto alela G související se zmasilostí možná hraje stěžejní roli a alela A související s vyšší tučností kvůli vyššímu příjmu krmiva doplňuje růst až po splnění nároků na záchovu, případně v závěrečné části výkrmu (po dosažení inflexního bodu růstové křivky). Do této interpretace zapadají výsledky Salajpala et al. (2009), kteří pozorovali signifikantní efekt p .Asp298Asn až ve vyšších porážkových hmotnostech. V podmínkách krmné restrikce by tedy zvířata genotypu AA byla oproti nositelům alely G znevýhodněna méně efektivním využitím záchovné dávky. Zato heterozygotní genotyp GA by mohl být během krmné restrikce ideální proto, že díky alele G dokáže efektivně využít záchovnou část krmné dávky pro růst a alela A poskytuje dodatečný růstový potenciál. Naopak pokud jsou zvířata krmena *ad libitum*, maximalizace efektivity konverze záchovné krmné dávky na růst svaloviny není tak podstatná, a tudíž genotyp AA předčí zbylé dva díky vyšší příjmu krmiva.

Studie Jokubky et al. (2006), která zahrnovala zvířata krmená restringovaně, však předchozí vysvětlení nepodporuje. Tito autoři pozorovali nejvyšší ADG spolu s nejvyšším LMP právě u genotypu AA a tyto výsledky připsali možnosti, že během krmné restrikce alela A využívá živiny více pro tvorbu libové svaloviny na úkor ukládání tuku. Ovšem je vhodné dodat, že Jokubka et al. (2006) studovali lokální plemeno, litevské bílé, což rovněž mohlo přispět k odlišnosti výsledků.

Co se týče parametrů kvality masa (pH, teplota a odkap), v naší práci nebyly pozorovány žádné statisticky významné rozdíly. Obdobně Van Den Broeke et al. (2015) nebo Van Den Maagdenberg et al. (2007) dospěli k závěru, že p.Asp298Asn má potenciální využití ve šlechtění na vlastnosti spojené s růstem bez ovlivnění kvalitativní stránky jatečné hodnoty. Nicméně Van Den Maagdenberg et al. (2007), Dvořáková et al. (2011) a Kwon et al. (2015) zároveň pozorovali významnou asociaci ($p < 0.05$) alely A s vyšším množstvím intramuskulárního tuku či vyšší křehkostí, což může korelovat s celkově vyšší tučností. Proto vliv p.Asp298Asn na některé ukazatele kvality masa nelze vyloučit.

Obrázek 5.1-A: 95% intervaly spolehlivosti odhadů genotypů MC4R pro hlavní vlastnosti.



Zkratky: ADG = průměrný denní přírůstek; ADL = *ad libitum* krmná dávka; BF = výška hřbetního tuku; CW = jatečná hmotnost; FCR = konverze krmiva; JV = jatečná výtěžnost; LMP = podíl libového masa; LSM = průměr podle metody nejmenších čtverců; RES = restringovaná krmná dávka

Tabulka 5.1-C: Rozdíly mezi genotypy MC4R v rámci kategorie krmné dávky.

Vlastnost	Krmná dávka <i>ad libitum</i> (ADL)						Krmná dávka restringovaná (RES)					
	GG (94)		GA (195)		AA (94)		GG (46)		GA (56)		AA (17)	
	LSM	SE	LSM	SE	LSM	SE	LSM	SE	LSM	SE	LSM	SE
ADG [g]	987,32	10,72	996,79	9,44	1003,6	11,78	893,62 xy	17,27	914,56 x	17,44	856,99 y	29,09
FCR [g]	2415,4	28,01	2416,9	24,66	2411,2	30,77	2331,2 X	45,13	2330,8 X	45,57	2575,2 Y	76,00
CW [kg]	85,53 x	0,92	86,71 xy	0,81	87,62 y	1,01	77,87 xy	1,47	80,18 x	1,49	75,65 y	2,48
JV [%]	79,22 xy	0,20	79,10 x	0,17	79,66 y	0,22	78,89	0,33	79,50	0,33	79,58	0,54
LMP [%]	59,17 x	0,30	58,52 y	0,26	58,60 xy	0,33	60,10	0,50	59,52	0,50	59,87	0,83
BF [mm]	23,13	0,42	23,24	0,38	23,26	0,46	21,28 x	0,77	22,68 y	0,78	22,15 xy	1,23
LC _R [%]	65,54	0,21	65,75	0,19	65,53	0,23	65,92	0,40	65,40	0,39	65,43	0,59
LC _F [%]	14,27	0,24	14,30	0,21	14,22	0,26	13,50	0,44	13,61	0,44	13,37	0,66
LC _M [%]	51,27	0,31	51,45	0,28	51,31	0,34	52,42	0,58	51,79	0,58	52,06	0,87
panenka [g]	535,59	12,20	553,52	10,73	549,10	13,33	533,52	22,69	517,19	22,55	515,71	33,84
pH _{MLT}	6,44	0,04	6,44	0,04	6,46	0,05	6,42	0,07	6,43	0,07	6,49	0,12
pH _{MS}	6,62	0,04	6,58	0,04	6,50	0,05	6,58	0,08	6,52	0,08	6,50	0,13
t _{MLT} [°C]	36,10	0,20	36,12	0,18	36,15	0,23	35,38	0,36	35,11	0,37	34,59	0,59
t _{MS} [°C]	36,58	0,26	36,77	0,24	36,61	0,30	36,52	0,47	36,49	0,48	36,22	0,77
DL [%]	5,82	0,35	5,70	0,31	5,86	0,39	6,68	0,64	6,07	0,65	5,47	0,97

Záhlaví: GG/GA/AA = genotypy podle p.Asp298Asn v MC4R (počet zvířat v závorce); LSM = průměr podle metody nejmenších čtverců; SE = střední chyba průměru; **vlastností** dle předchozí tabulky

Významnosti: x/y/xy (X/Y/XY) = průměry bez společného indexu v rámci daného řádku a kategorie KD se liší na hladině významnosti $p < 0.05$ (respektive $p < 0.001$)

Poznámka: LSM vycházejí z testování efektu interakce MC4R×KD, ovšem kvůli lepší přehlednosti *nejsou* uváděny statistické rozdíly napříč kategoriemi KD (např. ADL_GG vs. RES_AA).

5.2. Genetické hodnocení výšky tuku a svalu

5.2.1. Navržený BLUP-AM

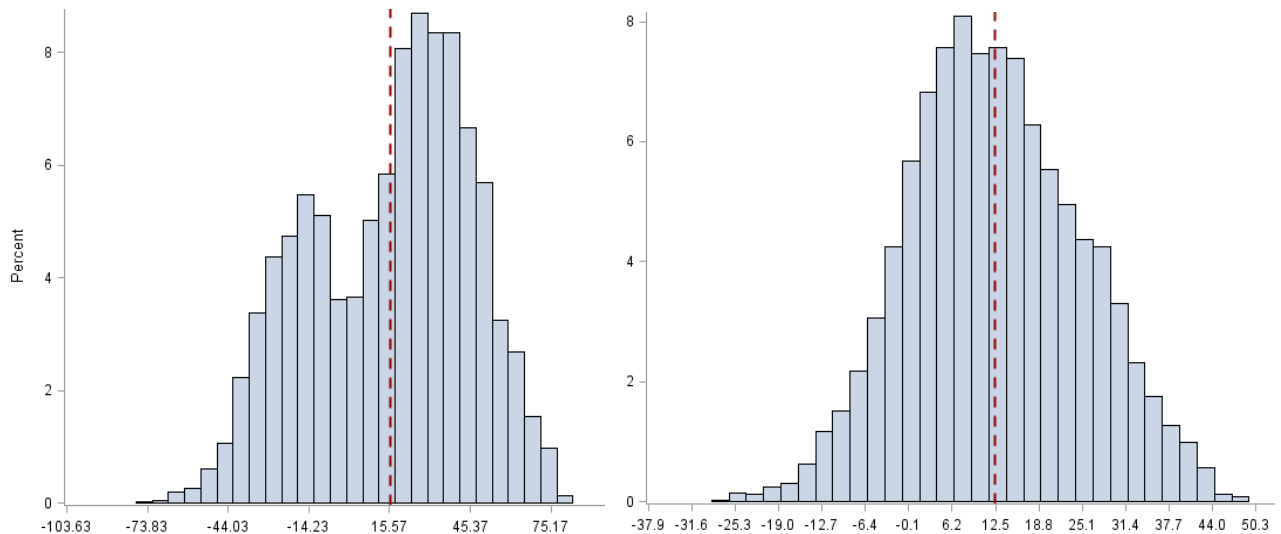
Jako významné prediktory pro průměrný denní přírůstek (ADG), hmotnost na konci testu (BW), výšku hřbetního tuku (BF) a výšku svalu (MD) byly vybrány: pohlaví, rok narození, ultrazvukový přístroj, věk na konci testu (dále jen věk), stádo-rok-období (SRO) a vrh. Tyto nebo podobné faktory se v dostupné literatuře objevují opakovaně (tabulka 2.4-B). Naopak počet narozených selat ani pořadí vrhu neměly na zkoumané vlastnosti průkazný vliv, ačkoliv někteří autoři (Hermesch et al. 2000; Salces et al. 2006; Khanal et al. 2019) je do svých modelů zahrnuli, a to zřejmě kvůli možné korelaci mezi porodní hmotností a růstem během výkrmu (Tang et al. 2008; Schinckel et al. 2010). Avšak Beaulieu et al. (2010) průkazný vliv velikosti vrhu ani porodní hmotnosti na BF ani MD nenašli. Takový výsledek může být připsán růstové kompenzaci (Whittemore & Kyriazakis 2007; Zotti et al. 2017) nebo zohlednění ID vrhu v modelové rovnici. Zároveň je potřeba podotknout, že čistokrevná selata procházejí předvýběrem.

Hlavní výsledky této kapitoly jsou zachyceny v tabulkách 5.2-A a 5.2-B. Analýza kolinearity mezi hmotností a věkem odhalila poměrně nízké parciální korelace (0,34 pro CLW a 0,28 pro CLA), které indikovaly, že by hmotnost mohla být jako doprovodná vlastnost vhodnější než přírůstek. Nicméně korelace mezi EBV ekvivalentních modelů (ADG–BW, BF–BF, MD–MD) byly blízké 1,00, s čímž souvisela rovněž i jejich srovnatelná výkonnost. Snowden & Fogarty (2009) za účelem výběru mezi základní a složenou vlastností (v našem případě BW vs. ADG) doporučili zohlednění vyšší koeficientu dědivosti, ale ty v naší analýze byly podobné (tabulka 5.2-C). Jako doprovodná vysvětlovaná proměnná tak byl preferován ADG, který je součástí nynějšího selekčního indexu a umožňuje použití jednoduššího modelu

Prediktor SRO byl uvažován jako pevný efekt (model A1) i jako náhodný efekt (A2). V populaci českého bílého ušlechtilého bylo náhodné SRO spojeno s lepší prediktivní výkonností, a to hlavně díky snížení průměrné absolutní disperze ($\overline{1 - b_{w,p}}$) z 0,23 (A1) na 0,11 (A2). Oproti tomu v populaci české landrase byly všechny pozorované metriky lepší v případě pevného SRO (A1). Jak bylo rozebráno v kapitole 2.4.2., efekt SRO je ze své podstaty náhodný (Schaeffer 2009). Ale zároveň takový přístup může vést k vychýleným a méně přesným předpovědím, pokud existují velké systematické rozdíly na vyšším stupni populační hierarchie (Frey et al. 1997; Babot et al. 2003; Wolf J & Wolfová M 2012; Škorput et al. 2018). Model A2 z tohoto důvodu obsahoval také samostatný pevný efekt roku narození a stáda. V případě varianty A1 by jejich zahrnutí bylo teoreticky správnější, ale prakticky se ukázalo být nadbytečné. Naopak ve variantě A2 při jejich absenci docházelo k umělému nadhodnocení LR metrik, protože metoda LR prakticky předpokládá očištění

environmentálních trendů (Macedo et al. 2020b). Zkreslení LR metrik vyplývá z porušení normality rozdělení EBV, které zachycuje obrázek 5.2-A. Nicméně ani pevné efekty stáda a roku v případě české landrase patrně nestačily na to, aby převážily výhody použití SRO jako náhodného efektu. Proto pro navazující komplexnější modely u české landrase bylo použito pevné SRO, zatímco u českého bílého ušlechtilého náhodné SRO.

Obrázek 5.2-A: Histogramy plemenných hodnot validační kohorty CLW. Vlevo model s náhodným efektem SRO bez pevných efektů stáda a roku, vpravo po přidání těchto efektů.



Dále byly analyzovány dvě interakce hlavních faktorů jako alternativa k vnořeným efektům. Tento přístup může vést ke zpřesnění výsledků, neboť lépe vystihuje hierarchické uspořádání chovu (Simsek & Firat 2011). Interakce vrh×SRO (modely označené _VS) byla přidána kvůli skutečnosti, že ve zkoumaných populacích 80 % sourozenců z jednoho vrhu zůstává pohromadě a spolu s dalšími vrhy tvoří nadřazenou skupinu (SRO). Vhodnost modelů (podle D_AIC) se tím u obou plemen významně zlepšila, avšak vliv na LR metriky a prediktivitu byl zanedbatelný. Interakce otec×stádo (modely označené _OS) pak měla za cíl zohlednit vysoký podíl kanců, kteří se vyskytli pouze v jednom stádu (73 % u CLW a 70 % u CLA). Podle Janga et al. (2022) tato interakce dokáže podchytit proměnlivost způsobenou nenáhodným pářením, tedy právě i nadměrným využíváním určitých otců v některých chovech. Nicméně v naší studii ani interakce otec×stádo nepřispěla ke zlepšení výkonnosti základního modelu – naopak způsobila vyšší vychýlení. Je tedy možné, že pevný efekt SRO (model A1 u CLA), případně pevný efekt stáda s náhodným SRO (A2 u CLW) vystihl nenáhodné asociace otců se stády dostatečně, jak naznačili Ugarte et al. (1992) nebo Visscher & Goddard (1993).

Zařazení efektu trvalého prostředí matky a maternálního genetického efektu (modely označené _MAT) vedlo k nejvýraznějšímu zvýšení vhodnosti modelů podle D_AIC (u CLW

–139 111, u CLA –64 615). K podobnému výsledku dospěli i Dong et al. (2020). V populaci českého bílého ušlechtilého navíc došlo ke snížení vychýlení a zvýšení populační přesnosti pro ADG a BF. Avšak vliv na prediktivní výkonnost pro MD byl negativní. V případě české landrase byl pozorován opačný trend: parametry výkonnosti modelu se pro ADG a BF zhoršily, zatímco pro MD se zlepšily. Tyto změny oproti základnímu modelu se rovněž promítly do pořadí zvířat podle výše EBV. Spearmanovy korelační koeficienty mezi EBV činily 0,97, 0,91 a 0,80 u českého bílého ušlechtilého (A2 vs. A2_MAT) a 0,95, 0,88 a 0,72 u české landrase (A1 vs. A1_MAT) pro ADG, BF, respektive MD.

Nekonzistentní vliv maternálních efektů na výkonnost modelu koresponduje s otázkou, jaká je jejich relevance ve vztahu k produkčním ukazatelům po odstavu. Některé předchozí studie vnímají maternální efekty jako důležité (Dube et al. 2014; Szyndler-Nędza et al. 2019; Dong et al. 2020), avšak v tabulce 2.4-B převažují jednodušší modely. Solanes et al. (2004) a Ogawa et al. (2021) v této souvislosti naznačují, že použití faktoru vrhu může do značné míry maternální efekty nahradit. Mimo to Gianola & Rosa (2015) či Paneru et al. (2022) upozorňují, že přílišná složitost modelu může vést ke zhoršení prediktivní výkonnosti, ačkoliv se vhodnost modelu zdá být lepší (problém overfittingu). Dále se také zvyšuje výpočetní čas (oproti základnímu modelu běžel REMLF90 zhruba 12krát déle a BLUPF90 zhruba čtyřikrát déle). Po zvážení těchto hledisek proto byly doporučeno použití základních modelů (A2 pro CLW a A1 pro CLA).

V modelu A2 pro české bílé ušlechtilé bylo dosaženo průměrného absolutního vychýlení $\overline{\mu_{p,w}} = 0,11$, průměrné absolutní disperze $\overline{1 - b_{w,p}} = 0,11$, průměrné inverzní korelace $\overline{\rho^{-1}} = 1,63$, průměrné populační přesnosti $\overline{acc_{LR}} = 0,36$ a průměrné prediktivity $\overline{PrA} = 0,13$. V modelu A1 pro českou landrase hodnoty činily: $\overline{\mu_{p,w}} = 0,19$, $\overline{1 - b_{w,p}} = 0,34$, $\overline{\rho^{-1}} = 2,18$, $\overline{acc_{LR}} = 0,31$, $\overline{PrA} = 0,08$. V tomto shrnutí vynikají především poměrně vysoké vychýlení a disperze u české landrase, které vyplynuly z nadhodnocených EBV v parciálním datovém souboru a promítly se do nízké prediktivity. Vysvětlením pro takové výsledky v porovnání s populací českého bílého ušlechtilého může být rozdíl ve velikosti vzorku, což souvisí s nižší spolehlivostí rodičovských průměrů. Přesto u obou studovaných plemen populační přesnosti zůstaly na vyhovující úrovni (i s ohledem na nízké dědivosti). Steyn et al. (2021) dosáhli populační přesnosti (podle LR metody) 0,36 pro ADG a 0,42 pro BFT, přičemž standardizovaná vychýlení a disperze se blížily 0,00. Hollifield et al. (2021b) odhadli přesnost pro jatečnou hmotnost ve výši 0,37 s disperzí 0,07. Podobné hodnoty lze pomocí koeficientu dědivosti dopočítat z výsledků autorů, kteří ve svých analýzách použili prediktivitu. Jedná se například o studie Lopez et al. (2019): 0,29 pro ADG, 0,32 pro BF a 0,38 pro MA; Tusell et al. (2016): 0,35 pro ADG a 0,23 pro LMP; a Poulsen et al. (2020): 0,34 pro ADG.

Tabulka 5.2-A: Parametry vhodnosti a prediktivní výkonnosti modelů pro české bílé ušlechtilé.

Model	Vlastnost	Validační kohorta $n = 7\,932$		D_{AIC}	$\mu_{p,w}$	$1 - b_{w,p}$	ρ^{-1}	acc_{LR}	PrA
		$EBV_w: \bar{x} (s)$	$EBV_p: \bar{x} (s)$						
A1	ADG	18,91 (11,40)	18,08 (7,38)	NA	-0,04	0,38	2,47	0,29	0,10
	BF	0,11 (0,11)	0,12 (0,08)		0,05	0,17	1,68	0,30	0,09
	MD	0,13 (0,64)	-0,02 (0,53)		-0,12	0,16	1,43	0,39	0,10
A2	ADG	18,81 (13,40)	17,17 (8,48)	0,0	-0,08	0,16	1,88	0,37	0,16
	BF	0,09 (0,14)	0,13 (0,10)		0,14	0,06	1,50	0,35	0,13
	MD	0,23 (0,66)	0,07 (0,49)		-0,13	0,11	1,50	0,36	0,10
A2 _VS	ADG	16,18 (12,47)	15,03 (7,73)	-1880,8	-0,06	0,11	1,82	0,37	0,16
	BF	0,09 (0,13)	0,12 (0,09)		0,14	0,04	1,48	0,35	0,13
	MD	0,25 (0,65)	0,08 (0,49)		-0,13	0,10	1,48	0,36	0,10
A2 _OS	ADG	18,82 (13,28)	16,00 (8,78)	-84,0	-0,13	0,11	1,69	0,40	0,17
	BF	0,09 (0,13)	0,13 (0,10)		0,16	0,14	1,54	0,33	0,11
	MD	0,19 (0,62)	0,01 (0,48)		-0,14	0,12	1,48	0,35	0,10
A2 _MAT	ADG	8,56 (11,98)	9,13 (7,69)	-139111,4	0,03	0,06	1,66	0,43	0,16
	BF	0,10 (0,12)	0,11 (0,09)		0,04	-0,01	1,32	0,38	0,13
	MD	0,30 (0,58)	0,16 (0,41)		-0,12	0,12	1,59	0,34	0,09
B1	BW	1,14 (1,66)	1,38 (1,07)	NA	0,08	0,41	2,63	0,28	0,09
	BF	0,04 (0,11)	0,05 (0,08)		0,04	0,16	1,74	0,29	0,09
	MD	0,16 (0,64)	0,05 (0,52)		-0,09	0,15	1,47	0,38	0,10
B2	BW	1,48 (1,97)	1,54 (1,21)	0,0	0,02	0,16	1,94	0,35	0,15
	BF	0,06 (0,14)	0,08 (0,10)		0,08	0,07	1,49	0,36	0,14
	MD	0,28 (0,68)	0,13 (0,49)		-0,11	0,09	1,51	0,36	0,10
B2 _VS	BW	1,26 (1,79)	1,32 (1,09)	-2054,8	0,02	0,12	1,87	0,36	0,15
	BF	0,06 (0,14)	0,08 (0,10)		0,07	0,04	1,48	0,36	0,13
	MD	0,26 (0,66)	0,11 (0,48)		-0,12	0,08	1,49	0,36	0,11
B2 _OS	BW	1,55 (1,93)	1,43 (1,26)	-89,1	-0,04	0,13	1,77	0,38	0,16
	BF	0,06 (0,13)	0,08 (0,10)		0,09	0,15	1,58	0,34	0,12
	MD	0,22 (0,62)	0,08 (0,46)		-0,12	0,10	1,50	0,35	0,10
B2 _MAT	BW	0,80 (1,06)	0,95 (0,73)	-139208,1	0,07	0,27	1,98	0,28	0,08
	BF	0,02 (0,12)	0,04 (0,09)		0,10	0,05	1,31	0,40	0,13
	MD	0,10 (0,60)	-0,05 (0,41)		-0,12	0,05	1,53	0,34	0,10

Zkratky: $1 - b_{w,p}$ = disperse EBV; acc_{LR} = populační přesnost podle metody LR;

ADG = průměrný denní přírůstek [g]; BF = výška hřbetního tuku [mm]; BW = živá hmotnost na konci testu [kg];
 D_{AIC} = diference v Akaikovu informačním kritériu od základního modelu (označeného jako 0,0); $EBV_{w/p}$ = plemenné
hodnoty v kompletním/parciálním souboru; MD = výška svalu v [mm]; n = počet zvířat; $\mu_{p,w}$ = standardizované
vychýlení EBV; NA = nedostupné (kvůli odlišným pevným efektům); PrA = prediktivita;
 ρ^{-1} = inverzní korelace EBV; s = směrodatná odchylka; \bar{x} = průměr

Tabulka 5.2-B: Parametry vhodnosti a prediktivní výkonnosti modelů pro českou landrase.

Model	Vlastnost	Validační kohorta <i>n</i> = 4 573		D_{AIC}	$\mu_{p,w}$	$1 - b_{w,p}$	ρ^{-1}	acc_{LR}	PrA
		$EBV_w: \bar{x} (s)$	$EBV_p: \bar{x} (s)$						
A1	ADG	14,36 (10,64)	14,92 (7,35)	0,0	0,03	0,36	2,26	0,31	0,11
	BF	0,11 (0,13)	0,17 (0,08)		0,28	0,22	2,03	0,31	0,09
	MD	0,94 (0,59)	1,25 (0,47)		0,27	0,44	2,25	0,31	0,05
A2	ADG	17,94 (11,06)	20,18 (7,39)	NA	0,11	0,42	2,59	0,29	0,09
	BF	0,06 (0,14)	0,13 (0,09)		0,27	0,53	3,46	0,24	0,05
	MD	0,85 (0,69)	1,21 (0,53)		0,30	0,51	2,70	0,30	0,05
A1 _VS	ADG	13,29 (10,20)	13,96 (7,04)	-443,5	0,03	0,38	2,32	0,29	0,10
	BF	0,07 (0,12)	0,15 (0,08)		0,36	0,22	2,05	0,31	0,09
	MD	0,78 (0,58)	1,26 (0,46)		0,42	0,45	2,28	0,31	0,05
A1 _OS	ADG	14,05 (10,59)	14,70 (7,40)	+0,3	0,03	0,35	2,21	0,31	0,11
	BF	0,08 (0,12)	0,18 (0,08)		0,43	0,21	1,99	0,30	0,08
	MD	0,79 (0,58)	1,25 (0,46)		0,39	0,44	2,27	0,30	0,05
A1 _MAT	ADG	15,04 (7,88)	13,16 (5,47)	-64615,0	-0,12	0,45	2,61	0,27	0,06
	BF	-0,12 (0,11)	-0,12 (0,07)		0,00	0,36	2,30	0,28	0,07
	MD	1,73 (0,47)	1,92 (0,51)		0,19	0,35	1,44	0,42	0,07
B1	BW	1,81 (1,64)	2,23 (1,18)	0,0	0,14	0,42	2,38	0,31	0,11
	BF	0,05 (0,13)	0,09 (0,09)		0,16	0,21	1,91	0,33	0,10
	MD	0,75 (0,63)	0,98 (0,48)		0,18	0,49	2,56	0,28	0,05
B2	BW	1,64 (1,67)	2,26 (1,14)	NA	0,21	0,46	2,73	0,29	0,09
	BF	0,05 (0,14)	0,10 (0,08)		0,23	0,56	3,73	0,23	0,04
	MD	0,71 (0,71)	1,07 (0,52)		0,28	0,58	3,23	0,27	0,03
B1 _VS	BW	1,46 (1,57)	2,13 (1,14)	-238,3	0,23	0,48	2,64	0,28	0,09
	BF	0,05 (0,13)	0,10 (0,08)		0,22	0,24	2,00	0,33	0,09
	MD	0,67 (0,62)	1,00 (0,47)		0,26	0,53	2,75	0,27	0,04
B1 _OS	BW	1,66 (1,63)	2,05 (1,16)	-3,0	0,13	0,40	2,33	0,31	0,11
	BF	0,05 (0,12)	0,09 (0,08)		0,14	0,19	1,86	0,33	0,10
	MD	0,73 (0,62)	0,92 (0,47)		0,15	0,50	2,65	0,27	0,05
B1 _MAT	BW	1,06 (0,93)	1,41 (0,75)	-64637,6	0,17	0,59	3,01	0,24	0,04
	BF	0,02 (0,11)	0,06 (0,07)		0,16	0,36	2,53	0,25	0,07
	MD	0,47 (0,44)	0,62 (0,46)		0,17	0,35	1,48	0,42	0,07

Zkratky: $1 - b_{w,p}$ = disperze EBV; acc_{LR} = populační přesnost podle metody LR;

ADG = průměrný denní přírůstek [g]; BF = výška hřbetního tuku [mm]; BW = živá hmotnost na konci testu [kg];
 D_{AIC} = diference v Akaikovu informačním kritériu od základního modelu (označeného jako 0,0); $EBV_{w/p}$ = plemenné
hodnoty v kompletním/parciálním souboru; MD = výška svalu v [mm]; *n* = počet zvířat; $\mu_{p,w}$ = standardizované
vychýlení EBV; NA = nedostupné (kvůli odlišným pevným efektům); PrA = prediktivita;
 ρ^{-1} = inverzní korelace EBV; *s* = směrodatná odchylka; \bar{x} = průměr

5.2.2. Populační genetické parametry

U obou plemen byly odhady přímého genetického rozptylu, reziduálního rozptylu, a tedy i koeficientů dědivosti konzistentní napříč analyzovanými modely s výjimkou variant s maternálními efekty (příloha 1). Maternální genetický rozptyl a kovariance mezi maternálním a přímým efektem totiž přetáhly část proměnlivosti od přímého efektu, aniž by snížily reziduální rozptyl (tabulka 5.2-C). Tento výsledek může vysvětlovat, proč se výkonnost modelu po přidání maternálních efektů změnila různými směry.

Odhadnuté dědivosti v navržených základních modelech byly u obou plemen podobné (tabulka 5.2-C): 0,23, 0,10 a 0,10 v populaci českého bílého ušlechtilého a 0,26, 0,10 a 0,09 v populaci české landrase pro ADG, BF, respektive MD. Odhady pro ADG zapadají do výsledků z literárního přehledu (tabulka 2.4-A), které se pohybovaly v rozmezí 0,10–0,80 s průměrem 0,34. Nicméně naše odhady pro BF a MD byly v tomto porovnání výrazně nižší, neboť ostatní autoři uvádějí dědivosti 0,27–0,61 s průměrem 0,44 pro BF a 0,20–0,55 s průměrem 0,32 pro MD.

Nesoulad těchto výsledků může být částečně připsán chybám měření, protože zejména BF nemůže být měřen se stejnou přesností jako ADG. Tato skutečnost je patrná z fenotypového variačního koeficientu, který byl nejvyšší právě v případě BF (18,4 % u CLW a 18,6 % u CLA, na základě optimalizovaných dat), následně pro MD (11,4 % a 10,2 %) a ADG (11,1 % a 7,8 %). Rozdíly oproti ostatním studiím mohly vzniknout také menší uniformitou podmínek prostředí. Avšak Csató et al. (2002), Lukač (2016) či Lopez et al. (2018) rovněž analyzovali data z polního testu se srovnatelným počtem pozorování, a přesto dosáhli vyšších odhadů dědivosti pro BF i MD. Další možné vysvětlení je, že v českých populacích na tyto vlastnosti působil nepřímý selekční tlak skrze LMP a zároveň dochází ke ztrátě genetické diverzity, jak ukázali Krupa et al. (2014). Ztrátu proměnlivosti potvrdilo i dodatečné použití vybraného modelu ve starším datovém souboru (2004–2012), kdy byly pro BF a MD odhadnuty koeficienty dědivosti přibližně 2,4krát vyšší než v nových datech. Nízké koeficienty dědivosti by proto měly být zohledněny při konstrukci selekčního indexu.

Odhady koeficientů genetické korelace dosáhly oproti dědivostem velmi vysokých hodnot (tabulka 5.2-C): 0,61, 0,61 a 0,66 v populaci českého bílého ušlechtilého a 0,72, 0,92 a 0,70 v populaci české landrase pro dvojice vlastností ADG–BF, ADG–MD, respektive BF–MD. V tabulce 2.4-A se korelace mezi ADG (nebo hmotností) a BF pohybovala v rozmezí od –0,31 do +0,63 a mezi ADG a MD od –0,40 do 0,77. Je dobré připomenout, že BF ani MD v naší analýze nebyly očištěny o vliv hmotnosti čili středně silné až silné kladné korelace s ADG jsou v takovém případě očekávatelné. Ovšem podobně vysoké korelace v populaci o srovnatelné velikosti publikovali pouze Ros-Freixedes et al. (2013). Navíc korelace mezi BF a MD obvykle bývá záporná (průměrně –0,23).

V tomto ohledu se naše odhady blížily pouze studii Csató et al. (2002). Lze tak konstatovat, že mezi lokálními subpopulacemi stejného plemene existuje značná proměnlivost pleiotropních efektů. V případě českých mateřských linií mohou být vysoké a zároveň kladné genetické korelace výhodné, jelikož zájmem šlechtitelů je zastavit pokles BF a zvýšit MD.

V modelech s maternálními efekty byly odhady přímých genetických korelací nižší, a tedy více konvenční. Také odhady maternální dědivosti (0,01–0,03) zapadají do předchozích výsledků (Solanes et al. 2004; Dong et al. 2020). Nicméně korelace mezi přímým a maternálním efektem pro ADG byly pochybně vysoké (0,83 u CLW a 0,78 u CLA). Také z toho důvodu byl doporučen model bez maternálních efektů.

Tabulka 5.2-C: Složky rozptylu spolu s koeficienty dědivosti (h^2 ; v závorce na diagonále přímých genetických efektů) a genetické korelace mezi vlastnostmi (r_A ; v závorce mimo diagonálu) v doporučených modelech u obou plemen. Pro porovnání přidány modely s maternálními efekty (v závorce ve sloupci matka maternální dědivost = m^2 ; v závorce ve sloupci AM korelace mezi přímým a maternálním efektem = r_{AM})

Plemeno (model)	VL	Přímé genetické efekty			Další náhodné efekty					
		ADG (h^2 či r_A)	BF (h^2 či r_A)	MD (h^2 či r_A)	SRO	vrh	EPM	matka	AM	R
CLW (A2)	ADG	463,50 (0,23)	3,56	17,05	397,20	595,50				1551,00
	BF	(0,61)	0,07 (0,10)	0,23	0,25	0,12				0,70
	MD	(0,61)	(0,66)	1,69 (0,10)	2,60	2,13				14,68
CLW (A2_MAT)	ADG	311,30 (0,15)	1,41	4,74	391,80	566,80	0,72	27,62 (0,01)	76,57 (0,83)	1622,00
	BF	(0,34)	0,06 (0,07)	0,10	0,24	0,11	0,01	0,01 (0,02)	0,00 (-0,14)	0,71
	MD	(0,23)	(0,38)	1,32 (0,08)	2,51	2,00	0,08	0,55 (0,03)	-0,29 (-0,34)	14,84
CLA (A1)	ADG	380,70 (0,26)	3,17	20,86		350,10				1097,00
	BF	(0,71)	0,05 (0,10)	0,19		0,06				0,47
	MD	(0,92)	(0,70)	1,35 (0,09)		1,37				12,95
CLA (A1_MAT)	ADG	243,60 (0,16)	1,40	8,41		336,10	0,73	22,04 (0,01)	57,08 (0,78)	1161,00
	BF	(0,41)	0,05 (0,09)	0,09		0,05	0,01	0,01 (0,02)	-0,01 (-0,27)	0,47
	MD	(0,54)	(0,42)	0,98 (0,07)		1,24	0,08	0,49 (0,03)	-0,19 (-0,27)	13,10

Zkratky: ADG = průměrný denní přírůstek; AM = kovariance mezi přímým a maternálním efektem; BF = výška hřbetního tuku; CLA = česká landrase; CLW = české bílé ušlechtilé; EPM = rozptyl trvalého prostředí matky; MD = výška svalu; R = reziduální rozptyl; SRO = rozptyl stáda-roku-období

5.2.3. Porovnání BLUP a ssGBLUP: mateřská plemena

5.2.3.1. Revize základního modelu

Jak vyplynulo z předchozích výsledků, statistické pojetí faktorů jako stádo, rok a období hraje ve validaci významnou roli. Vzhledem ke skutečnosti, že validace v této části práce vycházela z jiných souborů (optimalizovaných) než v předešlé části, byly znovu ověřeny modely s pevným a náhodným efektem SRO (A1, respektive A2). Ani u jednoho z plemen kvantilové grafy nenasvědčovaly tomu, že by některý model měl být preferován kvůli větší normalitě plemenných hodnot (příloha 2). Nicméně rozdíly se vyskytly ve validačních metrikách.

V populaci českého bílého ušlechtilého (tabulka 5.2-F) tentokrát byla u modelu A1 (pevné SRO) zaznamenána nižší průměrná absolutní disperze (A1: $\overline{1 - b_{w,p}} = 0,08$; A2: $\overline{1 - b_{w,p}} = 0,20$), ale také nižší populační přesnost (A1: $\overline{acc_{LR}} = 0,28$; A2: $\overline{acc_{LR}} = 0,33$). Z výsledků je tedy patrné, že $cov(EBV_w, EBV_p)$ byla výrazně vyšší u modelu A2 (náhodné SRO), nicméně za cenu relativně vysoké $var(EBV_p)$, a to zejména v případě BF a MD. Jinými slovy, použití náhodného SRO vede k nadhodnocení rozptylu EBV parciálních dat či k podhodnocení rozptylu EBV kompletních dat. Z tohoto důvodu v metodě LR hodnocení disperze předchází populační přesnosti (Legarra & Reverter 2018) a v populaci českého bílého ušlechtilého se pevné SRO jeví jako jistější volba. V populaci české landrase (tabulka 5.2-G) byly rozdíly mezi modely A1 a A2 zanedbatelné, a proto byl rovněž preferován jednodušší model A1.

Nesoulad s předcházejícími výsledky s největší pravděpodobností vznikl využitím menšího počtu jedinců, zato ve větších a lépe propojených skupinách SRO. Zaprvé se tedy změnila struktura validační kohorty, která přímo ovlivňuje prediktivní výkonnost, a zadruhé se snížila náchylnost odhadů pevných efektů vůči extrémním hodnotám (zejména pro SRO). V takovém případě může být pevné SRO robustnější, neboť i přes teoretické nedostatky (Schaeffer 2009) dokáže spolehlivě oddělit časově stabilní trendy (pakliže je to potřeba). To se týká nejen trendu v hlavních faktorech jako stádo a rok, nýbrž také třeba interakce stádo×rok, která by v tomto případě rovněž mohla být podstatná. Zatímco model s pevným SRO absence prediktorů vyššího řádu nezatěžuje takovou chybou, v případě modelu s náhodným SRO už může docházet k podcenění systematických rozdílů, jak bylo demonstrováno v kapitole 5.2.1.

Přidání náhodných genetických skupin (UPG) sloužilo spíše pro potvrzení, že nezpůsobí nestabilitu předpovědí. Genetické skupiny jsou totiž využívány v současné praxi (Krupa et al. 2019a) a zároveň mohou být důležité pro ssGBLUP, kde ovlivňují vychýlení (Misztal et al. 2013; Masuda et al. 2022). V populaci českého bílého ušlechtilého zůstaly validační metriky základních modelů (A1 vs. A1_UPG) téměř beze změny. Takový výsledek lze přičíst genetické vzdálenosti mezi

předpověďmi efektů UPG a sledovanými plemennými hodnotami, neboť z naší definice validační kohorty vyplývá, že rodiče sledovaných jedinců nemohou být UPG. Žádní přímí potomci UPG se nevyskytují ani v celkových užitkových datech. Z tohoto pohledu jsou pak nápadnější rozdíly v populaci české landrase, kde přidání UPG přineslo značené zlepšení prediktivní výkonnosti. Konkrétně došlo ke snížení průměrného standardizovaného vychýlení (A1: $\overline{\mu_{p,w}} = 0,16$; A1_UPG: $\overline{\mu_{p,w}} = 0,08$), snížení disperze (A1: $\overline{1 - b_{w,p}} = 0,48$ vs. A1_UPG: $\overline{1 - b_{w,p}} = 0,38$), snížení inverzní korelace (A1: $\overline{\rho^{-1}} = 2,21$ vs. A1_UPG: $\overline{\rho^{-1}} = 1,93$) a zvýšení populační přesnosti (A1: $\overline{acc_{LR}} = 0,30$ vs. A1_UPG: $\overline{acc_{LR}} = 0,32$) i prediktivity (A1: $\overline{Pr\bar{A}} = 0,09$ vs. A1_UPG: $\overline{Pr\bar{A}} = 0,11$). V naší analýze nebylo zřejmé, proč UPG v případě české landrase měly o tolik větší dopad než u českého bílého ušlechtilého. Hodnoty předpovědí UPG se mezi plemeny příliš nelišily (např. pro ADG činila předpověď průměrně 2,35 g s odchylkou 3,46 g u CLW a průměrně 2,24 g s odchylkou 3,14 g u CLA). Rozdíly nejsou patrné ani ze zastoupení potomků UPG v rodokmenu (rodokmen CLW jich po roce 2013 zahrnoval 70, zatímco rodokmen CLA 44). Nicméně roli může hrát skutečnost, že do chovů českého bílého ušlechtilého jsou importovány inseminační dávky pouze z Francie, zatímco do chovů české landrase také z Polska. Propagace efektů UPG napříč maticí příbuznosti by proto vyžadovala hlubší zkoumání propojenosti nebo vlivů na genetické trendy (Theron et al. 2002).

Použití náhodných UPG lze doporučit u obou plemen s upozorněním, že dochází k umělému snížení individuálních přesností, jak je ukázáno na průměrných hodnotách r (tabulky 5.2-F a 5.2-G). Příčinu tohoto jevu popsali Bermann et al. (2023) a navrhli řešení v podobě zohlednění kontrastů s nejstaršími zakladateli. Nicméně tento krok v některých případech včetně našeho není nutný, neboť se jedná o systematické snížení. Pořadí zvířat podle individuálních přesností v základním modelu proto bylo téměř stejné jako pořadí v modelu s UPG (Spearmanovy korelace přes 0,99). V praxi by tedy stačilo vzít v úvahu posunutou škálu přesností, anebo k plemenným hodnotám z modelu s UPG připojit přesnosti ze základního modelu.

5.2.3.2. Genomický model

Trend napříč genomickými modely korespondoval s konvenčními modely, tedy bylo znovu možné podpořit využití modelů s pevným SRO a náhodnými UPG (G1_UPG) u obou mateřských plemen. Ačkoliv Lourenco et al. (2020) upozorňují, že využití UPG v ssGBLUP není stoprocentně vyřešené z hlediska teorie ani výpočetní aplikace, v našem případě UPG nezpůsobily neočekávané chování modelu. Změna parametrů pro blending a tuning zato nepřinesla žádné zlepšení (příloha 2). Dále proto bude srovnána prediktivní výkonnost ekvivalentních modelů A1_UPG s G1_UPG a individuální přesnosti budou z výše uvedeného důvodu srovnány mezi A1 a G1 (tabulky 5.2-F a 5.2-G).

Přidání genotypů mělo na prediktivní výkonnost modelu pro české bílé ušlechtilé nejednoznačný vliv. Inverzní korelace $\overline{\rho^{-1}}$ (−9 % při přechodu z A1_UPG na G1_UPG) a populační přesnost $\overline{acc_{LR}}$ (+3 %) sice poukazují na mírné zlepšení, nicméně zároveň se zvýšila disperze $1 - \overline{b_{w,p}}$ (+3 %), což je u genomických předpovědí častý jev (Legarra & Reverter 2018). Očekávaným směrem se změnily také individuální přesnosti r , které se pro celou validační kohortu sice zvýšily jen nepatrně (průměrně +1 % pro kompletní data a +2 % pro parciální data při přechodu z A1 na G1), ovšem v podskupině genotypovaných zvířat byl efekt zřejmý zejména na parciálních datech čili při absenci vlastní užitkovosti (+11 %). V populaci české landarse zahrnutí genomické matice výrazně snížilo průměrnou disperzi $1 - \overline{b_{w,p}}$ (−0,12). Snížila se také inverzní korelace $\overline{\rho^{-1}}$ (−0,22), zato průměrná populační přesnost $\overline{acc_{LR}}$ zůstala beze změny. Rozdíl v individuálních přesnostech r byl znovu patrný pouze na parciálních datech u genotypovaných zvířat (+6 %).

Zřejmou limitací naší studie je nízký celkový počet genotypovaných zvířat a rovněž jejich malé zastoupení ve validační kohortě (tabulka 4.2-D). Li et al. (2019) na základě simulovaných dat doporučili genotypování 40–50 % selat z vrhu čistokrevných prasnic ve čtyřech generacích, jelikož tak dochází k maximalizaci predikční přesnosti při současném zachování hospodářského zisku. Hollifield et al. (2021a) na reálných datech z komerční populace prasat ukázali, že trénováním modelu na třech rodičovských generacích s podílem genotypovaných jedinců 16 % vůči rodokmenu (19 702/125 155) a 25 % vůči fenotypům (19 702/77 945) bylo v následné generaci (validační kohortě) docíleno populační přesnosti téměř 0,80. K těmto studiím se tedy české populace prasat zatím zdaleka nepřibližují.

Nicméně z dostupné literatury není zcela zřejmé, zda existuje nějaký minimální počet či podíl genotypovaných jedinců, který by vymezoval možnost využití ssGBLUP. Massender et al. (2022) pomocí LR metody analyzovali výkonnost modelů u dvou plemen koz (833 a 874 genotypů) a dospěli k podobným závěrům jako my, totiž že konstrukce matice \mathbf{H} ani při takto malém počtu jedinců se známými genotypy neohrožuje stabilitu předpovědí a pozorované rozdíly oproti tradiční předpovědi jsou spíše pozitivní (např. individuální přesnosti mladých genotypovaných zvířat). Mehrban et al. (2019) v populaci masného skotu pozorovali nárůst predikční přesnosti o 8–14 % pro vlastnosti CW, BF a MA při zahrnutí 1 151 genotypů. Mancisidor et al. (2021) podpořili aplikaci ssGBLUP ve šlechtitelském programu lamy alpaka dokonce na základě 431 genotypů.

Tabulka 5.2-E dále ukazuje studie, které porovnávaly BLUP a ssGBLUP na produkčních ukazatelích u prasat. Žádná z nich sice neaplikovala přímo metodu LR, ale použité korelační či regresní metriky rovněž vyjadřují prediktivní výkonnost. Co se týče počtu zvířat s užitkovostí a genotypem, nejvíce se naší analýze přibližují Lopez et al. (2019) a Melnikova et al. (2021). V obou

těchto studiích došlo ke snížení disperze, přičemž ve druhé jmenované zároveň došlo k výraznému zvýšení (průměrně o 19 %) korelačního koeficientu mezi EBV na kompletních a parciálních datech. Nabízí se tedy srovnání s populací české landarse, kde se tato korelace průměrně zvýšila o 6 % po přidání genotypů, a to rovněž v souvislosti se snížením disperze. Melnikova et al. (2021) ovšem využili velmi malou validační skupinu s vyšší průměrnou spolehlivostí a možná také s větší propojeností v matici genomické příbuznosti, což může jejich výsledky nadhodnocovat. Z tohoto pohledu naší analýze více odpovídá studie Lopeze et al. (2019), kteří pro validaci vybrali všechna mladá zvířata, takže rozdíly ve výkonnosti modelů byly mnohem více konzervativní. Hong et al. (2019) a Zhou et al. (2019) ve svých analýzách využili srovnatelný počet genotypů, ovšem v menších populacích. Zatímco Hong et al. (2019) aplikovali dopřednou validaci na mladých zvířatech a dosáhli tak pouze mírného zlepšení predikční přesnosti (2–5 %), Zhou et al. (2019) zvolili pro validaci podskupinu genotypovaných zvířat a pozorovali výraznější rozdíly. Nicméně tyto rozdíly zahrnovaly značné změny v disperzi negativním směrem.

Salek Ardestani et al. (2021) dosáhli průměrného zlepšení v přesnosti modelu o 15 % (napříč třemi populacemi a třemi produkčními ukazateli) při téže proporci jedinců s genotypem vůči jedincům s fenotypem (v absolutních číslech 5 008, 5 362 a 6 649 genotypů pro LW, LA, respektive DU). Podobně jako v předchozí zmíněné studii, Zhou et al. (2019), bylo zlepšení přesnosti doprovázeno výraznou disperzí při použití ssGBLUP (až $-1,36$ u DU). Avšak tento problém byl odstraněn při použití metody BayesC π , která připisuje genetický rozptyl menšímu počtu lokusů. Míra převahy v predikční přesnosti nad BLUP přitom byla zachována. Na druhou stranu se významně zvýšil výpočetní čas (z průměrných 19 minut pro ssGBLUP na téměř 16 hodin pro BayesC π). Obdobné zvýšení predikční přesnosti při použití ssGBLUP popsali také Jibrila et al. (2022) a Poulsen et al. (2020), ačkoliv pracovali s mnohem hustěji genotypovanými populacemi. Při těchto srovnáních je vhodné upozornit, že zásadní roli hraje definice validační kohorty, její velikost nebo výběr vysvětlované proměnné (Salek Ardestani et al. 2021), což bylo potvrzeno i v našich výsledcích. V každém případě dostupné publikace naznačují, že ve šlechtitelské praxi prasat je potenciální nárůst predikční (populační) přesnosti díky aplikaci ssGBLUP z velké části naplněn už při genotypování 15 % populace. V případě českého bílého ušlechtilého by proto bylo potřeba použít DNA čip zhruba na 900 zvířat ročně a v případě české landrase na 400 zvířat ročně, což by znamenalo až čtyřnásobné navýšení současného tempa.

Avšak míra genotypování se odvíjí od také toho, jakou strategii budou chovatelé považovat za rentabilní. Jestliže bude hlavním cílem navýšení spolehlivosti u mladých kandidátů pro selekci, a nikoliv maximalizace populační přesnosti, značného efektu je dosaženo už při genotypování relativně malého počtu prověřených a vhodně zvolených zvířat pro tvorbu referenční populace. Bauer

et al. (2015) toto demonstrovali na datech z první laktace holštýnského skotu ($h^2 = 0,23$), když porovnali individuální přesnost u mladých býků na základě pouze rodičovského průměru (BLUP) oproti kombinaci rodičovského průměru s informací o genotypu (ssGBLUP). S 217 prověřenými býky v referenční populaci vzrostla průměrná individuální přesnost u mladých býků jen o 2 % a po zahrnutí 1791 prověřených býků tento nárůst činil už 18 %. V této souvislosti zdůraznili důležitost referenční populace rovněž Melnikova et al. (2021), kteří se skupinou čítající 1 194 jedinců dosáhli okolo 13% zvýšení individuální přesnosti u prasnic a 9% zvýšení u kanců z validační kohorty. Těmto výsledkům v naší studii odpovídá srovnání podskupiny genotypovaných zvířat v parciálních datech ($r[\overline{GEBV_p}] - r[\overline{EBV_p}] = +11 \%$ pro CLW a +6 % pro CLA), z čehož lze usuzovat, že genotypovací schéma u českých mateřských plemen má odpovídající podobu a při zařazení dalších jedinců do referenční populace se dá se očekávat zvýšení pozitivního efektu.

V českých populacích byl pro ADG, BF a MD navíc odhadnut poměrně nízký koeficient dědivosti, což může dále zvýraznit přínos ssGBLUP (Guarini et al. 2018; Melnikova et al. 2021). Ovšem odhady populačních genetických parametrů se při využití matice genomické příbuznosti zásadně nezměnily (tabulka 5.2-D). Největší rozdíl byl pozorován u české landrase pro ADG a MD, kde byla odhadnuta o 10 % nižší genetická proměnlivost než s tradičním modelem (beze změny v reziduální proměnlivosti). Tento pokles může značit, že na základě SNP byla zvýšena příbuznost některých rodin nebo zvířat importovaných ze zahraničí, což zároveň může souviset se zmíněným snížením disperze. Obecně by se odhady složek rozptylu měly mezi metodami shodovat, ovšem v praxi dochází k odchýlkám v závislosti na statistickém modelu, velikosti populace nebo strategii selekce a genotypování (Bermann et al. 2022).

5.2.4. Porovnání BLUP a ssGBLUP: otcovská plemena

Výsledkům ze sdružené analýzy otcovských plemen je věnována samostatná podkapitola, protože se nepodařilo najít způsob výběru validační kohorty, který dostatečně vyhovoval nárokům na normalitu rozdělení EBV. Genetické rozdíly mezi plemeny či liniemi, a to i v rámci jednoho chovu, zůstaly příliš velké navzdory použití animal modelu s pevným efektem plemene nebo SRO (popř. pevného stáda, když SRO bylo náhodné). Proto jsou předloženy výsledky ze dvou variant validační kohorty, které se jeví jako nejvhodnější: 1) všechna mladá zvířata z největšího chovu a 2) pouze jedinci plemene pietrain se stejnými podmínkami jako u mateřských plemen.

První varianta (tabulka 5.2-H) měla výhodu ve vyšším počtu zvířat a vyšší normalitě EBV pro ADG a MD. Na druhou stranu rozdělení EBV pro BF poukazovalo na existenci dvou subpopulací, pročež se populační přesnost dané vlastnosti stala zkrácenou. Výrazný vliv u všech tří vlastností mělo přidání UPG, díky nimž byly zlepšeny zejména tradiční modely (A1 a A2). Konkrétně došlo ke

snížení průměrné absolutní disperze (A1: $\overline{1 - b_{w,p}} = 0,25$ vs. A1_UPG: $\overline{1 - b_{w,p}} = 0,04$; průměry platí pouze pro ADG a MD) při zachování poměrně nízkého absolutního vychýlení (A1: $\overline{\mu_{p,w}} = 0,06$ vs. A1_UPG: $\overline{\mu_{p,w}} = 0,12$), a tím byla snížena inverzní korelace (A1: $\overline{\rho^{-1}} = 2,24$ vs. A1_UPG: $\overline{\rho^{-1}} = 1,64$), zvýšena populační přesnost (A1: $\overline{acc_{LR}} = 0,35$ vs. A1_UPG: $\overline{acc_{LR}} = 0,46$) i prediktivita (A1: $\overline{PrA} = 0,14$ vs. A1_UPG: $\overline{PrA} = 0,18$). Jinými slovy, díky UPG vzrostla $cov(EBV_w, EBV_p)$ podobě jako v případě české landrase. Pozorovaný dopad zahrnutí UPG do tradičních modelů dává oproti české landrase větší smysl alespoň z toho pohledu, že výskyt UPG v rodokmenu byl značně vyšší (SIR: 157 UPG po roce 2013 vs. CLA: 44). Ovšem populační přesnost (0,50) a prediktivita (0,28) pro ADG dosáhla diskutabilní výše.

Naopak v genomických modelech (G1 a G2) byl vliv UPG spíše mírně záporný, neboť se zvýšilo absolutní vychýlení i disperze (G1: $\overline{\mu_{p,w}} = \overline{1 - b_{w,p}} = 0,05$ vs. G1_UPG: $\overline{\mu_{p,w}} = \overline{1 - b_{w,p}} = 0,12$) a zachováno bylo pouze zlepšení v populační přesnosti díky ADG (G1: $\overline{acc_{LR}} = 0,34$ vs. G1_UPG: $\overline{acc_{LR}} = 0,39$). Souhrnně lze porovnání tradičních i genomických modelů vyjádřit jako A1_UPG > G1 > G1_UPG > A1, tedy že genomický model je lepší než tradiční pouze bez UPG. Zdá se tedy, že zpřesnění příbuznosti díky matici **H** částečně suplovalo zpřesnění díky UPG v případě použití rodokmenové matice **A**. Tyto výsledky mohou souviset s odlišným managementem chovu než u mateřských plemen a také s vyšším poměrem známých genotypů vůči fenotypům (SIR: 2,1 %; CLW: 1,5 %; CLA: 1,2 %).

Avšak z validace pouze pro plemeno pietrain by byly vyvozeny jiné závěry (tabulka 5.2-I). Výhoda tohoto přístupu spočívala ve větší uniformitě kohorty, která se projevila větší normalitou EBV pro BF a celkově větší normalitou spolehlivostí. Nicméně vzhledem k malému počtu jedinců (324) byla distribuce EBV pro ADG a MD poměrně špičatá a zároveň se zvýrazňoval vliv extrémních hodnot. V tomto případě byl ze všech modelů nejlepší A2 (náhodné SRO, bez UPG). Ekvivalentní genomický model (G2) dosáhl podobného průměrného absolutního vychýlení i disperze (A2: $\overline{\mu_{p,w}} = 0,06$; $\overline{1 - b_{w,p}} = 0,08$ vs. G2: $\overline{\mu_{p,w}} = 0,06$; $\overline{1 - b_{w,p}} = 0,11$), ale v ostatních LR metrikách byl horší (A2: $\overline{\rho^{-1}} = 1,83$; $\overline{acc_{LR}} = 0,22$; $\overline{PrA} = 0,13$ vs. G2: $\overline{\rho^{-1}} = 2,09$; $\overline{acc_{LR}} = 0,18$; $\overline{PrA} = 0,11$). Převaha genomického modelu se tedy projevila pouze ve vyšších průměrných individuálních přesnostech genotypovaných zvířat z validační kohorty, a to navýšením o zhruba 6 % v kompletním souboru a o 17 % v parciálním souboru. Ve víceplemenné validační kohortě (první způsob validace) bylo zaznamenáno zvýšení o 11 %, respektive 17 %.

V porovnání s mateřskými plemeny byly u otcovských plemen pozorovány odlišné změny nejen v prediktivní výkonnosti, nýbrž také v populačních genetických parametrech (tabulka 5.2-D). Při aplikaci ssGREML došlo k nárůstu odhadu aditivního genetického rozptylu u všech tří vlastností,

zvláště pak pro BF (+42 %) a MD (+32 %), díky čemuž se zvýšily také koeficienty dědivosti. Angarita Barajas et al. (2023) tento jev připisují skutečnosti, že matice genomické příbuznosti by měla lépe vyjádřit proměnlivost uvnitř rodin (mendelistický výběr). S tím mohou souviset i uvedené rozdíly v LR metrikách v případě víceplemenné validační kohorty, kde se významně projevilo přidání genotypů či UPG.

Ověření výkonnosti modelu pro otcovská plemena celkově potvrdilo skutečnost, že i metoda LR má své limity, které plynou z nevhodného modelu (Macedo et al. 2020b) nebo nevhodně zvolené validační kohorty (Salek Ardestani et al. 2021). Sloučení více populací do jedné analýzy může být z hlediska prediktivní výkonnosti prospěšné (Zhou et al. 2019), ale současně klade větší nároky na správnost modelu (Jang et al. 2023). Proto by bylo vhodné zvážit složitější model. Macedo et al. (2020a) a Jang et al. (2023) také doporučili použít strategii metazakladatelů (angl. metafounders) namísto UPG, obzvláště v případě analýzy více linií najednou. Nicméně ani takový krok neodstraní nedostatek propojenosti, který nastává kvůli vysokému podílu importovaných inseminačních dávek, jež navíc nejsou sdíleny mezi chovy. Další problém může způsobit nevyvážená struktura dat. Zároveň zůstává otázkou, zda EBV plemen jako duroc a pietrain, která se významně liší právě v přírůstku nebo zmasilosti, mohou být zcela sloučeny do jednoho rozdělení, i když dochází ke křížení populací.

5.2.5. Efekty SNP

Tabulky 5.2-J až 5.2-L a příloha 3 zobrazují výsledky z *post hoc* analýzy efektů jednotlivých SNP. Na základě korigované *p*-hodnoty by nebylo možné žádný z těchto efektů považovat za statisticky průkazný, neboť předpovědi aditivního účinku byly velmi nízké. Například v populaci českého bílého ušlechtilého byly nejvyšší účinky předpovězeny pro marker ALGA0041464, přičemž hodnoty dosáhly pouze 0,122 g pro ADG, 0,001 mm pro BF a 0,007 mm pro MD. Přesto mezi účinky SNP existovaly řádové rozdíly, které se promítly do podílu vysvětlené proměnlivosti. Nejvýznamnější jednotlivé markery vysvětlovaly 0,10–0,15 % aditivního genetického rozptylu dané vlastnosti. Součet 50 nejvýznamnějších markerů pak vysvětloval kolem 3,92 %, 3,67 % a 3,01 % rozptylu u českého bílého ušlechtilého, české landrase, respektive otcovských plemen. Ve všech třech souborech platilo, že marker s největším podílem vysvětlené proměnlivosti byl stejný pro všechny tři sledované vlastnosti, ale zároveň se velikosti účinků markerů neopakovaly napříč soubory. Toto pozorování ukazuje, že efekty SNP pro kvantitativní vlastnosti jsou populačně specifické.

Naše výsledky jsou srovnatelné se studií Kabanova et al. (2022), kteří v populaci bílého ušlechtilého zjistili, že na vysvětlení 5 % genetické proměnlivosti ve vlastnosti věk při 100 kg hmotnosti, BF a MD stačilo 71–108 markerů. Tito autoři zároveň porovnali ssGBLUP a WssGBLUP v rámci individuálních přesností prověřených zvířat a dosáhli zlepšení z průměrných 0,85 na 0,95.

Ruan et al. (2021) se v ssGWAS zaměřili na okna markerů o délce 0,8 Mb a ta nejvýznamnější z nich (> 1 %) dohromady vysvětlovala 17–24 % proměnlivosti podobných vlastností. Také Nosková et al. (2023) nebo Park (2024) při QTL mapování identifikovali vysoce významné regiony (například i v genu *MC4R*) a poukázali na pleiotropní účinek těchto lokusů. Na druhou stranu Jang et al. (2023) ve sdružené analýze tří komerčních linií pomocí metody LR pozorovali, že WssGBLUP oproti klasické ssGBLUP přinesla takřka nulové zlepšení. Nicméně jako možné vysvětlení bylo uvedeno, že existuje nepřímá úměra mezi užitečností vážení markerů a velikostí genotypované populace. Tato studie totiž pracovala s velice rozsáhlým množstvím údajů o SNP: téměř 207 tisíci genotypovanými jedinci, pro něž byly pomocí imputace navíc doplněny genotypy z celogenomového sekvenování.

Tabulka 5.2-D: Složky rozptylu spolu s koeficienty dědivosti (h^2) a genetické korelace mezi vlastnostmi (r_A) v doporučených konvenčních a genomických modelech pro jednotlivá plemena.

Plemeno (model)	Vl.	Přímé genetické efekty			Další náhod. efekty		Koeficienty		
		1 (ADG)	2 (BF)	3 (MD)	<i>vrh</i>	<i>R</i>	h^2	r_A	
CLW (A1)	ADG	441,40	3,62	19,50	620,20	1597,00	0,22	1-2	0,69
	BF	3,62	0,06	0,24	0,12	0,70	0,08	1-3	0,70
	MD	19,50	0,24	1,74	2,24	14,73	0,11	2-3	0,74
CLW (G1)	ADG	419,20	3,51	19,06	626,70	1609,00	0,21	1-2	0,69
	BF	3,51	0,06	0,25	0,12	0,70	0,08	1-3	0,70
	MD	19,06	0,25	1,76	2,24	14,73	0,11	2-3	0,74
CLA (A1)	ADG	402,40	2,99	21,83	357,80	1125,00	0,26	1-2	0,67
	BF	2,99	0,05	0,18	0,06	0,48	0,09	1-3	0,93
	MD	21,83	0,18	1,37	1,50	13,34	0,09	2-3	0,70
CLA (G1)	ADG	364,40	2,98	19,78	364,60	1144,00	0,24	1-2	0,68
	BF	2,98	0,05	0,18	0,06	0,48	0,10	1-3	0,93
	MD	19,78	0,18	1,25	1,52	13,40	0,09	2-3	0,71
SIR (A1)	ADG	966,20	5,95	45,40	559,60	1619,00	0,37	1-2	0,53
	BF	5,95	0,13	0,38	0,16	0,73	0,15	1-3	0,92
	MD	45,40	0,38	2,53	4,03	23,98	0,10	2-3	0,66
SIR (G1)	ADG	1006,00	6,37	43,30	570,60	1597,00	0,39	1-2	0,47
	BF	6,37	0,18	0,28	0,15	0,70	0,21	1-3	0,75
	MD	43,30	0,28	3,34	3,85	23,61	0,12	2-3	0,36

Zkratky: ADG = průměrný denní přírůstek; BF = výška hřbetního tuku; CLA = česká landrase; CLW = české bílé ušlechtilé; MD = výška svalu; R = reziduální rozptyl; SIR = otcovská plemena

Tabulka 5.2-E: Porovnání prediktivní výkonnosti BLUP a ssGBLUP v podobných studiích.

Zdroj	Plem.	Fenotyp	Genotyp	VI.	Disperze A → H	Přesnost A → H	Poznámky
Angarita Barajas et al. (2023)	F2 DU×PE	930	612 (65,8 %)	BF		0,73 → 0,77	$acc = r(y_{obs}, y_{pred})$
				MA		0,67 → 0,72	
Hong et al. (2019)	LW	21 762	886 (4,1 %)	ADG		0,26 → 0,28	$acc = PrA/h$
	LA	13 166	1 029 (7,8 %)	ADG		0,47 → 0,53	
Jibrila et al. (2022)	a	160 426	50 895 (31,7 %)	ADG	0,00 → 0,12	0,24 → 0,38	$acc = r(EBV, PYD)$, kde PYD = odchylka v užítkovosti potomků; přepočtená disperze; nespecifikovaná mateřská (a) a otcovská (b) linie
				BF	-0,02 → 0,18	0,34 → 0,45	
	MD	-0,14 → -0,06	0,43 → 0,55				
	ADG	-0,10 → 0,04	0,41 → 0,47				
b	75 129	33 506 (44,6 %)	BF	0,08 → 0,12	0,48 → 0,52		
			MD	0,04 → 0,04	0,49 → 0,62		
Lopez et al. (2019)	DU	24 828	560 (2,3 %)	ADG	0,43 → 0,38	0,29 → 0,30	$acc = PrA/h$
				BF	0,19 → 0,17	0,32 → 0,33	
Melnikova et al. (2021)	LW	41 941	1 194 (2,8 %)	A100	0,46 → 0,27	0,39 → 0,60	$acc = r(EBV_w, EBV_p)$ pro 110 vysoce spolehlivých zvířat
				BF	0,17 → 0,12	0,56 → 0,76	
				MD	0,07 → 0,07	0,61 → 0,77	
Poulsen et al. (2020)	LA	11 255	10 988 (97,7 %)	ADG		0,34 → 0,48	$acc = PrA/h$
Salek Ardestani et al. (2021)	LA	32 305	5 362 (16,6 %)	ADG	0,00 → -0,27	0,16 → 0,35	$acc = r(EBV_w, EBV_p)$ deregistrované EBV_w
				BF	0,24 → -0,14	0,29 → 0,53	
	MD	0,25 → -0,01	0,17 → 0,23				
	ADG	0,02 → -1,36	0,06 → 0,24				
DU	50 049	6 649 (13,3 %)	BF	0,10 → -0,60	0,14 → 0,39		
			MD	0,42 → -0,48	0,04 → 0,13		
Zhou et al. (2019)	LW _{as}	23 213	1 209 (5,2 %)	A100	0,17 → 0,00	0,43 → 0,59	$acc = PrA/h$; dvě linie; ve validaci pouze genotypovaná zvířata (182 a 291)
				BF	0,25 → 0,28	0,50 → 0,63	
	LW _b	13 280	1 199 (9,0 %)	A100	0,05 → 0,58	0,38 → 0,32	
				BF	-0,28 → -0,37	0,80 → 0,93	

Zkratky: A → H = změna parametru z rodokmenové (BLUP) na genomickou (ssGBLUP) předpověď;

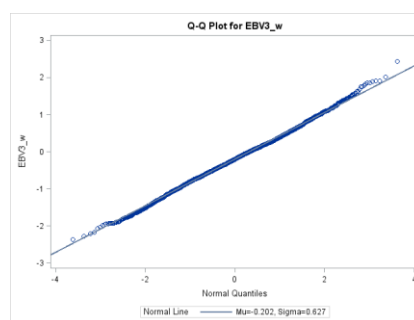
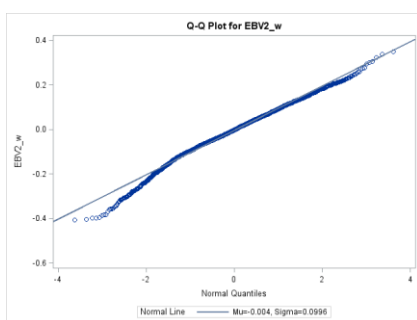
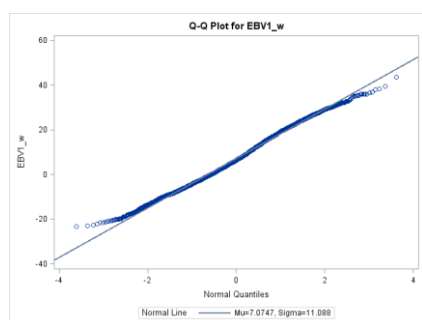
A100 = věk ve 100 kg; acc = způsob výpočtu kritéria přesnosti; ADG = průměrný denní přírůstek; BF = výška hřbetního tuku; DU = duroc; EBV_w a EBV_p = plemenná hodnota z kompletních a parciálních dat; LA = landrase;

LW = bílé ušlechtilé; MA = plocha svalu; MD = výška svalu; PE = pietrain; PrA = prediktivita;

y_{obs} a y_{pred} = fenotyp pozorovaný a předpovězený;

Tabulka 5.2-F: Parametry výkonnosti modelů a rozdělení EBV českého bílého ušlechtilého.

Model	Vlastnost	$\mu_{p,w}$	$1 - b_{w,p}$	ρ^{-1}	acc_{LR}	PrA	Celá val. kohorta $n = 4\ 267$		Genotyp. kohorta $n = 52$	
							r_w	r_p	r_w	r_p
A1	ADG	0,02	-0,02	1,76	0,30	0,14	0,62	0,46	0,72	0,46
	BF	0,02	0,15	1,74	0,25	0,08	0,58	0,43	0,68	0,43
	MD	-0,02	0,06	1,67	0,30	0,09	0,60	0,44	0,70	0,44
A1_UPG	ADG	0,01	0,01	1,78	0,30	0,14	0,55	0,35	0,65	0,34
	BF	0,00	0,17	1,75	0,25	0,08	0,50	0,29	0,60	0,29
	MD	-0,04	0,11	1,73	0,30	0,09	0,52	0,32	0,63	0,32
A2	ADG	-0,04	0,04	1,77	0,34	0,17	0,64	0,47	0,74	0,47
	BF	0,02	0,31	1,69	0,30	0,09	0,60	0,44	0,70	0,44
	MD	-0,01	0,25	1,69	0,33	0,10	0,61	0,45	0,71	0,45
A2_UPG	ADG	-0,09	0,06	1,80	0,34	0,16	0,57	0,37	0,68	0,37
	BF	0,06	0,35	1,73	0,30	0,09	0,52	0,32	0,63	0,32
	MD	-0,03	0,26	1,71	0,33	0,10	0,54	0,34	0,64	0,33
G1	ADG	0,01	0,01	1,61	0,34	0,15	0,62	0,48	0,75	0,57
	BF	0,03	0,17	1,68	0,27	0,08	0,59	0,45	0,72	0,53
	MD	0,00	0,11	1,65	0,31	0,09	0,60	0,46	0,74	0,55
G1_UPG	ADG	-0,01	0,02	1,61	0,34	0,15	0,55	0,37	0,69	0,47
	BF	-0,03	0,18	1,69	0,27	0,08	0,50	0,32	0,65	0,42
	MD	-0,02	0,16	1,69	0,31	0,08	0,52	0,34	0,67	0,45
G2	ADG	-0,06	0,07	1,65	0,37	0,16	0,64	0,49	0,77	0,58
	BF	0,00	0,31	1,70	0,29	0,09	0,61	0,47	0,74	0,55
	MD	-0,03	0,25	1,71	0,32	0,09	0,62	0,47	0,75	0,56
G2_UPG	ADG	-0,12	0,08	1,65	0,37	0,16	0,57	0,39	0,71	0,50
	BF	0,00	0,35	1,77	0,29	0,08	0,53	0,35	0,68	0,46
	MD	-0,04	0,27	1,74	0,32	0,09	0,54	0,36	0,69	0,47



Poznámky: modře zvýrazněn BLUP-AM a zeleně zvýrazněn ssGBLUP-AM s nejlepšími hodnotami pro danou metriku

(průměr za všechny tři vlastnosti, přesnost na tři desetinná místa);

kvantilové grafy (Q-Q ploty) na základě modelu A1_UPG (v pořadí ADG, BF, MD)

Zkratky: $1 - b_{w,p}$ = disperze EBV; acc_{LR} = populační přesnost podle metody LR; ADG = průměrný denní

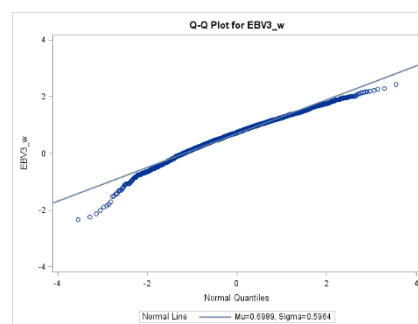
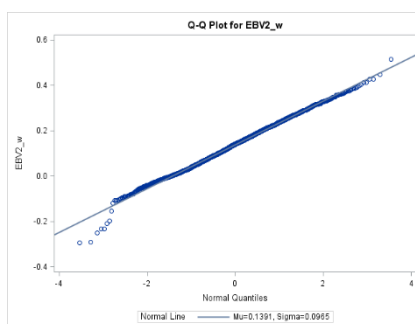
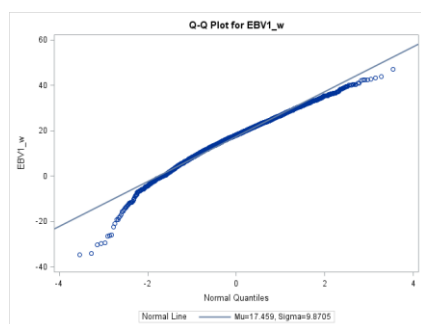
přírůstek [g]; BF = výška hřbetního tuku [mm]; MD = výška svalu [mm]; n = počet zvířat;

$\mu_{p,w}$ = standardizované vychýlení EBV; PrA = prediktivita; ρ^{-1} = inverzní korelace EBV;

$r_{w/p}$ = průměr individuálních přesností pro kompletní/parciální soubor

Tabulka 5.2-G: Parametry výkonnosti modelů a rozdělení EBV české landrase.

Model	Vlastnost	$\mu_{p,w}$	$1 - b_{w,p}$	ρ^{-1}	acc_{LR}	PrA	Celá val. kohorta $n = 3\ 151$		Genotyp. kohorta $n = 35$	
							r_w	r_p	r_w	r_p
A1	ADG	0,19	0,42	2,13	0,30	0,14	0,65	0,47	0,75	0,50
	BF	0,08	0,56	2,41	0,28	0,05	0,59	0,42	0,68	0,45
	MD	0,21	0,46	2,09	0,32	0,07	0,64	0,46	0,74	0,49
A1_UPG	ADG	0,11	0,28	1,81	0,32	0,17	0,57	0,31	0,68	0,36
	BF	0,00	0,53	2,24	0,29	0,05	0,48	0,21	0,60	0,27
	MD	0,12	0,32	1,73	0,36	0,09	0,54	0,28	0,66	0,34
A2	ADG	0,16	0,42	2,19	0,28	0,13	0,66	0,47	0,75	0,50
	BF	0,05	0,45	1,96	0,33	0,08	0,60	0,43	0,70	0,46
	MD	0,20	0,54	2,36	0,29	0,05	0,63	0,45	0,73	0,48
A2_UPG	ADG	0,13	0,32	2,05	0,28	0,14	0,57	0,32	0,68	0,37
	BF	0,02	0,43	1,86	0,35	0,08	0,50	0,24	0,62	0,30
	MD	0,15	0,47	2,13	0,31	0,06	0,54	0,28	0,65	0,34
G1	ADG	0,12	0,26	1,84	0,30	0,17	0,64	0,47	0,76	0,56
	BF	0,04	0,41	1,95	0,29	-0,01	0,59	0,43	0,71	0,52
	MD	0,14	0,28	1,77	0,32	0,00	0,63	0,46	0,75	0,55
G1_UPG	ADG	0,02	0,17	1,69	0,31	0,17	0,56	0,32	0,69	0,44
	BF	-0,04	0,39	1,86	0,30	0,07	0,49	0,24	0,63	0,38
	MD	0,03	0,22	1,59	0,35	0,09	0,53	0,30	0,67	0,43
G2	ADG	0,12	0,32	1,92	0,29	0,14	0,65	0,48	0,77	0,57
	BF	0,03	0,37	1,77	0,34	0,09	0,60	0,44	0,72	0,53
	MD	0,15	0,40	1,95	0,31	0,06	0,62	0,46	0,75	0,55
G2_UPG	ADG	0,04	0,21	1,84	0,28	0,15	0,56	0,33	0,70	0,45
	BF	0,00	0,34	1,68	0,35	0,09	0,50	0,26	0,65	0,40
	MD	0,08	0,34	1,80	0,32	0,07	0,53	0,30	0,67	0,43



Poznámky: modře zvýrazněn BLUP-AM a zeleně zvýrazněn ssGBLUP-AM s nejlepšími hodnotami pro danou metriku

(průměr za všechny tři vlastnosti, přesnost na tři desetinná místa);

kvantilové grafy (Q-Q ploty) na základě modelu A1_UPG (v pořadí ADG, BF, MD)

Zkratky: $1 - b_{w,p}$ = disperze EBV; acc_{LR} = populační přesnost podle metody LR; ADG = průměrný denní

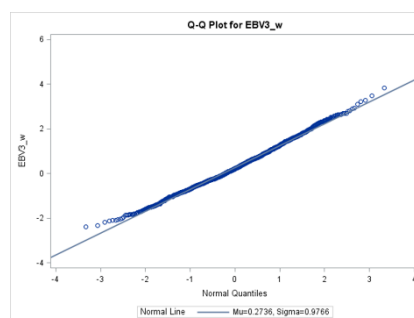
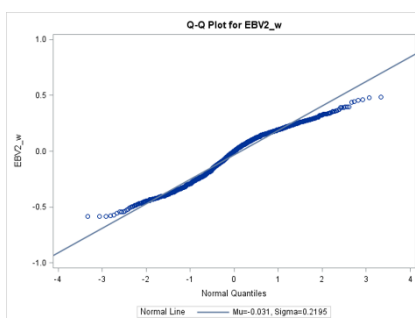
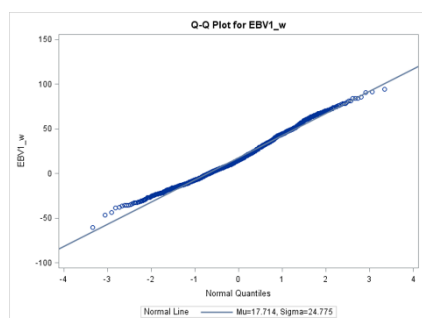
přírůstek [g]; BF = výška hřbetního tuku [mm]; MD = výška svalu [mm]; n = počet zvířat;

$\mu_{p,w}$ = standardizované vychýlení EBV; PrA = prediktivita; ρ^{-1} = inverzní korelace EBV;

r_w/p = průměr individuálních přesností pro kompletní/parciální soubor

Tabulka 5.2-H: Parametry výkonnosti modelů a rozdělení EBV otcovských plemen.

Model	Vlastnost	$\mu_{p,w}$	$1 - b_{w,p}$	ρ^{-1}	acc_{LR}	PrA	Celá val. kohorta $n = 1473$		Genotyp. kohorta $n = 30$	
							r_w	r_p	r_w	r_p
A1	ADG	-0,06	0,28	2,21	0,38	0,18	0,66	0,41	0,68	0,44
	BF	-0,07	-0,42	1,74	0,29	0,23	0,56	0,35	0,58	0,37
	MD	-0,07	0,22	2,27	0,31	0,11	0,64	0,40	0,66	0,42
A1_UPG	ADG	-0,11	0,01	1,61	0,50	0,28	0,56	0,23	0,63	0,43
	BF	0,00	-0,37	1,11	0,61	0,45	0,41	NA	0,51	0,32
	MD	-0,12	0,06	1,67	0,42	0,08	0,51	0,16	0,59	0,40
A2	ADG	-0,06	0,24	2,14	0,38	0,18	0,67	0,42	0,70	0,45
	BF	-0,03	-0,39	1,65	0,30	0,25	0,59	0,37	0,61	0,39
	MD	-0,07	0,08	2,06	0,28	0,11	0,63	0,39	0,65	0,42
A2_UPG	ADG	-0,11	0,01	1,57	0,52	0,29	0,57	0,25	0,64	0,45
	BF	0,04	-0,30	1,13	0,58	0,45	0,46	0,08	0,55	0,36
	MD	-0,13	0,19	1,86	0,35	0,07	0,50	0,15	0,58	0,39
G1	ADG	-0,05	-0,05	1,77	0,43	0,25	0,66	0,43	0,73	0,58
	BF	-0,02	-0,58	1,14	0,57	0,49	0,58	0,39	0,67	0,54
	MD	-0,05	0,05	2,12	0,26	0,09	0,57	0,38	0,66	0,53
G1_UPG	ADG	-0,10	-0,02	1,57	0,52	0,29	0,56	0,23	0,63	0,44
	BF	0,03	-0,28	1,10	0,61	0,48	0,46	0,08	0,55	0,36
	MD	-0,14	0,22	2,37	0,26	0,07	0,44	NA	0,53	0,34
G2	ADG	-0,04	-0,05	1,73	0,43	0,25	0,67	0,44	0,74	0,59
	BF	0,03	-0,49	1,15	0,55	0,48	0,60	0,40	0,68	0,55
	MD	-0,07	0,12	2,13	0,27	0,08	0,59	0,39	0,67	0,54
G2_UPG	ADG	-0,11	0,00	1,57	0,52	0,29	0,58	0,26	0,65	0,45
	BF	0,08	-0,24	1,12	0,59	0,47	0,48	0,14	0,57	0,39
	MD	-0,17	0,26	2,26	0,29	0,08	0,47	0,08	0,55	0,36



Poznámky: modře zvýrazněn BLUP-AM a zeleně zvýrazněn ssGBLUP-AM s nejlepšími hodnotami pro danou metriku

(průměr za všechny tři vlastnosti, přesnost na tři desetinná místa);

kvantilové grafy (Q-Q ploty) na základě modelu A1_UPG (v pořadí ADG, BF, MD)

Zkratky: $1 - b_{w,p}$ = disperze EBV; acc_{LR} = populační přesnost podle metody LR; ADG = průměrný denní

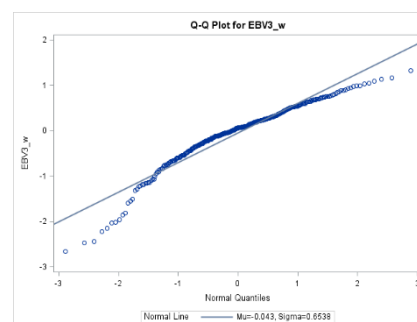
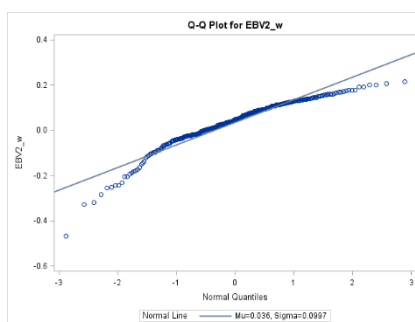
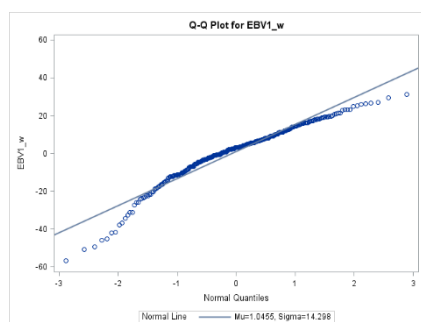
přírůstek [g]; BF = výška hřbetního tuku [mm]; MD = výška svalu [mm]; n = počet zvířat;

$\mu_{p,w}$ = standardizované vychýlení EBV; PrA = prediktivita; ρ^{-1} = inverzní korelace EBV;

r_w/p = průměr individuálních přesností pro kompletní/parciální soubor

Tabulka 5.2-I: Parametry výkonnosti modelů a rozdělení EBV pietraina.

Model	Vlastnost	$\mu_{p,w}$	$1 - b_{w,p}$	ρ^{-1}	acc_{LR}	PrA	Celá val. kohorta $n = 324$		Genotyp. kohorta $n = 14$	
							r_w	r_p	r_w	r_p
A1	ADG	-0,10	-0,28	1,99	0,23	0,17	0,63	0,41	0,65	0,40
	BF	-0,14	-0,14	1,60	0,17	0,14	0,51	0,32	0,53	0,30
	MD	-0,07	-0,16	1,93	0,22	0,06	0,60	0,39	0,62	0,38
A1_UPG	ADG	0,03	0,19	2,72	0,19	0,10	0,56	0,30	0,64	0,45
	BF	-0,09	0,02	1,78	0,16	0,12	0,40	0,08	0,53	0,35
	MD	0,05	0,22	2,72	0,18	0,01	0,52	0,26	0,61	0,43
A2	ADG	0,04	-0,11	2,04	0,23	0,16	0,65	0,44	0,68	0,42
	BF	-0,04	-0,04	1,57	0,19	0,16	0,56	0,36	0,58	0,34
	MD	0,10	0,09	1,89	0,23	0,06	0,60	0,40	0,63	0,38
A2_UPG	ADG	0,08	0,16	2,52	0,20	0,11	0,59	0,34	0,67	0,48
	BF	-0,04	-0,02	1,64	0,19	0,15	0,47	0,21	0,58	0,40
	MD	0,13	0,17	2,21	0,20	0,04	0,52	0,27	0,62	0,43
G1	ADG	-0,06	0,04	3,05	0,16	0,09	0,64	0,45	0,71	0,56
	BF	-0,11	0,00	1,85	0,14	0,12	0,55	0,39	0,65	0,52
	MD	0,02	0,22	2,63	0,16	0,02	0,55	0,38	0,64	0,51
G1_UPG	ADG	0,01	0,21	2,84	0,19	0,09	0,56	0,30	0,65	0,45
	BF	-0,11	0,00	1,67	0,16	0,14	0,45	0,17	0,57	0,38
	MD	0,13	0,39	2,97	0,16	0,00	0,44	0,16	0,56	0,38
G2	ADG	-0,01	-0,13	2,50	0,18	0,13	0,66	0,47	0,74	0,58
	BF	-0,09	-0,06	1,64	0,17	0,16	0,58	0,41	0,67	0,54
	MD	0,07	0,13	2,12	0,19	0,05	0,58	0,41	0,67	0,53
G2_UPG	ADG	0,08	0,17	2,62	0,20	0,11	0,60	0,35	0,68	0,48
	BF	-0,08	-0,02	1,57	0,19	0,17	0,50	0,24	0,60	0,42
	MD	0,18	0,29	2,30	0,20	0,03	0,48	0,22	0,59	0,41



Poznámky: modře zvýrazněn BLUP-AM a zeleně zvýrazněn ssGBLUP-AM s nejlepšími hodnotami pro danou metriku

(průměr za všechny tři vlastnosti, přesnost na tři desetinná místa);

kvantilové grafy (Q-Q ploty) na základě modelu A1_UPG (v pořadí ADG, BF, MD)

Zkratky: $1 - b_{w,p}$ = disperze EBV; acc_{LR} = populační přesnost podle metody LR; ADG = průměrný denní

přírůstek [g]; BF = výška hřbetního tuku [mm]; MD = výška svalu [mm]; n = počet zvířat;

$\mu_{p,w}$ = standardizované vychýlení EBV; PrA = prediktivita; ρ^{-1} = inverzní korelace EBV;

r_w/p = průměr individuálních přesností pro kompletní/parciální soubor

Tabulka 5.2-J: Pět markerů s největším podílem vysvětlené proměnlivosti pro průměrný denní přírůstek (ADG), výšku tuku (BF) a výšku svalů (MD) v populaci českého bílého ušlechtilého.

Vlastnost	Pořadí	Název SNP	Chromozom	Pozice	$-\log_{10}p$	Vysvětlený rozptyl [%]
ADG	1	ALGA0041464	7	52279313	4,02	0,15
ADG	2	ALGA0041501	7	52574518	3,39	0,12
ADG	3	MARC0092956	2	27214992	2,83	0,10
ADG	4	DRGA0016654	17	25366971	3,07	0,10
ADG	5	H3GA0000476	1	7211765	3,28	0,09
BF	1	ALGA0041464	7	52279313	3,05	0,14
BF	2	ALGA0047701	8	40219544	3,14	0,12
BF	3	MARC0015472	3	37832622	2,85	0,11
BF	4	ASGA0068611	15	13218467	2,78	0,10
BF	5	ALGA0079177	14	86517116	2,39	0,10
MD	1	ALGA0041464	7	52279313	3,65	0,15
MD	2	ALGA0041501	7	52574518	2,87	0,12
MD	3	DRGA0007276	7	19232432	2,67	0,11
MD	4	DRGA0007294	7	20439045	2,58	0,11
MD	5	SIRI0001291	14	24132815	3,03	0,09

Tabulka 5.2-K: Pět markerů s největším podílem vysvětlené proměnlivosti pro průměrný denní přírůstek (ADG), výšku tuku (BF) a výšku svalů (MD) v populaci české landrase.

Vlastnost	Pořadí	Název SNP	Chromozom	Pozice	$-\log_{10}p$	Vysvětlený rozptyl [%]
ADG	1	MARC0011216	7	108972110	4,10	0,12
ADG	2	DRGA0011974	13	11399800	4,28	0,10
ADG	3	MARC0012283	6	44004031	3,89	0,10
ADG	4	MARC0011192	15	91900610	3,52	0,10
ADG	5	DRGA0012131	13	18324629	3,13	0,09
BF	1	MARC0011216	7	108972110	4,43	0,13
BF	2	MARC0012227	12	4276148	4,39	0,11
BF	3	MARC0011801	9	153484548	4,19	0,11
BF	4	MARC0012294	18	280810	3,43	0,10
BF	5	MARC0012272	1	44136139	3,90	0,10
MD	1	MARC0011216	7	108972110	4,32	0,13
MD	2	MARC0012283	6	44004031	4,12	0,10
MD	3	MARC0012294	18	280810	3,03	0,10
MD	4	MARC0011192	15	91900610	3,40	0,09
MD	5	DRGA0012131	13	18324629	2,90	0,09

Tabulka 5.2-L: Pět markerů s největším podílem vysvětlené proměnlivosti pro průměrný denní přírůstek (ADG), výšku tuku (BF) a výšku svalů (MD) v souboru otcovských plemen.

Vlastnost	Pořadí	Název SNP	Chromozom	Pozice	$-\log_{10}P$	Vysvětlený rozptyl [%]
ADG	1	ASGA0050421	11	24892871	5,32	0,13
ADG	2	MARC0009254	6	16449770	3,08	0,09
ADG	3	H3GA0007165	2	93318735	4,08	0,09
ADG	4	ALGA0077805	14	60273398	3,65	0,08
ADG	5	ALGA0008113	1	249012536	4,43	0,08
BF	1	H3GA0013062	4	81984659	3,14	0,10
BF	2	ASGA0050421	11	24892871	2,98	0,09
BF	3	DIAS0004390	13	87988310	2,90	0,09
BF	4	DRGA0016667	17	28147853	2,67	0,08
BF	5	H3GA0012910	4	76954048	2,25	0,07
MD	1	ASGA0050421	11	24892871	4,56	0,12
MD	2	ALGA0008933	1	275078557	4,42	0,09
MD	3	ALGA0008113	1	249012536	4,35	0,08
MD	4	DRGA0003055	2	85073563	3,01	0,08
MD	5	ASGA0038189	8	22805463	3,05	0,08

6. Závěr

Tato dizertační práce si kladla za cíl posoudit přínos molekulárních markerů pro genetické hodnocení prasat v českém šlechtitelském programu. Hlavní část práce proto propojovala požadavek na návrh nového statistického modelu s daty z DNA čipů, aby mohl být srovnán konvenční BLUP animal model a jednokrokový genomický model.

Pro praktické využití byl doporučen tříznakový model pro současné hodnocení průměrného denního přírůstku, výšky hřbetního tuku a výšky svalu, v němž figurují pevné efekty plemene (v případě otcovských linií), pohlaví, SRO a ultrazvukového přístroje a náhodné efekty vrhu a zvířete. Výběr modelu byl založen na parametrech dopředné validace, které vyjadřují prediktivní výkonnost statistických modelů. Smysluplnost těchto parametrů závisí na rozdělení plemenných hodnot, takže se odvíjí od výběru skupiny zvířat pro validaci nebo od definice efektů v modelu. Jmenovitě může výsledky výrazně ovlivnit efekt společného prostředí zvířat (např. SRO). Tento efekt je obvykle interakcí hlavních faktorů, která zohledňuje nejen známé, nýbrž i skryté systematické trendy napříč chovy a časem. V praxi mnohdy efekt společného prostředí funguje lépe jako pevný, neboť náhodný efekt společného prostředí působí zkreslení výsledků, pokud nejsou očištěny systematické rozdíly na vyšších stupních hierarchie. V této souvislosti obecně platí, že modely používané ve šlechtění zvířat mohou dosáhnout vyšší přesnosti při dodržení podstaty víceúrovňového modelování. Ovšem příliš složitý model už zlepšení nepřináší. V naší analýze se takto ukázalo být nadbytečné vnoření efektu vrhu do SRO a vnoření otce do stáda. Striktní dodržování hierarchie v animal modelech zřejmě není nutné, protože na rozdíl od jiných smíšených modelů pracují s maticí příbuznosti, která propojuje celou populaci.

Zahrnutí SNP pomocí matice realizované příbuznosti mělo v českých populacích prasat zatím nízký účinek, který byl ovšem předpokládáný vzhledem k omezenému počtu genotypovaných zvířat. V hodnocení mateřských plemen zároveň lze tento účinek považovat za kladný, neboť se projevila převaha ssGBLUP v individuálních přesnostech, přičemž celková prediktivní výkonnost modelu byla v případě českého bílého ušlechtilého zachována a v případě české landrase dokonce zlepšena. Ve sdruženém hodnocení otcovských plemen nebylo možné učinit jednoznačný závěr, jelikož výsledky patrně zkreslila nevhodná struktura dat.

Přesto je možné konstatovat, že hlavní cíl této práce – navržení metodického postupu pro předpověď genomických plemenných hodnot v českém šlechtitelském programu prasat – byl naplněn. Současně lze podpořit hypotézu o tom, že molekulárněgenetické informace přispívají ke zpřesnění genetického hodnocení. Nicméně jak napovídá dostupná literatura, praktický dopad genomické

selekce na genetický zisk nastane až poté, co budou nashromážděna genotypová data od několikanásobně většího množství jedinců.

Dále naše analýza ukázala, že výška hřbetního tuku a výška svalu, tedy dva z hlavních produkčních ukazatelů, jsou doprovázeny velmi nízkými koeficienty dědivosti. Příčinou může být ztráta genetické proměnlivosti nebo také nepřesnosti při sběru užitkových dat. Navržený model pro přímé hodnocení těchto vlastností by sice měl zajistit lepší kontrolu jejich vývoje, ale zároveň bude klíčové pravidelně sledovat genetické trendy čili složky rozptylu a kovariance.

Vedlejší studie týkající se SNP v genu *MC4R* poukázala na možné nedostatky podobných asociačních analýz, ovšem také na potenciální efekt, který nelze považovat za nevýznamný. Ačkoliv přímé využití takových SNP pro šlechtění na kvantitativní vlastnosti prostřednictvím markery asistované selekce se jeví jako překonaný postup, asociační studie stále mohou působit jako doplněk rutinního genetického hodnocení pomocí animal modelů. Užitečné může být zejména vážení markerů prostřednictvím WssGBLUP. Sice převaha WssGBLUP nad jednodušší ssGBLUP není pravidlem, ale při nynější dostupnosti různých metod je možné poměrně snadno zvolit tu, která nejvíce vyhovuje podmínkám daného šlechtitelského programu.

7. Seznam použité literatury

- Afrazandeh M, Abdolahi-Arpanahi R, Abbasi MA, Kashan NEJ, Torshizi RV. 2022. Comparison of different response variables in genomic prediction using GBLUP and ssGBLUP methods in Iranian Holstein cattle. *Journal of Dairy Research* **89**:121–127.
- Agresti A. 2019. *An Introduction to Categorical Data Analysis* Third edition. Wiley, Hoboken.
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* **93**:743–752.
- Aguilar I, Tsuruta S, Masuda Y, Lourenco D. 2018. BLUPF90 suite of programs for animal breeding with focus on genomics. Page World Congress on Genetics Applied to Livestock Production. New Zealand.
- Amaya Martínez A, Martínez Sarmiento R, Cerón Muñoz M. 2019. Evaluaciones genéticas usando el mejor predictor lineal insesgado genómico en una etapa en bovinos. *Ciencia & Tecnología Agropecuaria* **21**:1–13.
- Angarita Barajas BK, Cantet RJC, Steibel JP, Schrauf MF, Forneris NS. 2023. Heritability estimates and predictive ability for pig meat quality traits using identity-by-state and identity-by-descent relationships in an F2 population. *Journal of Animal Breeding and Genetics* **140**:13–27.
- Babot D, Noguera JL, Alfonso L, Estany J. 2003. Fixed or random contemporary groups in genetic evaluation for litter size in pigs using a single trait repeatability animal model. *Journal of Animal Breeding and Genetics* **120**:12–22.
- Bao W-B, Ye L, Pan Z-Y, ZHU J, ZHU G-Q, HUANG X-G, WU S-L. 2011. Beneficial genotype of swine FUT1 gene governing resistance to E. coli F18 is associated with important economic traits. *Journal of Genetics* **90**:315–318.
- Bauer J, Přibyl J, Vostrý L. 2015. Short communication: Reliability of single-step genomic BLUP breeding values by multi-trait test-day model analysis. *Journal of Dairy Science* **98**:4999–5003.
- Beaulieu AD, Aalhus JL, Williams NH, Patience JF. 2010. Impact of piglet birth weight, birth order, and litter size on subsequent growth performance, carcass quality, muscle composition, and eating quality of pork1. *Journal of Animal Science* **88**:2767–2778.
- Bermann M, Aguilar I, Lourenco D, Misztal I, Legarra A. 2023. Reliabilities of estimated breeding values in models with metafounders. *Genetics Selection Evolution* **55**:6.

- Bermann M, Cesarani A, Misztal I, Lourenco D. 2022. Past, present, and future developments in single-step genomic models. *Italian Journal of Animal Science* **21**:673–685.
- Bermann M, Lourenco D, Breen V, Hawken R, Brito Lopes F, Misztal I. 2021. Modeling genetic differences of combined broiler chicken populations in single-step GBLUP. *Journal of Animal Science* **99**.
- Bizelis J, Kominakis A, Rogdakis E, Georgadopoulou F. 2000. Genetic parameters of production and reproductive traits in on a farm tested Danish Large White and Landrace swine in Greece. Page Arch. Tierz., Dummerstorf.
- Blume JD, Greevy RA, Welty VF, Smith JR, Dupont WD. 2019. An Introduction to Second-Generation *p*-Values. *The American Statistician* **73**:157–167.
- Bourdon RM. 2014. *Understanding Animal Breeding* Second Edition. Pearson, Harlow.
- BrzÁková M, Bauer J, Steyn Y, ŠplÍchal J, Fulínová D. 2022. The prediction accuracies of linear-type traits in Czech Holstein cattle when using ssGBLUP or wssGBLUP. *Journal of Animal Science* **100**.
- Calta J, Zadinová K, Čítek J, KluzÁková E, OkrouhlÁ M, Stupka R, Tichý L, Machová K, Stratil A, Vostrý L. 2023. Possible effects of the *MC4R* Asp298Asn polymorphism on pig production traits under ad libitum versus restricted feeding. *Journal of Animal Breeding and Genetics* **140**:207–215.
- Calta J, ŽÁková E, Krupa E, Čítek J, DvořÁková Machová K, Tichý L, Stibal J, Vostrý L. 2024. Selecting an appropriate statistical model for estimating genetic parameters: A case study of pig maternal breeds in Czechia. *Livestock Science* **283**:105457.
- Causser D, Daumas G, Dhorne T, Engel B, Font-i-Furnols M, Højsgaard S. 2003. Statistical handbook for assessing pig classification methods: Recommendations from the EUPIGCLASS project group. Dostupné z https://agriculture.ec.europa.eu/system/files/2019-08/statistical-handbook-for-assessing-pig-classification-methods_en_0.pdf (07.07.2023).
- Chen P, Baas TJ, Mabry JW, Dekkers JCM, Koehler KJ. 2002. Genetic parameters and trends for lean growth rate and its components in U.S. Yorkshire, Duroc, Hampshire, and Landrace pigs. *Journal of Animal Science* **80**:2062.
- Cheng J, Dekkers JCM, Fernando RL. 2021. Cross-validation of best linear unbiased predictions of breeding values using an efficient leave-one-out strategy. *Journal of Animal Breeding and Genetics* **138**:519–527.

- Christensen OF, Lund MS. 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* **42**:2.
- Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. 2012. Single-step methods for genomic evaluation in pigs. *animal* **6**:1565–1571.
- Csató L, Nagy I, Farkas J, Radnóczy L. 2002. Genetic parameters of production traits of Hungarian Pig populations evaluated in separate and joint (field and station) tests. *Page Arch. Tierz., Dummerstorf*.
- Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM. 2013. Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics* **193**:347–365.
- Davoli R, Catillo G, Serra A, Zappaterra M, Zambonelli P, Zilio DM, Steri R, Mele M, Buttazzoni L, Russo V. 2019. Genetic parameters of backfat fatty acids and carcass traits in large white pigs. *Animal* **13**:924–932. Cambridge University Press.
- Dekkers JCM. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J Anim Sci* **82**:313–348.
- Dekkers JCM, Mathur PK, Knol EF. 2011. Genetic Improvement of the Pig. Pages 390–425 in Rothschild MF, Ruvinsky A, editors. *The Genetics of the Pig*, 2nd edition. CABI, London.
- Dekkers JCM, Su H, Cheng J. 2021. Predicting the accuracy of genomic predictions. *Genetics Selection Evolution* **53**:55.
- Dong L, Tan C, Cai G, Li Y, Wu D, Wu Z. 2020. Estimates of variance components and heritability using different animal models for growth, Backfat, litter size, and healthy birth ratio in Large White Pigs. *Canadian Journal of Animal Science* **100**:330–336. Agricultural Institute of Canada.
- Dube B, Mulugeta SD, Dzama K. 2013. Genetic relationship between growth and carcass traits in Large White pigs. *South African Journal of Animal Sciences* **43**:482–492. South African Bureau for Scientific Publications.
- Dube B, Mulugeta SD, Dzama K. 2014. Investigating maternal effects on production traits in Duroc pigs using animal and sire models. *Journal of Animal Breeding and Genetics* **131**:279–293.
- Dvořáková V, Stupka R, Šprysl M, Čítek J, Okrouhlá M, Kluzáková E, Kratochvílová H. 2011. Effect of the missense mutation Asp298Asn in MC4R on growth and fatness traits in commercial pig crosses in the Czech Republic. *Czech Journal of Animal Science* **56**:176–180.

- Erbe M, Gredler B, Seefried FR, Bapst B, Simianer H. 2013. A Function Accounting for Training Set Size and Marker Density to Model the Average Accuracy of Genomic Prediction. *PLoS ONE* **8**:e81046.
- Esfandyari H, Thekkoot D, Kemp R, Plastow G, Dekkers J. 2020. Genetic parameters and purebred–crossbred genetic correlations for growth, meat quality, and carcass traits in pigs. *Journal of Animal Science* **98**.
- Evropská komise. 2013. Prováděcí rozhodnutí Komise ze dne 18. dubna 2013, kterým se mění rozhodnutí 2005/1/ES, kterým se schvalují metody třídění jatečně upravených těl prasat v České republice, pokud jde o vzorce používané u povolených metod a o obchodní úpravu těchto jatečně upravených těl.
- Evropský parlament a Rada EU. 2013. Nařízení Evropského parlamentu a Rady (EU) č. 1308/2013 ze dne 17. prosince 2013, kterým se stanoví společná organizace trhů se zemědělskými produkty a zrušují nařízení Rady (EHS) č. 922/72, (EHS) č. 234/79, (ES) č. 1037/2001 a (ES) č. 1234/2007.
- Falconer DS. 1989. Variance. Pages 125–147 *Introduction to quantitative genetics*, 3rd edition. Longman Scientific & Technical, Harlow.
- Fan B, Onteru SK, Plastow GS, Rothschild MF. 2009. Detailed characterization of the porcine *MC4R* gene in relation to fatness and growth. *Animal Genetics* **40**:401–409.
- Fangmann A, Sharifi RA, Heinkel J, Danowski K, Schrade H, Erbe M, Simianer H. 2017. Empirical comparison between different methods for genomic prediction of number of piglets born alive in moderate sized breeding populations¹. *Journal of Animal Science* **95**:1434–1443.
- Fernando R, Grossman M. 1989. Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* **21**:467.
- Field A, Miles J. 2010. *Discovering statistics using SAS*. SAGE Publications.
- Fontanesi L, Buttazzoni L, Galimberti G, Calò DG, Scotti E, Russo V. 2013. Association between melanocortin 4 receptor (MC4R) gene haplotypes and carcass and production traits in Italian Large White pigs evaluated with a selective genotyping approach. *Livestock Science* **157**:48–56. Elsevier. Dostupné z <http://dx.doi.org/10.1016/j.livsci.2013.07.006>.
- Font-i-Furnols M, Čandek-Potokar M, Daumas G, Gispert M, Judas M, Seynaeve M. 2016. Comparison of national ZP equations for lean meat percentage assessment in SEUROP pig classification. *Meat Science* **113**:1–8. Elsevier Ltd.

- Frey M, Hofer A, Künzi N. 1997. Comparison of models with a fixed or a random contemporary group effect for the genetic evaluation for litter size in pigs. *Livestock Production Science* **48**:135–141.
- Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, O'Brien PJ, MacLennan DH. 1991. Identification of a Mutation in Porcine Ryanodine Receptor Associated with Malignant Hyperthermia. *Science* **253**:448–451.
- Garnett T et al. 2013. Sustainable Intensification in Agriculture: Premises and Policies. *Science* **341**:33–34.
- Gelman A. 2005. Analysis of variance - why it is more important than ever. *The Annals of Statistics* **33**.
- Gelman A, Hill J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Genest P, Daniel P. 1966. Genomic Modifications in Cell Line Cultures Chronically Infected With a Myxovirus. *Experimental Biology and Medicine* **123**:722–725.
- Getmantseva L, Leonova M, Usatov A, Bakoev S, Klimenko A, Vasilenko V, Kolosov A, Shirockova N, Makarenko M. 2017. The single and combined effect of MC4R and GH genes on productive traits of pigs. *American Journal of Agricultural and Biological Sciences* **12**:28–32.
- Gianola D, Rosa GJM. 2015. One Hundred Years of Statistical Developments in Animal Breeding. *Annual Review of Animal Biosciences* **3**:19–56.
- Gomes DGE. 2022. Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? *PeerJ* **10**:e12794.
- Groenen MAM et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**:393–398.
- Guarini AR, Lourenco DAL, Brito LF, Sargolzaei M, Baes CF, Miglior F, Misztal I, Schenkel FS. 2018. Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. *Journal of Dairy Science* **101**:8076–8086.
- Guarini AR, Lourenco DAL, Brito LF, Sargolzaei M, Baes CF, Miglior F, Tsuruta S, Misztal I, Schenkel FS. 2019. Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. *Journal of Dairy Science* **102**:8175–8183.

- Haley CS, Visscher PM. 1998. Strategies to Utilize Marker-Quantitative Trait Loci Associations. *Journal of Dairy Science* **81**:85–97.
- Hashim HO, Al-Shuhaib MB. 2019. Exploring the Potential and Limitations of PCR-RFLP and PCR-SSCP for SNP Detection: A Review. *Journal of Applied Biotechnology Reports* **6**:137–144.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* **92**:433–443.
- Henryon M, Berg P, Ostersen T, Nielsen B, Sørensen AC. 2012. Most of the benefits from genomic selection can be realized by genotyping a small proportion of available selection candidates¹. *Journal of Animal Science* **90**:4681–4689.
- Hermesch S, Luxford BG, Graser H-U. 2000. Genetic parameters for lean meat yield, meat quality, reproduction and feed efficiency traits for Australian pigs 1. Description of traits and heritability estimates. *Page Livestock Production Science*. Dostupné z www.elsevier.com/locate/livprodsci.
- Hickey JM, Veerkamp RF, Calus MP, Mulder HA, Thompson R. 2009. Estimation of prediction error variances via Monte Carlo sampling methods using different formulations of the prediction error variance. *Genetics Selection Evolution* **41**:23.
- Hidalgo J, Tsuruta S, Lourenco D, Masuda Y, Huang Y, Gray KA, Misztal I. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. *Journal of Animal Science* **98**.
- Hill WG, Goddard ME, Visscher PM. 2008. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genetics* **4**:e1000008.
- Hillel J, Schaap T, Haberfeld A, Jeffrey AJ, Plotzky Y, Cahaner A, Lavi U. 1990. DNA fingerprints applied to gene introgression in breeding programs. *Genetics* **124**:783–789.
- Hollifield MK, Lourenco D, Bermann M, Howard JT, Misztal I. 2021a. Determining the stability of accuracy of genomic estimated breeding values in future generations in commercial pig populations. *Journal of Animal Science* **99**.
- Hollifield MK, Lourenco D, Tsuruta S, Bermann M, Howard JT, Misztal I. 2021b. Impact of including the cause of missing records on genetic evaluations for growth in commercial pigs. *Journal of Animal Science* **99**.
- Hong JK, Kim YS, Cho KH, Lee DH, Min YJ, Cho ES. 2019. Application of single-step genomic evaluation using social genetic effect model for growth in pig. *Asian-Australasian Journal of Animal Sciences* **32**:1836–1843.

- Houde AA, Méthot S, Murphy BD, Bordignon V, Palin MF. 2010. Relationships between backfat thickness and reproductive efficiency of sows: A two-year trial involving two commercial herds fixing backfat thickness at breeding. *Canadian Journal of Animal Science* **90**:429–436.
- Houston RD, Cameron ND, Rance KA. 2004. A melanocortin-4 receptor (MC4R) polymorphism is associated with performance traits in divergently selected large white pig populations. *Animal Genetics* **35**:386–390.
- Huang Y, Hickey JM, Cleveland MA, Maltecca C. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genetics Selection Evolution* **44**:25.
- Ibáñez-Escriche N, Forni S, Noguera JL, Varona L. 2014. Genomic information in pig breeding: Science meets industry needs. *Livestock Science* **166**:94–100.
- Ibanez-Escriche N, Simianer H. 2016. From the Editors: Animal breeding in the genomics era. *Animal Frontiers* **6**:4–5.
- Illumina. (2024). Bead-Based Microarray Technology. Dostupné z <https://emea.illumina.com/science/technology/microarray.html>
- Illumina. 2019. Infinium HTS Assay – Reference guide. San Diego.
- Jakubec V, Říha J, Golda J, Majzlík I. 1999. Odhad plemenné hodnoty hospodářských zvířat. VÚCHS, Rapotín.
- Jakubec V, Říha J, Majzlík I, Bjelka M. 2003. Teorie a praxe selekce hospodářských zvířat. Asociace chovatelů masných plemen, Rapotín.
- Jang S, Lourenco D, Miller S. 2022. Inclusion of sire by herd interaction effect in the genomic evaluation for weaning weight of American Angus. *Journal of Animal Science* **100**.
- Jang S, Ros-Freixedes R, Hickey JM, Chen C-Y, Herring WO, Holl J, Misztal I, Lourenco D. 2023. Multi-line ssGBLUP evaluation using preselected markers from whole-genome sequence data in pigs. *Frontiers in Genetics* **14**.
- Jibrila I, Vandenplas J, ten Napel J, Bergsma R, Veerkamp RF, Calus MPL. 2022. Impact of genomic preselection on subsequent genetic evaluations with ssGBLUP using real data from pigs. *Genetics Selection Evolution* **54**:48.
- Johnson ZB, Chewning JJ, Nugent RA. 1999. Genetic Parameters for Production Traits and Measures of Residual Feed Intake in Large White Swine 1. Page J. *Anim. Sci.* Dostupné z <https://academic.oup.com/jas/article/77/7/1679/4625502>.

- Jokubka R, Maak S, Kerziene S, Swalve HH. 2006. Association of a melanocortin 4 receptor (MC4R) polymorphism with performance traits in Lithuanian White pigs. *Journal of Animal Breeding and Genetics* **123**:17–22.
- Kabanov A, Melnikova E, Nikitin S, Somova M, Fomenko O, Volkova V, Kostyunina O, Karpushkina T, Martynova E, Trebunskikh E. 2022. Weighted Single-Step Genomic Best Linear Unbiased Prediction Method Application for Assessing Pigs on Meat Productivity and Reproduction Traits. *Animals* **12**:1693.
- Kachman SD et al. 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution* **45**:30.
- Kanis E, De Greef KH, Hiemstra A, van Arendonk JAM. 2005. Breeding for societally important traits in pigs. *Journal of Animal Science* **83**:948–957.
- Kaps M, Lamberson WR. 2004. *Biostatistics for animal science*. CABI Publishing, Oxfordshire.
- Khanal P, Maltecca C, Schwab C, Gray K, Tiezzi F. 2019. Genetic parameters of meat quality, carcass composition, and growth traits in commercial swine. *Journal of Animal Science* **97**:3669–3683. Oxford University Press.
- Khlestkina EK. 2014. Molecular markers in genetic studies and breeding. *Russian Journal of Genetics: Applied Research* **4**:236–244.
- Khodarahmi M, Kahroba H, Jafarabadi MA, Mesgari-Abbasi M, Farhangi MA. 2020. Dietary quality indices modifies the effects of melanocortin-4 receptor (MC4R) rs17782313 polymorphism on cardio-metabolic risk factors and hypothalamic hormones in obese adults. *BMC Cardiovascular Disorders* **20**:1–10. *BMC Cardiovascular Disorders*.
- Kim KS, Larsen N, Short T, Plastow G, Rothschild MF. 2000a. A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome* **11**:131–135.
- Kim KS, Larsen NJ, Rothschild MF. 2000b. Rapid communication: Linkage and physical mapping of the porcine melanocortin-4 receptor (MC4R) gene. *Journal of Animal Science* **78**:791–792.
- Kim KS, Reecy JM, Hsu WH, Anderson LL, Rothschild MF. 2004. Functional and phylogenetic analyses of a melanocortin-4 receptor mutation in domestic pigs. *Domestic Animal Endocrinology* **26**:75–86.
- Knol EF, Nielsen B, Knap PW. 2016. Genomic selection in commercial pig breeding. *Animal Frontiers* **6**:15–22.

- Kočárek E. 2008. Genetika – obecná genetika a cytogenetika, molekulární biologie, biotechnologie, genomika, 2nd edition. Scientia, Brno.
- Koivula M, Strandén I, Su G, Mäntysaari EA. 2012. Different methods to calculate genomic predictions—Comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *Journal of Dairy Science* **95**:4065–4073.
- Koopae HK, Koshkoiyeh AE. 2014. SNPs genotyping technologies and their applications in farm animals breeding programs: review. *Brazilian Archives of Biology and Technology* **57**:87–95.
- Kreft I, de Leeuw J. 1998. *Introducing Multilevel Modeling*. SAGE Publications, Ltd, London.
- Krupa E, Krupová Z, Žáková E. 2019a. Praktické postupy šlechtění prasat v národním šlechtitelském programu CzePig. Agrární komora České republiky, Prague.
- Krupa E, Schröffelová D, Žáková E, Němcová L, Krupová Z, Vrtková I. 2019b. Možnosti sběru, analýzy a správy SNP dat. *Náš chov* **79**:31–34.
- Krupa E, Žáková E, Krupová Z. 2014. Evaluation of Inbreeding and Genetic Variability of Five Pig Breeds in Czech Republic. *Asian-Australasian Journal of Animal Sciences* **28**:25–36.
- Krupová Z, Kašná E, Zavadilová L, Krupa E, Bauer J, Wolfová M. 2024. Udder, Claw, and Reproductive Health in Genomic Selection of the Czech Holstein. *Animals* **14**:864.
- Kruschke JK. 2015. *Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and Stan*, 2nd edition. Elsevier, Oxford.
- Kruuk LEB. 2004. Estimating genetic parameters in natural populations using the ‘animal model.’ *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**:873–890.
- Kruuk LEB, Hadfield JD. 2007. How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology* **20**:1890–1903.
- Kwon R, Cahyadi M, Park HB, Seo DW, Jin S, Kim SW, Choi Y Il, Kim KS, Gotoh T, Lee JH. 2015. Association of variation in the MC4R gene with meat quality traits in a commercial pig population. *Journal of the Faculty of Agriculture, Kyushu University* **60**:113–118.
- Lakens D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* **4**.

- Landgridge P. 2019. Innovation in Breeding and Biotechnology. Pages 245–284 in Serraj R, Pingali P, editors. *Agriculture & Food Systems to 2050: Global Trends, Challenges and Opportunities*. World Scientific, Singapore.
- Lazic SE. 2010. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neuroscience* **11**:5.
- Lebret B. 2008. Effects of feeding and rearing systems on growth, carcass composition and meat quality in pigs. *Animal* **2**:1548–1558.
- Legarra A, Aguilar I, Misztal I. 2009. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* **92**:4656–4663.
- Legarra A, Christensen OF, Aguilar I, Misztal I. 2014. Single Step, a general approach for genomic selection. *Livestock Science* **166**:54–65.
- Legarra A, Reverter A. 2017. Can we frame and understand cross-validation results in animal breeding? *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22**:73–80.
- Legarra A, Reverter A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution* **50**:53.
- Li X, Ye J, Han X, Qiao R, Li X, Lv G, Wang K. 2020. Whole-genome sequencing identifies potential candidate genes for reproductive traits in pigs. *Genomics* **112**:199–206.
- Li X, Zhang Z, Liu X, Chen Y. 2019. Impact of genotyping strategy on the accuracy of genomic prediction in simulated populations of purebred swine. *Animal* **13**:1804–1810.
- Lidauer M, Matilainen K, Mäntysaari E, Pitkänen T, Taskinen M, Strandén I. 2019. MiX99: Technical reference guide for MiX99 solver. Release XI/2019.
- Littell RC, Pendergast J, Natarajan R. 2000. Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* **19**:1793–1819.
- Liu Z, Goddard ME, Reinhardt F, Reents R. 2014. A single-step genomic model with direct estimation of marker effects. *Journal of Dairy Science* **97**:5833–5850.
- Lopes MS. 2016. Genomic selection for improved crossbred performance. Wageningen University.
- Lopez BI, Viterbo V, Song CW, Seo KS. 2019. Estimation of genetic parameters and accuracy of genomic prediction for production traits in duroc pigs. *Czech Journal of Animal Science* **64**:160–165. Czech Academy of Agricultural Sciences.

- Lopez BIM, Song C, Seo K. 2018. Genetic parameters and trends for production traits and their relationship with litter traits in Landrace and Yorkshire pigs. *Animal Science Journal* **89**:1381–1388. Blackwell Publishing.
- López-Romero P, Carabaño MJ. 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein–Friesian cattle. *Livestock Production Science* **82**:81–96.
- Lourenco D. 2018. Theoretical Accuracies. Dostupné z http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=lourenco_theoretical_accuracies.pdf.
- Lourenco D, Legarra A, Tsuruta S, Masuda Y, Aguilar I, Misztal I. 2020. Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. *Genes* **11**:790.
- Lourenco DAL, Fragomeni BO, Bradford HL, Menezes IR, Ferraz JBS, Aguilar I, Tsuruta S, Misztal I. 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. *Journal of Animal Breeding and Genetics* **134**:463–471.
- Lukač D. 2016. Use of different models for estimation of genetic parameters and genetic trends of performance test traits of gilts. *Thai Journal of Veterinary Medicine* **46**:49–58.
- Lyadskiy IK, Getya AA, Pochernyaev KF. 2011. Association of the Asp298Asn polymorphism in the mc4r gene with back fat thickness in pigs of the Large White breed. *Cytology and Genetics* **45**:106–109.
- Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Oxford.
- Macedo FL, Christensen OF, Astruc J-M, Aguilar I, Masuda Y, Legarra A. 2020a. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetics Selection Evolution* **52**:47.
- Macedo FL, Reverter A, Legarra A. 2020b. Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *Journal of Dairy Science* **103**:529–544.
- MacNeil MD, Nkrumah JD, Woodward BW, Northcutt SL. 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators¹. *Journal of Animal Science* **88**:517–522.

- Mancisidor B, Cruz A, Gutiérrez G, Burgos A, Morón JA, Wurzinger M, Gutiérrez JP. 2021. ssGBLUP Method Improves the Accuracy of Breeding Value Prediction in Huacaya Alpaca. *Animals* **11**:3052.
- Massender E, Brito LF, Maignel L, Oliveira HR, Jafarikia M, Baes CF, Sullivan B, Schenkel FS. 2022. Single-step genomic evaluation of milk production traits in Canadian Alpine and Saanen dairy goats. *Journal of Dairy Science* **105**:2393–2407.
- Masuda Y. 2019. Introduction to BLUPF90 suite programs. University of Georgia.
- Masuda Y, VanRaden PM, Tsuruta S, Lourenco DAL, Misztal I. 2022. Invited review: Unknown-parent groups and metafounders in single-step genomic BLUP. *Journal of Dairy Science* **105**:923–939.
- Mehrban H, Lee DH, Naserkheil M, Moradi MH, Ibáñez-Escriche N. 2019. Comparison of conventional BLUP and single-step genomic BLUP evaluations for yearling weight and carcass traits in Hanwoo beef cattle using single trait and multi-trait models. *PLOS ONE* **14**:e0223352.
- Meidtner K, Wermter AK, Hinney A, Renschmidt H, Hebebrand J, Fries R. 2006. Association of the melanocortin 4 receptor with feed intake and daily gain in F2 Mangalitsa x Piétrain pigs. *Animal Genetics* **37**:245–247.
- Melnikova E et al. 2021. Application of Genomic Data for Reliability Improvement of Pig Breeding Value Estimates. *Animals* **11**:1557.
- Merks JWM, Mathur PK, Knol EF. 2012. New phenotypes for new breeding goals in pigs. *Animal* **6**:535–543.
- Mészáros G. 2021. Genomics Boot Camp. Dostupné z <https://genomicsbootcamp.github.io/book/> (10.08.2023).
- Meuwissen T, Hayes B, Goddard M. 2016. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers* **6**:6–14.
- Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**:1819–1829.
- Meyer K. 1989. Approximate accuracy of genetic evaluation under an animal model. *Livestock Production Science* **21**:87–100.
- Miar Y et al. 2014. Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs 1. *J. Anim. Sci* **92**:2869–2884. Dostupné z <https://academic.oup.com/jas/article/92/7/2869/4702156>.

- Milan D et al. 2000. A Mutation in *PRKAG3* Associated with Excess Glycogen Content in Pig Skeletal Muscle. *Science* **288**:1248–1251.
- Missoum A. 2018. DNA Microarray and Bioinformatics Technologies: A mini-review. *Proceedings of the Nature Research Society* **2**:02010.
- Misztal I. 2008. Reliable computing in estimation of variance components. *Journal of Animal Breeding and Genetics* **125**:363–370.
- Misztal I, Legarra A. 2017. Invited review: efficient computation strategies in genomic selection. *Animal* **11**:731–736.
- Misztal I, Legarra A, Aguilar I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science* **92**:4648–4655.
- Misztal I, Lourenco D, Legarra A. 2020. Current status of genomic evaluation. *Journal of Animal Science* **98**.
- Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, Vitezica Z. 2022. Manual for BLUPF90 Family Programs. University of Georgia. Dostupné z http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf (12.02.2024).
- Misztal I, Vitezica ZG, Legarra A, Aguilar I, Swan AA. 2013. Unknown-parent groups in single-step genomic evaluation. *Journal of Animal Breeding and Genetics* **130**:252–258.
- Misztal I, Wiggans GR. 1988. Approximation of Prediction Error Variance in Large-Scale Animal Models. *Journal of Dairy Science* **71**:27–32.
- Mote BE, Rothschild MF. 2020. Modern genetic and genomic improvement of the pig. Pages 249–262 *Animal Agriculture*. Elsevier.
- Mrode RA. 2005. *Linear Models for the Prediction of Animal Breeding Values* Second Edition. CABI Publishing.
- Muñoz G, Alcázar E, Fernández A, Barragán C, Carrasco A, de Pedro E, Silió L, Sánchez JL, Rodríguez MC. 2011. Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc×Iberian crossbred pigs. *Meat Science* **88**:169–173.
- Nagy I, Sölkner J, Csató L, Farkas J, Radnóczy L. 2004. Analysis of alternative models treating herd × year effects as fixed or random. *Czech Journal of Animal Science* **49**:349–356.
- Neeteson-van Nieuwenhoven A-M, Knap P, Avendaño S. 2013. The role of sustainable commercial pig and poultry breeding for food security. *Animal Frontiers* **3**:52–57.

- Nilforooshan MA. 2022. A Note on the Conditioning of the H-1 Matrix Used in Single-Step GBLUP. *Animals* **12**:3208.
- Nilforooshan MA. 2023. Extension of the reduced animal model to single-step methods. *Journal of Animal Science* **101**.
- Nosková A, Mehrotra A, Kadri NK, Lloret-Villas A, Neuenschwander S, Hofer A, Pausch H. 2023. Comparison of two multi-trait association testing methods and sequence-based fine mapping of six additive QTL in Swiss Large White pigs. *BMC Genomics* **24**:192.
- OECD/FAO. 2022. OECD-FAO Agricultural Outlook 2022-2031. OECD Publishing, Paris.
- Ogawa S, Yazaki N, Ohnishi C, Ishii K, Uemoto Y, Satoh M. 2021. Maternal effect on body measurement and meat production traits in purebred Duroc pigs. *Journal of Animal Breeding and Genetics* **138**:237–245.
- Oh J-D, Na C-S, Park K-D. 2016. Validation of selection accuracy for the total number of piglets born in Landrace pigs using genomic selection. *Asian-Australasian Journal of Animal Sciences* **30**:149–153.
- Paneru U, Brown DJ, Moghaddar N, van der Werf JHJ. 2022. 724. Impact of the order of polynomials in random regression model on the accuracy of genetic evaluation for weight traits. Pages 2984–2987 Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Wageningen Academic Publishers, The Netherlands.
- Park HB, Carlborg Ö, Marklund S, Andersson L. 2002. Melanocortin-4 receptor (MC4R) genotypes have no major effect on fatness in a large white x wild boar intercross. *Animal Genetics* **33**:155–157.
- Park J. 2024. Genome-wide association study to reveal new candidate genes using single-step approaches for productive traits of Yorkshire pig in Korea. *Animal Bioscience* **37**:451–460.
- Patry C, Ducrocq V. 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *Journal of Dairy Science* **94**:1011–1020.
- Pérez-Cabal MA, Vazquez AI, Gianola D, Rosa GJM, Weigel KA. 2012. Accuracy of Genome-Enabled Prediction in a Dairy Cattle Population using Different Cross-Validation Layouts. *Frontiers in Genetics* **3**.
- Perneger T V. 1998. What's wrong with Bonferroni adjustments. *BMJ* **316**:1236–1238.

- Piccoli ML, Brito LF, Braccini J, Brito F V., Cardoso FF, Cobuci JA, Sargolzaei M, Schenkel FS. 2018. A comprehensive comparison between single- and two-step GBLUP methods in a simulated beef cattle population. *Canadian Journal of Animal Science* **98**:565–575.
- PigQTLdb. 2023. Pig QTLdb data summary. Dostupné z <https://www.animalgenome.org/cgi-bin/QTLdb/SS/summary> (12.07.2023).
- PigQTLdb. 2024. Selected QTL on Pig Chromosome 1. Dostupné z <https://www.animalgenome.org/cgi-bin/QTLdb/SS/chromap?QTLid=28198,125088,9798,125108,125111,9801,125609,10621,126066,10624,258447,10627,285902,18521,21339,21342,21345,125084,125093,125109,9799,125519,9802,126064,10622,139265,10625,258493,10628,18522,21340,21343,28197,125085,125107,125110,9800,125595,9803,126065,10623,216362,10626,285897,10629,21338,21341,21344&chromos=1> (13.04.2024).
- Piórkowska K, Tyra M, Rogoz M, Ropka-Molik K, Oczkiewicz M, Rózycki M. 2010. Association of the melanocortin-4 receptor (MC4R) with feed intake, growth, fatness and carcass composition in pigs raised in Poland. *Meat Science* **85**:297–301.
- Pomar C, Marcoux M, Gispert M, Font i Furnols M, Daumas G. 2009. Determining the lean content of pork carcasses. Pages 493–518 *Improving the Sensory and Nutritional Quality of Fresh Meat*. Elsevier.
- Poulsen BG, Ask B, Nielsen HM, Ostersen T, Christensen OF. 2020. Prediction of genetic merit for growth rate in pigs using animal models with indirect genetic effects and genomic information. *Genetics Selection Evolution* **52**:58.
- Pravia MI, Navajas EA, Aguilar I, Ravagnolo O. 2023. Prediction ability of an alternative multi-trait genomic evaluation for residual feed intake. *Journal of Animal Breeding and Genetics* **140**:508–518.
- Příbyl J, Bauer J, Čermák V, Pešek P, Příbylová J, Šplíchal J, Vostrá-Vydrová H, Vostrý L, Zavadilová L. 2015. Domestic estimated breeding values and genomic enhanced breeding values of bulls in comparison with their foreign genomic enhanced breeding values. *Animal* **9**:1635–1642.
- Pulkrábek J. 2005. *Chov prasat*. Profi Press, Prague.
- Purcell S, Chang C. 2023. PLINK 1.9. Dostupné z www.cog-genomics.org/plink/1.9/ (18.08.2023).
- Ramos AM et al. 2009. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. *PLoS ONE* **4**:e6524.

- Raudsepp T, Chowdhary BP. 2011. Cytogenetics and chromosome maps. Page in Rothschild MF, Ruvinsky A, editors. *The genetics of the pig*, 2nd edition. CABI, Oxfordshire.
- Robinson GK. 1991. That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science* **6**.
- Roongsitthichai A, Tummaruk P. 2014. Importance of Backfat Thickness to Reproductive Performance in Female Pigs. *The Thai Journal of Veterinary* **44**:171–178.
- Ros-Freixedes R, Reixach J, Bosch L, Tor M, Estany J. 2013. Response to selection for decreased backfat thickness at restrained intramuscular fat content in Duroc pigs¹. *Journal of Animal Science* **91**:3514–3521.
- Rothschild M et al. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proceedings of the National Academy of Sciences* **93**:201–205.
- Ruan D et al. 2021. Weighted Single-Step GWAS Identified Candidate Genes Associated with Growth Traits in a Duroc Pig Population. *Genes* **12**:117.
- Runcie D, Cheng H. 2019. Pitfalls and Remedies for Cross Validation with Multi-trait Genomic Prediction Methods. *G3 Genes|Genomes|Genetics* **9**:3727–3741.
- Salajpal K, Dikic M, Karolyi D, Janjecic Z, Juric I. 2009. The effect of MC4R polymorphism on carcass composition and meat quality traits in pigs slaughter at different live weights. *Italian Journal of Animal Science* **8**:98–100.
- Salces AJ, Seo KS, Cho KH, Kim S, Lee YC. 2006. Genetic Parameter Estimation of Carcass Traits of Duroc Predicted Using Ultrasound Scanning Modes. *Asian-Australasian Journal of Animal Sciences* **19**:1379–1383.
- Salek Ardestani S, Jafarikia M, Sargolzaei M, Sullivan B, Miar Y. 2021. Genomic Prediction of Average Daily Gain, Back-Fat Thickness, and Loin Muscle Depth Using Different Genomic Tools in Canadian Swine Populations. *Frontiers in Genetics* **12**.
- Samorè AB, Fontanesi L. 2016. Genomic selection in pigs: state of the art and perspectives. *Italian Journal of Animal Science* **15**:211–232.
- Schaeffer LR. 1991. C. R. Henderson: Contributions to Predicting Genetic Merit. *Journal of Dairy Science* **74**:4052–4066.
- Schaeffer LR. 2009. Contemporary groups are always random. Dostupné z <https://animalbiosciences.uoguelph.ca/~lrs/piksLRS/ranfix.pdf> (05.01.2023).
- Schaeffer LR. 2010. *Linear Models and Animal Breeding*. University of Guelph, Guelph.

- Schaeffer LR. 2018. Necessary changes to improve animal models. *Journal of Animal Breeding and Genetics* **135**:124–131.
- Schinckel AP, Einstein ME, Stewart TS, Schwab C, Olynkf NJ. 2010. Use of a Stochastic Model to Evaluate the Growth Performance and Profitability of Pigs from Different Litter Sizes and Parities of Dams. *The Professional Animal Scientist* **26**:547–560.
- Scholz AM, Bünger L, Kongsro J, Baulain U, Mitchell AD. 2015. Non-invasive methods for the determination of body and carcass composition in livestock: dual-energy X-ray absorptiometry, computed tomography, magnetic resonance imaging and ultrasound: invited review. *Animal* **9**:1250–1264.
- SCHP. 2010. Metodické pokyny pro kontrolu užítkovosti prasat. Svaz chovatelů prasat v Čechách a na Moravě. Dostupné z http://www.schpcm.cz/slechtění/metodiky/03_Metodika_KU.pdf (23.11.2022).
- SCHP. 2021. Ročenka 2020. Praha. Dostupné z <http://www.schpcm.cz/publikace/rocenka2020.pdf> (16.07.2023).
- Schulz K, Norda C, Eisenreich R, Mäurer H, Paulke T. 2019. Richtlinie für die Stationsprüfung auf Mastleistung, Schlachtkörperwert und Fleischbeschaffenheit beim Schwein. Bonn. Dostupné z <https://www.rind-schwein.de/services/files/brs/richtlinien/20190624%20Richtlinie%20NKP%20Deckblatt%20benSans.pdf>.
- Serão NVL, Kemp RA, Mote BE, Willson P, Harding JCS, Bishop SC, Plastow GS, Dekkers JCM. 2016. Genetic and genomic basis of antibody response to porcine reproductive and respiratory syndrome (PRRS) in gilts and sows. *Genetics Selection Evolution* **48**:51.
- Silva DA da, Costa CN, Silva AA da, Silva FF e, Lopes PS, Santos GG dos, Thompson G, Carvalheira J. 2019. Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. *Livestock Science* **220**:1–7.
- Simsek B, Firat MZ. 2011. Application of multilevel analysis in animal sciences. *Applied Mathematics and Computation* **218**:1067–1071.
- Škorput D, Špehar M, Luković Z. 2018. Connectedness between contemporary groups in Black Slavonian pig. *Livestock Science* **216**:6–8.
- Snowder GD, Fogarty NM. 2009. Composite trait selection to improve reproduction and ewe productivity: a review. *Animal Production Science* **49**:9.

- Solanes FX, Grandinson K, Rydhmer L, Stern S, Andersson K, Lundeheim N. 2004. Direct and maternal influences on the early growth, fattening performance, and carcass traits of pigs. *Livestock Production Science* **88**:199–212.
- Song H, Zhang J, Zhang Q, Ding X. 2019. Using Different Single-Step Strategies to Improve the Efficiency of Genomic Prediction on Body Measurement Traits in Pig. *Frontiers in Genetics* **9**.
- Stachowiak M, Szydlowski M, Obarzanek-Fojt M, Switonski M. 2006. An effect of a missense mutation in the porcine melanocortin-4 receptor (MC4R) gene on production traits in Polish pig breeds is doubtful. *Animal Genetics* **37**:55–57.
- Steyn Y, Lourenco DA, Chen C-Y, Valente BD, Holl J, Herring WO, Misztal I. 2021. Optimal definition of contemporary groups for crossbred pigs in a joint purebred and crossbred genetic evaluation. *Journal of Animal Science* **99**.
- Stibal J. 2016. Metodika uchování genetického zdroje zvířat. Dostupné z <http://www.genetickezdroje.cz/wp-content/uploads/2016/12/Metodika-GZ-prase-pc.pdf> (29.11.2022).
- Stock J, Bennewitz J, Hinrichs D, Wellmann R. 2020. A Review of Genomic Models for the Analysis of Livestock Crossbred Data. *Frontiers in Genetics* **11**.
- Stupka R, Šprysl M, Cítek J. 2009. *Základy chovu prasat*. PowerPrint, Prague.
- Su G, Christensen OF, Ostersen T, Henryon M, Lund MS. 2012. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS ONE* **7**:e45293.
- Szyndler-Nędza M, Eckert R, Tyra M, Zak G, Szulc K, Blicharski T. 2019. Analysis of Genetic Parameters of Carcass Traits and Daily Gain of Native Breed Pigs Raised in Poland. *Annals of Animal Science* **19**:595–604. Sciendo.
- Tang Z, Peng Z, Liu B, Fan B, Zhao S, Li X, Xu S, Li K. 2008. Effect of breed, sex and birth parity on growth, carcass and meat quality in pigs. *Frontiers of Agriculture in China* **2**:331–337.
- Tang Z, Xu J, Yin L, Yin D, Zhu M, Yu M, Li X, Zhao S, Liu X. 2019. Genome-Wide Association Study Reveals Candidate Genes for Growth Relevant Traits in Pigs. *Frontiers in Genetics* **10**.
- Taylor JF, Everett RW. 1985. Estimation of Variance Components by the Expectation-Maximization Algorithm for Restricted Maximum Likelihood in a Repeatability Model for Semen Production. *Journal of Dairy Science* **68**:2948–2953.

- Teama S. 2018. DNA Polymorphisms: DNA-Based Molecular Markers and Their Application in Medicine. Page Genetic Diversity and Disease Susceptibility. InTech.
- Theron HE, Kanfer FHJ, Rautenbach L. 2002. The effect of phantom parent groups on genetic trend estimation. Page South African Journal of Animal Science. Dostupné z <http://sasas.co.za/Sajas.html>.
- TIS SZIF. 2024. Zpráva o trhu hovězího a vepřového masa 3. 5. 2024. Praha. Dostupné z <https://www.szif.cz/cs/zpravy-o-trhu> (07.05.2024).
- Tusell L, Gilbert H, Riquet J, Mercat M-J, Legarra A, Larzul C. 2016. Pedigree and genomic evaluation of pigs using a terminal-cross model. *Genetics Selection Evolution* **48**:32.
- Ugarte E, Alenda R, Carabaño MJ. 1992. Fixed or Random Contemporary Groups in Genetic Evaluations. *Journal of Dairy Science* **75**:269–278.
- Uniporc. 2021. Regles de classification des carcasses: Sites de mesures.
- Urban T. 2006. Biometrické metody v genetice - lineární modely. Dostupné z https://user.mendelu.cz/urban/vyuka/genslech/6_genetika_slechtenu_modely_06_w.pdf (05.04.2023).
- van Bebber J, Reinsch N, Junge W, Kalm E. 1997. Accounting for herd, year and season effects in genetic evaluations of dairy cattle: a review. *Livestock Production Science* **51**:191–203.
- Van Den Broeke A, Aluwé M, Tuytens FAM, Ampe B, Vanhaecke L, Wauters J, Janssens S, Coussé A, Buys N, Millet S. 2015. An intervention study demonstrates effects of MC4R genotype on boar taint and performances of growing–finishing pigs. *Journal of Animal Science* **93**:934–943.
- Van Den Maagdenberg K, Stinckens A, Claeys E, Seynaeve M, Clinquart A, Georges M, Buys N, De Smet S. 2007. The Asp298Asn missense mutation in the porcine melanocortin-4 receptor (MC4R) gene can be used to affect growth and carcass traits without an effect on meat quality. *Animal* **1**:1089–1098.
- Van Eenennaam AL, Weigel KA, Young AE, Cleveland MA, Dekkers JCM. 2014. Applied Animal Genomics: Results from the Field. *Annual Review of Animal Biosciences* **2**:105–139.
- van Laack RLJM, Smulders FJM. 1992. On the assessment of water-holding capacity of hot- vs cold-boned pork. *Meat Science* **32**:139–147.
- Van Laere A-S et al. 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**:832–836.

- Van Vleck LD. 1987. Contemporary Groups for Genetic Evaluations. *Journal of Dairy Science* **70**:2456–2464.
- Vandenplas J, Veerkamp RF, Calus MPL, Lidauer MH, Strandén I, Taskinen M, Schrauf M, ten Napel J. 2022. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs. Pages 1498–1501 *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Wageningen Academic Publishers, The Netherlands.
- VanRaden PM. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* **91**:4414–4423.
- VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* **92**:16–24.
- Verbyla AP. 2019. A note on model selection using information criteria for general linear models estimated using REML. *Australian & New Zealand Journal of Statistics* **61**:39–50.
- Veroneze R, Lopes PS, Guimarães SEF, Silva FF, Lopes MS, Harlizius B, Knol EF. 2013. Linkage disequilibrium and haplotype block structure in six commercial pig lines. *Journal of Animal Science* **91**:3493–3501.
- Visscher PM, Goddard ME. 1993. Fixed and Random Contemporary Groups. *Journal of Dairy Science* **76**:1444–1454.
- Vitezica ZG, Aguilar I, Misztal I, Legarra A. 2011. Bias in genomic predictions for populations under selection. *Genetics Research* **93**:357–366.
- Wakchaure R, Ganguly S. 2015. Marker Assisted Selection (MAS) in Animal Breeding: A Review. *Journal of Drug Metabolism & Toxicology* **06**.
- Walstra P, Merkus GS. 1996. Procedure for assessment of the lean meat percentage as a consequence of the new EU reference dissection method in pig carcass classification : based on discussion in the EU Management Committee on Pig Meat and based on discussions with dissection experts during a meeting on May 18-19, 1994 at Zeist, NL. ID-DLO.
- Wang B et al. 2019. Association of twelve candidate gene polymorphisms with the intramuscular fat content and average backfat thickness of chinese suhuai pigs. *Animals* **9**. MDPI AG.
- Whittemore C. 2006. Development and improvement of pigs by genetic selection. Pages 184–261 in Kyriazakis I, Whittemore CT, editors. *Whittemore’s Science and Practice of Pig Production*, 3rd edition. Blackwell Publishing, Oxford.

- Whittemore CT, Kyriazakis I. 2007. Growth and body composition changes in pigs. Pages 65–104 in Whittemore CT, Kyriazakis I, editors. *Whittemore's Science and Practice of Pig Production*. Blackwell Publishing, Hoboken, NJ.
- Wientjes YCJ, Bijma P, van den Heuvel J, Zwaan BJ, Vitezica ZG, Calus MPL. 2023. The long-term effects of genomic selection: 2. Changes in allele frequencies of causal loci and new mutations. DOI: 10.1093/genetics/iyad141.
- Willham RL. 1972. The Role of Maternal Effects in Animal Breeding: III. Biometrical Aspects of Maternal Effects in Animals. *Journal of Animal Science* **35**:1288–1293.
- Wolf J, Wolfová M. 2012. Impact of genetic groups and herd-year-season fixed/random on genetic parameter estimates from large data sets in pigs. Page RESEARCH IN PIG BREEDING.
- Wright AF. 2005. Genetic Variation: Polymorphisms and Mutations. Page eLS. Wiley.
- Xu S. 2022. *Quantitative Genetics*. Springer International Publishing, Cham.
- Yang W, Kang X, Yang Q, Lin Y, Fang M. 2013. Review on the development of genotyping methods for assessing farm animal diversity. *Journal of Animal Science and Biotechnology* **4**:2.
- Žáková E, Krupa E, Krupová Z. 2022. Metodika k certifikaci: Hodnocení příbuznosti prasat prostřednictvím SNP dat. VÚŽV, Praha Uhřetěves.
- Žáková E, Stibal J. 2021. Plemenné standardy a chovné cíle pro plemena prasat v plemenné knize. Praha.
- Zhou L, Wang Y, Yu J, Yang H, Kang H, Zhang S, Wang C, Liu J. 2019. Improving genomic prediction for two Yorkshire populations with a limited size using the single-step method. *Animal Genetics* **50**:391–394.
- Zhou Z-Y, Li A, Otecko NO, Liu Y-H, Irwin DM, Wang L, Adeola AC, Zhang J, Xie H-B, Zhang Y-P. 2017. PigVar: a database of pig variations and positive selection signatures. *Database* **2017**.
- Zotti E, Resmini FA, Schutz LG, Volz N, Milani RP, Bridi AM, Alfieri AA, da Silva CA. 2017. Impact of piglet birthweight and sow parity on mortality rates, growth performance, and carcass traits in pigs. *Revista Brasileira de Zootecnia* **46**:856–862. Sociedade Brasileira de Zootecnia.

8. Seznam použitých zkratek

- ADFI** = průměrný denní příjem krmiva (angl. average daily feed intake)
- ADG** = průměrný denní přírůstek (angl. average daily gain)
- ADL** = *ad libitum* krmná dávka
- AI** = algoritmus průměrné informace pro odhad maximální věrohodnosti (angl. average information)
- AIC** = Akaikovo informační kritérium
- AM** = animal model
- ANOVA** = analýza rozptylu (angl. analysis of variance)
- BF** = výška hřbetního tuku (angl. backfat thickness)
- BIC** = Bayesovo/Schwarzovo informační kritérium
- BLUP** = nejlepší lineární nevychýlená předpověď (angl. best linear unbiased prediction)
- BO** = bílé otcovské
- BW** = živá hmotnost (angl. body weight)
- CG** = skupina zvířat se společnými podmínkami prostředí (angl. contemporary group)
- CPH** = celková plemenná hodnota (selekční index)
- CW** = jatečná hmotnost, hmotnost JUT
- DF** = algoritmus bez derivace pro odhad maximální věrohodnosti (angl. derivative-free)
- DL** = ztráta vody odkapem (angl. drip loss)
- DU** = duroc
- EBV** = předpovězená/odhadnutá plemenná hodnota (angl. estimated breeding value)
- EM** = algoritmus očekávání-maximalizace pro odhad maximální věrohodnosti (angl. expectation-maximization)
- EPM** = efekt trvalého prostředí matky
- GBLUP** = genomická BLUP
- GEV** = genomická plemenná hodnota (angl. genomic estimated/enhanced breeding value)
- GLM** = obecný lineární model (angl. general linear model)
- GLMM** = zobecněný lineární smíšený model (angl. generalized linear mixed model)
- HA** = hampshire
- JUT** = jatečně upravené tělo
- LA** = landrase (česká populace jako CLA)
- LC** = hlavní masité části (angl. lean cuts)
- LMP** = podíl libové svaloviny, zmasilost (angl. lean meat percentage)
- LR** = lineární regrese
- LSE** = odhady metody nejmenších čtverců (angl. least squares estimators)
- LSM** = průměry nejmenších čtverců (angl. least squares means)
- LW** = bílé ušlechtilé (angl. large white; české mateřské linie CLW; otcovské linie jako LWS)
- MA** = plocha svalu *longissimus thoracis et lumborum* (angl. loin muscle area)
- MAS** = markery asistovaná selekce
- MC4R** = melanokortin 4 receptor
- MD** = výška svalu *longissimus thoracis et lumborum* (angl. loin muscle depth)
- ML/MLE** = metoda maximální věrohodnosti (angl. maximum likelihood) / její odhady (angl. estimators)
- MME** = rovnice smíšeného modelu (angl. mixed model equations)
- ms/tsGBLUP** = víceřadová/dvouřadová GBLUP (angl. multi-step/two-step GBLUP)
- MVN** = vícerozměrné normální rozdělení (angl. multivariate normal distribution)
- OPH** = odhadnutá plemenná hodnota (vizte EBV)
- PCA** = analýza hlavních komponent (angl. principal component analysis)

PCG = metoda konjugovaných gradientů (angl. preconditioned conjugate gradients)
PCR = polymerázová řetězová reakce (angl. polymerase chain reaction)
PEV = rozptyl chyby předpovědi (angl. prediction error variance)
PE = pietrain
PrA = prediktivní schopnost, prediktivita (angl. predictive ability)
QTL = lokusy kvantitativních vlastností (angl. quantitative trait loci)
REML = restringovaná/reziduální maximální věrohodnost (angl. restricted maximum likelihood)
RES = restringovaná krmná dávka
RFLP = polymorfismus délky restričních fragmentů (angl. restriction fragment length polymorphism)
SNP = jednonukleotidový polymorfismus (angl. single nucleotide polymorphism)
SRO = stádo-rok-období
ssGBLUP = jedнокroková GBLUP (angl. single-step GBLUP)
TBV = skutečná plemenná hodnota (angl. true breeding value)
UPG = skupina neznámého rodiče, genetická skupina (angl. unknown parent group)
VIF = faktor zvětšení rozptylu (angl. variance inflation factor)
WssGBLUP = vážená ssGBLUP (angl. weighted ssGBLUP)
YAMS = Yet Another Mixed-Model Solver
ZP = dvoubodová metoda (něm. Zwei-Punkt)

Jazykové poznámky:

- 1) Zkratky až na výjimky vycházejí z angličtiny ze dvou důvodů. Prvním a hlavním je zachování návaznosti na původní literaturu a matematicko-statistická odvození, popř. využití mezinárodně unifikovaných zkratk jako u názvů plemen (převzaté od CAAB a NAAB). Druhým je skutečnost, že se v textu objevují pojmy (zejména metody), které přesný český ekvivalent nebo univerzálně používanou českou zkratku nemají. Proto je například plemenná hodnota označena jako EBV namísto OPH (byť tento pojem také není úplně přesný kvůli absenci předpovědi).
- 2) Některá odborná slovní spojení jsou záměrně vyjádřena jako přívlastek postupně rozvíjející namísto složeného přídavného jména, které bylo v minulosti upřednostňováno. Příkladem je aditivní genetický efekt. Pokud by se jednalo o složené přídavné jméno, zřejmě by se dnes už měl použít tvar aditivněgenetický efekt (nikoliv dříve používaný aditivně genetický). Avšak z pravidel pravopisu v současnosti není jasné, zda se tento případ už s jistotou dá považovat za ustálené spojení (podobně jako třeba molekulárněgenetické metody).
- 3) Latinské pojmy se v práci vyskytují výjimečně, ovšem vzhledem k praktickému výstupu je vhodné zmínit, že odborné označení kotlety se vychází z označení podle Kauffmana et al. (1990), tedy *musculus longissimus thoracis et lumborum*.
- 4) Snahou práce bylo ze statistického hlediska rozlišit pojmy „faktor“ a „efekt“. Faktorem se rozumí vysvětlující (nezávislá) proměnná, kdežto efektem se rozumí vliv na vysvětlovanou (závisle) proměnnou při změně úrovně faktoru, popř. odhadnutý regresní koeficient. Výraz „efekt“ v této práci převažuje z důvodu zaměření na smíšené modely (v jejichž terminologii je běžnější než „faktor“) nebo z důvodu kontextu testování statistické významnosti (která je založena na porovnání rozdílů mezi úrovněmi faktoru, tedy mezi efekty).
- 5) Zkratka GLM je v této práci použita pro obecný lineární model, aby byla zachována návaznost na použité procedury v programu SAS. Mnozí autoři však používají tuto zkratku pro zobecněný lineární model, a to i s odůvodněním, že obecné modely jako regrese a ANOVA jsou pouze zvláštním případem zobecněných.
- 6) BLUP je skloňován podle rodu ženského (podle „předpovědi“), kdežto BLUP-AM podle rodu mužského (podle „modelu“).

9. Seznam příloh

Všechny přílohy jsou dostupné online na adrese <https://figshare.com/s/b5f9f34b61f4a3d021d4>

Příloha 1: výběr základního BLUP animal modelu vč. genetických parametrů (kapitola 5.2.1. a 5.2.2.)

Příloha 2: srovnání BLUP a ssGBLUP (kapitola 5.2.3. a 5.2.4.)

Příloha 3: výsledky z ssGWAS (kapitola 5.2.5.)