

**Česká zemědělská univerzita v Praze**

**Fakulta agrobiologie, potravinových a přírodních zdrojů**

**Katedra genetiky a šlechtění**



**Genetická analýza SNP lokusů u dojeného skotu  
v návaznosti na genomickou selekci**

**Diplomová práce**

**Autor práce: Bc. Petr Pešek**

**Vedoucí práce: doc. Ing. Luboš Vostrý, Ph.D.**

**2014 ČZU v Praze**

## **Čestné prohlášení**

Prohlašuji, že svou diplomovou práci " Genetická analýza SNP lokusů u dojeného skotu v návaznosti na genomickou selekci" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne

---

**Poděkování:**

Tímto bych rád poděkoval rodině a přítelkyni za podporu při psaní této práce a dále doc. Ing. Luboši Vostrému, Ph.D. za vysoce odborné vedení této diplomové práce a prof. Ing. Josefu Příbylovi, DrSc. za cenné připomínky a náměty.

# Genetická analýza SNP lokusů u dojeného skotu v návaznosti na genomickou selekci

## Souhrn

Cílem předkládané práce je analýza a výběr nejvýznamnějších SNP lokusů pro vybrané užitkové vlastnosti v návaznosti na předpověď genomické plemenné hodnoty v populaci holštýnského skotu v ČR. Hypotéza: Analýza a kvantifikace nejvýznamnějších lokusů povede k přesnějšímu stanovení genetického založení jedinců dojeného skotu v České republice.

Vstupní soubor s plemennými hodnotami (PH) poskytla Českomoravská společnost chovatelů, a. s. a vstupní soubor s genotypy býků poskytl Svaz chovatelů holštýnského skotu, o. s. Soubor s PH obsahoval předpovědi pro celkem 131 312 jedinců. Analýza byla provedena pro dojivost, procentický obsah bílkovin a tuku v mléce.

Pro předpověď regresních koeficientů jednotlivých bodových mutací v DNA (SNP – Single nucleotide poly morphism) byla použita metoda RR-BLUP. Do předpovědi byly zahrnuty pouze ty SNP, které neměly četnost ani jedné z alel nižší než 0,05, a ty, které měly neznámý genotyp nejvýše u 10 % býků. Jako závisle proměnná vstupovaly do předpovědi regresních koeficientů odregresované PH býků (DRP) se spolehlivostí předpovědi pro danou vlastnost 0,6 a vyšší. Předpovězené regresní koeficienty byly statisticky vyhodnoceny a byly popsány lokusy s nejvyššími absolutními hodnotami předpovězených regresních koeficientů. Regresní koeficienty byly použity pro výpočet genetického rozptylu jednotlivých lokusů a pro předpověď přímé genetické hodnoty (PGH) a genomické plemenné hodnoty (GEPH), které byly porovnány se vstupními DRP a PH.

V případě dojivosti byly regresní koeficienty předpovězeny pro 40 890 SNP. Průměrná absolutní hodnota předpovězeného regresního koeficientu činila 0,506 kg se směrodatnou odchylkou 0,409 kg. Nejvyšší absolutní hodnotu předpovězeného regresního koeficientu (7,472 kg) měl SNP ARSBFGLNGS4939 na chromozomu 14 uvnitř genu DGAT1. Vysoké hodnoty regresních koeficientů měly i další SNP v blízkosti tohoto genu. Genetický rozptyl činil v průměru 0,170 se směrodatnou odchylkou 0,384.

U procentického obsahu bílkovin byl regresní koeficient předpovězen pro 40 925 SNP. Průměrná absolutní hodnota předpovězeného regresního koeficientu pro procentický obsah bílkovin činila  $1,206 \times 10^{-4} \%$  se směrodatnou odchylkou  $1,008 \times 10^{-4} \%$ . Nejvyšší absolutní hodnotu předpovězeného regresního koeficientu ( $19,090 \times 10^{-4} \%$ ) mělo opět SNP

uvnitř genu DGAT1 a SNP v jeho okolí. Vysoké hodnoty předpovězeného regresního koeficientu měly dále SNP uvnitř a v blízkosti genu GHR na chromozomu 20. Genetický rozptyl vysvětlený jedním SNP činil v průměru  $99,64 \times 10^{-10}$  a směrodatná odchylka genetického rozptylu  $2,41 \times 10^{-8}$ .

Průměrná absolutní hodnota předpovězeného regresního koeficientu pro procentický obsah tuku v mléce činila  $2,708 \times 10^{-4} \%$ , směrodatná odchylka  $2,575 \times 10^{-4} \%$ . I u procentického obsahu tuku měl nejvyšší předpovězenou hodnotu regresního koeficientu ( $98,070 \times 10^{-4} \%$ ) SNP ARSBFGLNGS4939 uvnitř genu DGAT1, stejně jako u předchozích vlastností. Průměrný genetický rozptyl pro procentický obsah tuku činil  $57,268 \times 10^{-9}$  se směrodatnou odchylkou  $428,694 \times 10^{-9}$ . U procentického obsahu tuku byl regresní koeficient předpovězen pro 40 911 SNP.

O vysoké přesnosti předpovědi regresních koeficientů svědčí vysoké korelace mezi vstupními DRP a PGH, které byly předpovězeny s použitím regresních koeficientů. Korelace činily 0,951, 0,975 a 0,974 u dojivosti, procentického obsahu bílkovin a tuku. GEPH byly předpovězeny kombinací PH a PGH podle indexu. Korelace mezi vstupní PH býků a předpovězenou GEPH byly 0,969, 0,985, 0,984 u dojivosti a procentického obsahu bílkovin a tuku. Vysoké korelace mezi vstupními PH a předpovězenými GEPH u prověřených býků svědčí o možnosti použití regresních koeficientů předpovězených na referenční populaci prověřených býků pro přesnější předpověď PH mladých zvířat oproti rodokmenové PH.

**Klíčová slova:** genomická plemenná hodnota, SNP, hřebenová regrese, dojený skot, lokus

# Genetic analysis of SNP loci in dairy cattle in relation to the genomic selection

## Summary

The aim of this study is analysis and identification of the most important SNP loci for selected traits in relation to prediction of genomic breeding value in population of Holstein cattle in Czech Republic. Hypothesis: Analysis and quantification of the most important loci will lead to more accurate prediction of individual genetic merit in population of dairy cattle in Czech Republic.

Input file with breeding values (EBV) provided Českomoravská společnost chovatelů, a. s. and input file with bull genotypes provided Svaz chovatelů holštýnského skotu, o. s. The analysis was performed for milk yield and milk protein and fat content.

For prediction of single nucleotide polymorphisms (SNP) regression coefficients was employed RR-BLUP method. Into prediction of regression coefficients were included only SNP with minor allele frequency at least 0.05 and SNPs with maximal frequency of missing genotypes across bulls 10%. As dependent variable into prediction of SNP regression coefficients were used deregressed breeding values (DRP) of bulls with reliability of EBV prediction at least 0.6. Predicted regression coefficients underwent statistical analysis and were identified loci with the highest absolute values of predicted regression coefficients. Using predicted regression coefficient it were computed genetic variances and predicted direct genetic values (DGV) and genomic breeding values (GEBV) and these were compared with input DRP and EBV.

For milk yield were predicted regression coefficients for 40,890 SNP. Mean absolute value of predicted regression coefficient was 0.506 kg with standard deviation 0.409 kg. The highest value of predicted regression coefficient had SNP ARSBFGLNGS4939 on chromosome 14 within DGAT1 gene. High value of predicted regression coefficients had other SNPs in linkage disequilibrium with this gene. Mean genetic variance 0.170 with standard deviation 0.384.

It was predicted 40,890 regression coefficients form milk protein content. Mean absolute value of predicted regression coefficients was  $1.206 \times 10^{-4}\%$  with standard deviation  $1.008 \times 10^{-4}\%$ . The highest absolute value of regression coefficient ( $19.090 \times 10^{-4}\%$ ) had the SNP within the DGAT1 gene and other SNPs in linkage disequilibrium with this gene

and SNP within GHR gene on chromosome 20 and SNPs in close proximity of GHR. Genetic variance explained in average by one SNP was  $99.64 \times 10^{-10}$ , standard deviation of genetic variance was  $2.41 \times 10^{-8}$ .

Mean absolute value of predicted regression coefficient for milk fat content was  $2.708 \times 10^{-4}\%$ , standard deviation  $2.575 \times 10^{-4}\%$ . Regression coefficients were predicted for 40,911 SNP. As in previous traits had the highest absolute value regression coefficient ( $98.070 \times 10^{-4}\%$ ) SNP ARSBFGLNGS4939. Mean genetic variance was  $57.268 \times 10^{-9}$  with standard deviation  $428.694 \times 10^{-9}$ .

High correlation between input DRP and predicted DGV using regression coefficients indicate high accuracy of SNP regression coefficients prediction. Correlations were 0.951, 0.975 and 0.974 in milk yield, milk protein and fat content. GEBV were predicted by combination of EBV and DGV using selection index. Correlations between input EBV and predicted GEBV were 0.969, 0.985 and 0.984 in milk yield, milk protein and fat content. High correlation between input EBV and predicted GEBV in proven bulls indicated possibility of using regression coefficients predicted in reference population of proven bulls for more accurate prediction of young animals breeding values.

**Key words:** genomic breeding value, SNP, ridge regression, dairy cattle, locus

# Obsah

1	Úvod .....	1
2	Vědecká hypotéza a cíle práce .....	2
3	Přehled literatury .....	3
3.1	Genetické hodnocení jedinců .....	3
3.1.1	Plemenná hodnota .....	3
3.1.2	DYD, YD – Odchyly užitkovosti dcer, odchyly vlastní užitkovosti .....	4
3.1.3	DRP – deregressed proof .....	6
3.2	Efektivní počet dcer .....	7
3.3	Mezinárodní hodnocení býků.....	8
3.3.1	MACE.....	9
3.4	Selekce s využitím markerů – MAS.....	11
3.5	Genotypování jedinců .....	13
3.6	Lokusy ovlivňující zkoumané vlastnosti.....	14
3.7	Metody předpovědi genomické plemenné hodnoty .....	15
4	Materiál a metody.....	18
4.1	Vstupní soubory .....	18
4.1.1	Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi.....	18
4.1.2	Výběr SNP pro předpověď regresních koeficientů.....	18
4.2	Předpověď regresních koeficientů .....	19
4.2.1	Řešení soustavy rovnic pomocí Preconditioned conjugate gradient.....	21
4.3	Výpočet genetických rozptylů a heterozygotností v jednotlivých lokusech...22	
4.4	Předpověď PGH a GEPH.....	23
5	Výsledky.....	25
5.1	Dojivost .....	25
5.1.1	Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi.....	25
5.1.2	Výběr SNP pro předpověď regresních koeficientů.....	25
5.1.3	Předpověď regresních koeficientů .....	28
5.1.4	Genetický rozptyl lokusů .....	32
5.1.5	Předpověď PGH a GEPH.....	34
5.2	Procentický obsah bílkovin .....	35



5.2.1	Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi.....	35
5.2.2	Výběr SNP pro předpověď regresních koeficientů.....	35
5.2.3	Předpověď regresních koeficientů .....	36
5.2.4	Genetický rozptyl lokusů .....	41
5.2.5	Předpověď PGH a GEPH.....	43
5.3	Procentický obsah tuku .....	44
5.3.1	Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi.....	44
5.3.2	Výběr SNP pro předpověď regresních koeficientů.....	44
5.3.3	Předpověď regresních koeficientů .....	45
5.3.4	Genetický rozptyl lokusů .....	49
5.3.5	Předpověď PGH a GEPH.....	51
6	Diskuse .....	52
7	Závěr.....	56
8	Seznam literatury.....	57
9	Seznam použitých zkratk a symbolů .....	64

# 1 Úvod

Genomická selekce je jedním z nejdiskutovanějších témat v oblasti šlechtění skotu a ostatních hospodářských zvířat v posledních letech. Genotypování jedinců pro tisíce až statisíce bodových mutací napříč celým jejich genomem se stává cenově stále dostupnější a statistické metody zahrnující informace z genetických čipů stále propracovanější. Vyspělé země v současné době zavádějí, nebo už zavedly předpověď genomické plemenné hodnoty do praxe.

Předpověď genomické plemenné hodnoty mladých zvířat bez vlastní užitkovosti nebo potomků je spolehlivější než předpověď pouze rodokmenové plemenné hodnoty, takže je možné selekci provádět již v raném věku s vyšší spolehlivostí předpovědi PH, je možno omezit testaci zvířat a dosáhnout tak nejen peněžních úspor, ale také vyššího genetického zisku díky zkrácení generačního intervalu.

K tomu, aby bylo možno přesně předpovídat genomickou plemennou hodnotu, je potřeba u některých metod nejdříve předpovědět regresní koeficienty jednotlivých SNP u referenční populace. Tyto regresní koeficienty jsou pak následně použitelné pro předpověď GEPH mladých zvířat. U jiných metod mohou předpovězené regresní koeficienty sloužit jako váha konkrétního SNP, jako například při tvorbě vážené genomické matice příbuznosti.

Velikost regresních koeficientů může rovněž sloužit výrobcům SNP čipů pro výběr SNP, jejichž genotypy budou těmito čipy „čteny“.

Vzhledem k velikosti a objemu dat k tomu potřebných je toto problém nejen teoretických znalostí, ale také výpočetní.

Práce byla vypracována v rámci řešení výzkumného projektu MZE QI111A167, „Genomická selekce u dojeného skotu“, řešeného ve Výzkumném ústavu živočišné výroby, v. v. i. Soubor s genotypy býků poskytl Svaz chovatelů holštýnského skotu, o. s., soubor s plemennými hodnotami býků poskytla Českomoravská společnost chovatelů, a. s.

## **2 Vědecká hypotéza a cíle práce**

Cílem práce je analýza a výběr nejvýznamnějších SNP lokusů pro vybrané užitkové vlastnosti v návaznosti na předpověď genomické plemenné hodnoty v populaci holštýnského skotu v ČR. Vybranými vlastnostmi byla dojivost, procentický obsah bílkovin a tuku v mléce.

Hypotéza: Analýza a kvantifikace nejvýznamnějších lokusů povede k přesnějšímu stanovení genetického založení jedinců dojeného skotu v České republice.

### 3 Přehled literatury

#### 3.1 Genetické hodnocení jedinců

##### 3.1.1 Plemenná hodnota

Základním genetickým parametrem jedince je plemenná hodnota, kterou Jakubec *et al.* (2010) definují následovně: „Plemenná hodnota jedince, který je pářen s určitým počtem jedinců vybraných náhodně z populace, se rovná dvojnásobku průměrné odchylky vlastnosti jeho potomstva od populačního průměru, tj. dvojnásobek průměrného genového efektu“.

Pro předpověď plemenné hodnoty (PH) se dnes téměř celosvětově používá metoda BLUP (best – nejlepší, linear – lineární, unbiased – nevychýlená, prediction – předpověď) (Henderson, 1975), která umožňuje současnou předpověď PH i odhad fixních efektů (Jakubec *et al.*, 1999).

Obecný lineární model používaný pro předpověď PH má následující podobu (1) (Mrode, 2005):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \quad (1)$$

Kde:

$\mathbf{X}$  je incidenční matice přiřazující jednotlivá pozorování k fixním efektům,

$\mathbf{Z}$  je incidenční matice přiřazující pozorování k náhodným efektům,

$\mathbf{b}$  je vektor odhadovaných fixních efektů,

$\mathbf{a}$  je vektor předpovídaných náhodných efektů,

$\mathbf{y}$  je vektor naměřených užitkovostí,

$\mathbf{e}$  je vektor residuálních efektů.

Soustava rovnic pro víceznakový model má následující podobu (2) (Mrode, 2005):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{M}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (2)$$

Kde:

$\mathbf{R}$  je diagonální matice residuálních rozptylů,

$\mathbf{M} = \mathbf{G} \otimes \mathbf{A}$ ,

$\mathbf{G}$  je variančně-kovarianční matice náhodných (genetických) efektů,

$\mathbf{A}$  je matice příbuznosti,

$\otimes$  značí přímé násobení matic,

$\hat{\mathbf{b}}$  je vektor odhadovaných fixních efektů,

$\hat{\mathbf{a}}$  je vektor předpovídaných náhodných efektů.

V případě jednoznakového modelu je možné soustavu rovnic zjednodušit vykrácením residuálních rozptylů. Soustava rovnic metody BLUP pro jednoznakový model má následující podobu (3) (Mrode, 2005):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\boldsymbol{\alpha} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (3)$$

Kde:

$\boldsymbol{\alpha}$  je poměr residuálního a aditivně genetického rozptylu ( $\boldsymbol{\alpha} = \sigma_e^2/\sigma_a^2$ ).

V mezinárodním měřítku není možné jedince porovnávat přímo podle plemenných hodnot předpovězených v jednotlivých zemích. Jednotlivé země se značně liší v chovatelských podmínkách, používaných modelech a efektech do nich zahrnutých. Pro potřeby mezinárodního porovnání je podle Szydy *et al.* (2008) vhodnější používat odchylky ve vlastní užitkovosti (YD), v případě býků jsou to odchylky v užitkovostech dcer (DYD) nebo DRP, které jsou počítány z již předpovězených plemenných hodnot a jsou používány při přepočtu PH mezi zeměmi pomocí metody MACE. Sigurdsson *et Banos* (1995) uvádějí, že hodnoty DYD a DRP nabývají podobných hodnot, ale DRP je vhodnější pro mezinárodní hodnocení, jelikož do výpočtu DRP nevstupuje tak vysoké množství dat, jako je tomu při výpočtu DYD a YD.

### 3.1.2 DYD, YD – Odchylky užitkovosti dcer, odchylky vlastní užitkovosti

VanRaden *et Wiggans* (1991) uvádějí, že DYD je průměrná užitkovost dcer očištěná od efektů stáda-roku-období otelení, věku, interakce otec-stádo, plemenných hodnot matek a trvalých prostředí dcer. Szyda *et al.* (2008) uvádějí, že na rozdíl od PH jsou DYD méně závislé na modelu, který je použit pro předpověď PH. První vztah pro výpočet DYD uvedli

VanRaden *et al.* (1991). Liu *et al.* (2004) rozšířili vztah pro potřeby RR-TDM. Szyda *et al.* (2008) shrnují podstatu odchylek užítkovosti do následujících vztahů (4,5,6):

$$\mathbf{y} = \boldsymbol{\beta} + \mathbf{p} + \frac{1}{2}\mathbf{a}_d + \frac{1}{2}\mathbf{a}_s + \mathbf{ms} + \mathbf{e} \quad (4)$$

Kde:

$\mathbf{y}$  je užítkovost jedince,  
 $\boldsymbol{\beta}$  je vektor fixních efektů,  
 $\mathbf{p}$  je efekt trvalého prostředí,  
 $\mathbf{a}_d$  je aditivní efekt matky jedince,  
 $\mathbf{a}_s$  je aditivní efekt otce jedince,  
 $\mathbf{ms}$  je náhodný mendelistický výběr,  
 $\mathbf{e}$  je reziduální chyba.

$$\mathbf{YD} = \mathbf{y} - (\hat{\boldsymbol{\beta}} + \hat{\mathbf{p}} + \frac{1}{2}\hat{\mathbf{a}}_d) \quad (5)$$

Kde:

$\mathbf{YD}$  je odhad odchylky užítkovosti jedince, ostatní jsou odhady a předpověď výše uvedených efektů.

$$\mathbf{DYD} = \frac{\sum \mathbf{YD}_i \mathbf{w}_i}{\sum \mathbf{w}_i} \quad (6)$$

Kde:

$\mathbf{DYD}$  je průměrná odchylka užítkovosti dcer,  
 $\mathbf{YD}_i$  je odchylka užítkovosti i-té dcery,  
 $\mathbf{w}_i$  je váha i-té dcery.

Pro studie zabývající se mapováním genů je použití DYD výhodnější než použití jiných vstupních údajů, protože na rozdíl od PH zahrnuje nejen aditivní efekt, ale také efekt reziduální. U vlastností, které nejsou determinovány pouze polygeny, ale také majorgeny, může být jejich efekt zahrnut v reziduální chybě (Szyda *et al.*, 2008).

### 3.1.3 DRP – deregressed proof

Při mezinárodním hodnocení skotu nejsou vždy k dispozici naměřené užitkovosti pro výpočet DYD nebo YD, proto se častěji používají DRP, které jsou počítány, jak již bylo uvedeno dříve, z již předpovězených PH.

Pro výpočet DRP je možné použít několik postupů, asi nejpřesnější je odregresování PH. Výpočet DRP odregresováním PH je v podstatě opačný postup než je tomu při řešení soustavy rovnic při předpovědi PH. Jsou popsány postupy pro jednoznakové i víceznakové modely, kdy je navíc při odregresování zahrnuta korelace mezi vlastnostmi.

Zjednodušeně lze vztah pro odregresování v maticové podobě zapsat následovně (7) (upraveno dle Ducrocq *et* Liu, 2009; VanRaden *et* Sullivan, 2010):

$$\begin{bmatrix} \Psi_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \Psi_m \end{bmatrix} + \alpha \mathbf{A}^{-1} \begin{bmatrix} \vdots \\ \hat{\mathbf{a}}_i \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \Psi_i \mathbf{DRP}_i \\ \vdots \end{bmatrix} \quad (7)$$

Kde:

$\Psi_i$  je efektivní počet dcer i-tého býka,

$\alpha$  je poměr reziduálního a aditivně genetického rozptylu,

$\mathbf{A}^{-1}$  je inverze matice příbuznosti,

$\hat{\mathbf{a}}_i$  je předpovězená plemenná hodnota i-tého býka,

$\mathbf{DRP}_i$  je neznámá odchylka užitkovosti i-tého býka.

Na pravé straně soustavy po pronásobení získáme součty odchylek dcer jednotlivých býků. Vynásobením pravé strany soustavy inverzí matice obsahující počty dcer býků získáme DRP (8):

$$\begin{bmatrix} \vdots \\ \mathbf{DRP}_i \\ \vdots \end{bmatrix} = \begin{bmatrix} \Psi_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \Psi_m \end{bmatrix}^{-1} \begin{bmatrix} \vdots \\ \Psi_i \mathbf{DRP}_i \\ \vdots \end{bmatrix} \quad (8)$$

V této podobě jsou do soustavy zahrnuti pouze jedinci se známou plemennou hodnotou a spolehlivostí předpovědi PH. Pokud by byli v matici příbuznosti zahrnuti i např. předci bez plemenné hodnoty, měla by soustava složitější podobu. Odregresování složitějších

soustav, do kterých vstupují nejen jedinci se známou PH, ale také jejich předci, případně genetické skupiny, uvádí například Jairath *et al.* (1998).

Nejjednodušším způsobem výpočtu DRP (nebo spíše jeho aproximace) je vydělení PH jedince spolehlivostí předpovědi PH. Při tomto způsobu se odchylka užitkovosti v podstatě počítá, jako by PH byla předpovídána pouze na základě vlastní užitkovosti, nebo na základě velkého množství potomků. V případě, že má jedinec PH předpovězenou na základě užitkovosti velkého množství potomků, je tento postup výpočtu DRP poměrně přesný, protože ostatní zdroje informací jsou v podstatě zanedbatelné. Tento způsob vychází z rovnice (9):

$$PH = b * (y - \mu) \quad (9)$$

Kde:

**b** je regrese PH na odchylku užitkovosti,

**y** je naměřená užitkovost očištěná o systematické vlivy prostředí,

**μ** je průměrná užitkovost v populaci.

Hodnota v závorce odpovídá DRP.

Regresní koeficient **b** je v případě, že se jedná o vlastní užitkovost, roven spolehlivosti předpovědi PH. V případě užitkovosti potomstva se **b** rovná dvojnásobku spolehlivosti předpovědi PH (Jakubec *et al.* 2010), protože býk svým potomkům předá pouze polovinu genetického založení. Odchylku užitkovosti tak lze získat vydělením plemenné hodnoty její spolehlivostí.

### 3.2 Efektivní počet dcer

EDC je odpovídajícím efektivním počtem dcer na býka, který je užíván při národním i mezinárodním hodnocení býků, kombinaci více zdrojů informací a také jako „váha“ býka při předpovědi regresních koeficientů metodou RR-BLUP, o které bude dále pojednáno.

Při genetickém hodnocení není přesné používat skutečný počet dcer na býka. Jednotlivé dcery se značně liší fází laktace, počtem ukončených laktací nebo počtem vrstevnic (Fikse *et Banos*, 2001). EDC je možné zpětně odvodit ze známé spolehlivosti předpovědi PH býka a dědivosti dané vlastnosti (10):



$$\Psi_i = \frac{\alpha r_i^2}{1 - r_i^2} \quad (10)$$

Kde:

$\Psi_i$  je EDC i-tého býka,

$r_i^2$  je spolehlivost předpovědi PH i-tého býka,

$\alpha$  je poměr reziduálního a genetického rozptylu  $\alpha = (4 - h^2)/h^2$ ,

$h^2$  je koeficient dědivosti.

Pro stanovení spolehlivosti předpovědi PH je potřeba invertovat část levé strany soustavy rovnic pro předpověď PH, což mnohdy vzhledem k velkému množství rovnic není možné. Proto se spolehlivost předpovědi PH stanovuje nepřímou z vypočteného EDC na jedince (Plemdat, 2010).

Fikse *et* Banos (2001) uvádějí další možnosti a porovnání výpočtu efektivního počtu dcer.

Vedle přepočteného počtu dcer u býků se používá také přepočtený počet vlastních užitkovostí na jedince (ERC). Výpočet se liší pouze v poměru rozptylů  $\alpha$  (11), který je v případě ERC:

$$\alpha = \frac{1 - h^2}{h^2} \quad (11)$$

### 3.3 Mezinárodní hodnocení býků

Jak už bylo uvedeno v kapitole o genetickém hodnocení jedinců, není možné jedince porovnávat v mezinárodním měřítku na základě národních plemenných hodnot, jelikož se jednotlivé země značně liší mimo jiné v chovatelských podmínkách, v použitých modelech a efektech zahrnutých do předpovědi PH. PH jedince je vždy platná pouze pro populaci, ve které je předpovězena a není možné přímo přejímat PH býka, která byla předpovězena v zemi jiné.

Nejjednodušší metodou pro přepočet plemenné hodnoty býků mezi zeměmi je použití regrese, kde PH v jedné zemi je závisle proměnnou a PH v zemi druhé je nezávisle proměnnou. Převod plemenných hodnot mezi zeměmi pomocí regrese se používal především

v minulosti před zavedením metody MACE. Weigel *et* Lin (1999) uvádějí, že v současné době se regrese používá pro převod mezi zeměmi v případě, že nejsou členy Interbullu, nebo pro vlastnosti, které nejsou jinak hodnoceny na mezinárodní úrovni.

Vztah pro přepočítání PH lze obecně vyjádřit následovně (12):

$$PH_{IMP} = a + b PH_{EXP} \quad (12)$$

Kde:

$PH_{IMP}$  je neznámá PH (závisle proměnná) v dovozní zemi,

$PH_{EXP}$  je známá PH (nezávisle proměnná) ve vývozní zemi,

$a$  je konstanta (genetická báze),

$b$  je regresní koeficient.

Existuje několik metod pro převod informací mezi zeměmi založených na regresi. Jednotlivé metody se liší především použitou závisle a nezávisle proměnnou. Je to například Goddardova metoda (Goddard, 1985) založená na regresi DYD v importující zemi na PH v exportující zemi. Wilminkova metoda (Wilmink *et al.*, 1986) je založená na regresi PH v importující zemi na odchylce PH od průměru vynásobené spolehlivostí v zemi exportující. Dále je možné použít prostou regresi PH v zemi importující na PH v zemi exportující, nebo regresi DYD v zemi importující na DYD v zemi exportující. Wiegeler *et* Lin (1999) uvádějí, že je vhodnější použít robustní regresi, aby nedocházelo k vychýlení předpovědi v důsledku odchýlených hodnot.

Interbull (1990) uvádí, že pro dostatečnou přesnost výše uvedených metod je potřeba, aby pro odhad regresních koeficientů bylo známo alespoň 20 býků, kteří mají v obou zemích (importující a exportující) spolehlivost předpovědi PH alespoň 0,75 a korelaci mezi PH v obou zemích 0,75.

### 3.3.1 MACE

Pro potřeby mezinárodního hodnocení býků navrhl Schaeffer (1994) metodu MACE. Podstatou metody MACE je hodnocení býků na základě DRP, odregresovaných z plemenných hodnot předpovězených v jednotlivých zemích. K DRP z jednotlivých zemí se přistupuje, jako by se jednalo o samostatnou vlastnost. Jako váha DRP se používá EDC

daného býka, vypočtený na základě spolehlivosti předpovědi PH (Mrode, 2005). V případě, že je hodnocena pouze jedna vlastnost z každé země, se jedná o metodu ST-MACE. Výhodou metody MACE je, že do hodnocení býků jsou zahrnuty informace ze všech zemí zapojených v mezinárodním hodnocení, ve kterých má býk potomstvo. Dochází tak k významnému nárůstu spolehlivosti předpovědi PH. I v případě, že v některé zemi býk potomstvo nemá, může mu být PH pro tuto zemi předpovězena.

Model metody MACE (13) má následující podobu (Schaeffer, 1994):

$$\mathbf{y}_i = \boldsymbol{\mu}_i \mathbf{1} + \mathbf{Z}_i \mathbf{Q} \mathbf{g}_i + \mathbf{Z}_i \mathbf{s}_i + \mathbf{e}_i \quad (13)$$

Kde:

$\mathbf{y}_i$  je vektor DRP z i-té země,

$\boldsymbol{\mu}_i$  je průměr i-té země,

$\mathbf{Z}_i$  je incidenční matice přiřazující DRP z i-té země plemeníkům,

$\mathbf{Q}$  je matice přiřazující plemeníky ke skupině neznámých předků,

$\mathbf{g}_i$  je vektor plemenných hodnot skupin neznámých předků v i-té zemi,

$\mathbf{s}_i$  je vektor plemenných hodnot býků v i-té zemi,

$\mathbf{e}_i$  je vektor reziduálních chyb býků v i-té zemi.

Rovnice smíšeného modelu (14) (Mrode, 2005):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \otimes \mathbf{G}^{-1} \\ \mathbf{0} & -\mathbf{Q}'\mathbf{A}^{-1} \otimes \mathbf{G}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ \mathbf{Q}\hat{\mathbf{w}} + \hat{\mathbf{a}} \\ \hat{\mathbf{w}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{0} \end{bmatrix} \quad (14)$$

Kde:

$\mathbf{X}$ ,  $\mathbf{Z}$  a  $\mathbf{Q}$  jsou incidenční matice přiřazující jednotlivá pozorování k průměru dané země, plemeníkovi a skupině neznámých předků,

$\mathbf{A}$  je matice příbuznosti,

$\mathbf{G}$  je genetická variančně-kovarianční matice mezi zeměmi,

$\mathbf{R}$  je matice reziduálních rozptylů, zahrnující EDC býků,

$\hat{\mathbf{c}}$  je vektor obsahující odhady průměrů jednotlivých zemí,

$\hat{\mathbf{a}}$  je vektor předpovědi PH býků v jednotlivých zemích,

$\hat{\mathbf{w}}$  je vektor PH skupin neznámých předků v jednotlivých zemích,

$\mathbf{y}$  je vektor DRP,

$\otimes$  je symbol pro přímé násobení matic.

Pro hodnocení více vlastností navrhli Schaeffer (2001), Sullivan *et* Wilton (2001) a Liu *et al.* (2004) metodu MT-MACE. Základní rozdíl mezi metodami ST-MACE a MT-MACE je, že při metodě MT-MACE je navíc počítáno s genetickými kovariancemi mezi současně hodnocenými vlastnostmi, což vede k nárůstu spolehlivosti předpovědi PH.

### 3.4 Selektce s využitím markerů – MAS

Pro předpověď plemenné hodnoty mladých zvířat bez vlastní užitkovosti nebo potomstva je obvykle dostupné jen málo informací. Důsledkem nízkého množství informací je nízká spolehlivost předpovězené plemenné hodnoty. Na konci 20. Století se pro zvýšení spolehlivosti předpovědi PH začaly využívat poznatky z molekulární genetiky v podobě MAS – selektce s využitím genetických markerů.

Podstatou MAS je nalezení QTL a markerů, které jsou ve vazbové nerovnováze s QTL, a zahrnutí těchto informací do předpovědi plemenné hodnoty. Jako marker většinou slouží jednoduše zjizitelný gen, nebo polymorfní úsek DNA a v některých případech může být markerem samotný QTL.

Síla vazby mezi markerem a QTL je dána rekombinačním poměrem, který udává poměr gamet rekombinovaných ze všech gamet. Žádoucí je pevná vazba resp. malá rekombinační četnost mezi markerem a QTL, což zpřesňuje předpověď substitučních efektů a PH. Dle simulačních studií může navýšení genetického zisku s využitím MAS pohybovat od 2 % (Ruane *et* Colleau, 1994) do 30 % (Kashi *et al.*, 1990).

PH se zahrnutím polygenního efektu a markerů lze vyjádřit následovně (15) (Mrode, 2005):

$$\mathbf{a}_i = \mathbf{v}_i^p + \mathbf{v}_i^m + \mathbf{u}_i \quad (15)$$

Kde:

$\mathbf{a}_i$  je plemenná hodnota  $i$ -tého jedince,

$\mathbf{v}_i^p$  je polovina plemenné hodnoty otce  $i$ -tého jedince pro marker QTL,

$\mathbf{v}_i^m$  je polovina plemenné hodnoty matky  $i$ -tého jedince pro marker QTL,

$u_i$  je plemenná hodnota i-tého jedince pro polygeny.

Plemenná hodnota se zahrnutím markerů je předpovídána metodou BLUP. Model má následující podobu (16) (Mrode, 2005):

$$y = X\beta + Za + Wv + e \quad (16)$$

Kde:

$y$  je vektor naměřených užitekostí,

$\beta$  je vektor fixních efektů,

$a$  je vektor PH předpovězených na základě polygenních efektů,

$v$  je vektor PH předpovězených na základě alel markeru,

$e$  je reziduální chyba,

$X, Z, W$  jsou matice plánu pokusu pro fixní efekt, polygenní efekt a substituční efekt markeru.

Rovnice smíšeného modelu (17) (Mrode, 2005):

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A_u^{-1}\alpha_1 & Z'W \\ W'X & W'Z & W'W + G_v^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix} \quad (17)$$

Kde:

$\hat{\beta}, \hat{a}, \hat{v}$  jsou odhad a předpovědi výše uvedených efektů,

$A_u$  je matice příbuznosti,

$G_v$  je gametická matice příbuznosti (Příbyl, 1995),

$\alpha_1$  (18),  $\alpha_2$  (19) jsou poměry rozptylů vypočtené podle následujících vztahů:

$$\alpha_1 = \sigma_e^2 / \sigma_a^2 \quad (18)$$

$$\alpha_2 = \sigma_e^2 / \sigma_v^2 \quad (19)$$

Kde:

$\sigma_e^2$  je reziduální rozptyl,

$\sigma_a^2$  je genetický rozptyl vysvětlený polygeny,

$\sigma_v^2$  je genetický rozptyl vysvětlený markery.

Mimo toho, že existuje vazba mezi markerem a QTL existuje také korelace mezi QTL a polygeny. Při selekci jedinců na QTL se mění kovariance mezi QTL a polygeny, což vede k poklesu odezvy na selekci u polygenů (Příbyl, 1995).

Genomická selekce je vyústěním MAS. Oproti MAS je při genomické selekci zahrnut podstatně větší počet markerů a nebývá zahrnut polygenní efekt. Předpokládá se, že vysoký počet hustě rozmístěných markerů napříč genomem vysvětlí celou genetickou proměnlivost. Calus *et* Veerkamp (2007) však uvádějí, že v případě, že není zahrnut v modelu vedle SNP i polygenní efekt, dochází k podhodnocení celkového genetického rozptylu.

### 3.5 Genotypování jedinců

Oproti MAS, kdy se do hodnocení jedince zapojuje jeden, nebo jen několik málo markerů, se při genomické selekci používají tisíce až statisíce markerů, které jsou rozmístěny napříč téměř celým genomem. Genotypování jedince pro vysoký počet markerů vyžaduje jednoduchou a rychlou technologii. Jako nejúčelnější se ukázaly být čipy, které jsou schopny v kombinaci se „čtečkou“ přečíst SNP napříč téměř celým genomem levně a v krátkém čase.

Asi nejúspěšnější na trhu v oblasti genetických čipů, nejen pro skot, ale i pro ostatní hospodářská zvířata a člověka je společnost Illumina ([www.illumina.com](http://www.illumina.com)). Jejím v současné době nejpoužívanějším čipem pro genotypování skotu je BovineSNP50 v2 BeadChip, který je schopný rozlišit bodové mutace (SNP) v 54 609 lokusech.

Na každém čipu jsou tisíce „korálků“, na kterých jsou uchyceny oligonukleotidové úseky (75 nukleotidů). Část každého úseku (50 nukleotidů) je komplementární s konkrétním úsekem genomické DNA. Genomická DNA, která je přidána na čip, hybridizuje v jednovláknové podobě s komplementárním úsekem. Hledaná bodová mutace se nachází na genomické DNA jednu bázi za hybridizovaným úsekem. Na čipu jsou volné, barevně značené nukleotidy, které jsou přidány polymerázami do místa bodové mutace podle pravidel komplementarity. Konkrétní alela je pak rozlišena podle barvy zabudovaného značeného nukleotidu pomocí „čtečky“ (Illumina, 2014).

Průměrná vzdálenost mezi jednotlivými SNP je 49,4 kb, medián 36,9 kb. Je možné s ním genotypovat 24 vzorků zároveň s opakovatelností vyšší než 99,9 %. Čip je použitelný pro všechna plemena skotu. Jednotlivá plemena se výrazně liší počtem polymorfních lokusů

(MAF > 0,05). U holštýnského skotu je průměrný počet polymorfních lokusů 42 849 a například u Jersey je to pouze 35 346. Průměr MAF je u holštýnského skotu 0,22 (Illumina, 2012).

### 3.6 Lokusy ovlivňující zkoumané vlastnosti

Prakticky všechny užitkové vlastnosti napříč celým spektrem hospodářských zvířat jsou polygenně založeny, nicméně dá se říci, že pro každou vlastnost lze nalézt lokusy, které mohou mít na danou vlastnost významnější vliv, který se může projevit tak, že mutace v těchto lokusech způsobí větší, či menší diskontinuitu v projevu vlastnosti v populaci. Lokusy ovlivňující užitkové vlastnosti se označují jako QTL (quantitative trait loci). Ve vztahu k dojivosti, procentickému obsahu bílkovin a tuku v mléce bylo uveřejněno několik studií a byly nalezeny lokusy, které by měly mít na tyto vlastnosti významný vliv (Khatkar *et al.*, 2004; Meuwissen *et al.*, 2004; Boichard *et al.*, 2003).

Khatkar *et al.* (2004) našli i QTL ovlivňující dojivost na 20 z 29 autosomů, statisticky nejvýznamnější ( $P < 0,001$ ) byly lokusy na chromozomech 1, 3, 6, 7, 14 a 20, na chromozomech 2, 3, 6, 14 a 20 našel statisticky významné lokusy ( $P < 0,001$ ) pro procentický obsah bílkovin v mléce a na chromozomech 2, 5, 6, 9, 10, 14, 19, 20 a 26 statisticky významné ( $P < 0,001$ ) QTL pro procentický obsah tuku v mléce. Deatwyler *et al.* (2008) objevili při použití 9 919 SNP významné lokusy ovlivňující dojivost na chromozomech 3, 5 a 16.

Arranz *et al.* (1998) publikovali substituční efekt 308 kg na chromozomu 20 při využití DYD jako závisle proměnné, k podobnému výsledku dospěl Georges *et al.* (1995), kteří zjistili substituční efekt 342 kg na stejném chromozomu.

Několik autorů publikovalo QTL uprostřed chromozomu 6 se substitučním efektem 0,07 % (Ron *et al.*, 2001) až 0,15 % (Spleman *et al.*, 1996). Další studie uvádí QTL ovlivňující především procentický obsah bílkovin na chromozomu 20 (Boichard *et al.*, 2003; Zhang *et al.*, 1998) a na chromozomu 14 se substitučním efektem v hodnotě 1,25 násobku celkové genetické směrodatné odchylky (Boichard *et al.*, 2003)

Lokusy ovlivňující procentický obsah tuku byly lokalizovány na centromerickém konci chromozomu 14 (Heyen *et al.*, 1999). Boichard *et al.* (2003) uvádějí substituční efekt ve výši 1,25 násobku genetické směrodatné odchylky. Někteří autoři uvádějí významný QTL pro procentický obsah tuku na chromozomu 3 (Plante *et al.*, 2001; Ron *et al.*, 1998).

Vedle hodnot substitučních efektů jednotlivých QTL je důležitý rovněž genetický rozptyl vysvětlený daným QTL. Boichard *et al.* (2003) uvádějí, že QTL na chromozomu 14 vysvětluje 13 % celkové genetické proměnlivosti. Celková genetická proměnlivost procentického zastoupení bílkovin je ze 7 %, 18 %, 30 %, 9 % a 7 % vysvětlena mutacemi v QTL na chromozomech 7, 18, 14, 18 a 20. 40 % celkové genetické proměnlivosti procentického zastoupení tuku vysvětluje QTL na chromozomu 14.

Jak již bylo uvedeno dříve, všechny zkoumané vlastnosti jsou polygenně řízené velkým množstvím genů. Pimentel *et al.* (2011) odhadují množství genů, které ovlivňují dojivost na 462-1593, v případě procentického zastoupení bílkovin na 337-1213 a u procentického zastoupení tuku je to 157-696.

Analýzu významu jednotlivých SNP u holštýnského skotu prováděli Szyda *et al.* (2011) a Jiang *et al.* (2010). Szyda *et al.* (2011) zjistili nejvyšší hodnotu regresního koeficientu pro dojivost 3,67 kg a potvrdili vliv genu DGAT1 na chromozomu 14 na dojivost.

Jiang *et al.* (2010) prováděli analýzu pro dojivost, nádoj a procentický obsah bílkovin a tuku. Pro dojivost uvádí 20 statisticky významných ( $p < 0,05$ ) SNP, z čehož 14 SNP je na chromozomu 14 a dále SNP na chromozomech 1, 3, 5, 9, 26. Pro procentický obsah bílkovin uvádějí 28 statisticky významných ( $p < 0,05$ ). Z tohoto počtu je 14 významných SNP na chromozomu 20, sedm SNP je na chromozomu 14, pět SNP je na chromozomu 6 a dvě SNP na chromozomu 5. U procentického zastoupení tuku bylo zjištěno 65 statisticky významných SNP, z čehož většina (59 SNP) se nachází na chromozomu 14.

### **3.7 Metody předpovědi genomické plemenné hodnoty**

Bylo vyvinuto několik různých metod pro zahrnutí informací z genetických čipů do hodnocení jedinců vycházející z metody BLUP. Metody lze jednoduše dělit na metody jednokrokové a vícekrokové. Jiménez-Montero *et al.* (2013) uvádějí jiný způsob třídění metod předpovědi genomické plemenné hodnoty, a to na metody založené na předpovědi regresních koeficientů jednotlivých SNP a na metody založené na nahrazení matice příbuznosti, nebo její úpravě.

V případě vícekrokových metod je potřeba nejdříve získat vstupní údaje v podobě DYD, YD nebo DRP referenční populace, což je skupina býků s vysokou spolehlivostí předpovězené PH. Tyto údaje slouží buď k předpovědi PGH (metoda GBLUP), nebo k předpovědi regresních koeficientů jednotlivých SNP u referenční populace a následně



předpovědi PGH hodnocených jedinců (metody RR-BLUP, BayesA, BayesB (Meuwissen *et al.* 2001)). Předpovězené PGH jsou pak pomocí indexu (Szyda *et al.*, 2011) kombinovány s rodokmenovou PH pro získání GEPH, jelikož regresní koeficienty nevysvětlují celou genetickou proměnlivost.

Jednokroková metoda ssGBLUP (Miształ *et al.*, 2009; Aguilar *et al.*, 2010; Christensen *et al.*; 2010) je založená na opravě matice příbuznosti s použitím matice realizované genomické příbuznosti. Závisle proměnnými vstupujícími do předpovědi GEPH jsou na rozdíl od ostatních genomických metod přímo naměřené užitkovosti jedinců. Výhodou této metody oproti ostatním je nárůst spolehlivosti předpovědi i pro jedince, kteří nejsou genotypováni.

V případě metod GBLUP a ssGBLUP je potřeba vytvořit genomickou matici příbuznosti mezi genotypovanými zvířaty. Při vytváření této matice se vychází z četností druhých alel v jednotlivých lokusech. Sestavení genomické matice příbuznosti je možné podle následujícího vztahu (20) (Forni *et al.*, 2011):

$$\mathbf{G} = \frac{(\mathbf{T} - \mathbf{Q})(\mathbf{T} - \mathbf{Q})'}{\text{tr}((\mathbf{T} - \mathbf{Q})(\mathbf{T} - \mathbf{Q})')/n} \quad (20)$$

Kde:

**T** je matice o rozměrech  $n \times m$  (počet jedinců  $\times$  počet SNP), která obsahuje počty druhých alel u  $n$ -tého jedince v  $m$ -tém SNP,

**Q** je matice o rozměrech  $n \times m$  obsahující průměrné počty druhých alel,

**n** je počet jedinců se známým genotypem.

Průměrné počty druhých alel v matici **Q** by měly být vypočítány z neselektované populace, což ale není prakticky možné, proto se počítá s četností druhé alely 0,5, resp. průměrným počtem druhé alely 1, nebo se používají četnosti druhých alel z vyhodnocované populace (Forni *et al.*, 2011; Příbyl *et al.*, 2012).

V případě metody GBLUP matice **G** v podstatě nahrazuje klasickou matici příbuznosti **A** (VanRaden, 2008). Při metodě ssGBLUP je pomocí matice **G** upraven blok klasické matice příbuznosti **A**, ve kterém se nacházejí genotypovaní jedinci za vzniku matice **H** podle následujících vztahů (21, 22) (Aguilar *et al.*, 2010; Příbyl *et al.*, 2012):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \quad (21)$$

$$\mathbf{F} = ((\mathbf{1} - \mathbf{w})\mathbf{G} + \mathbf{w}\mathbf{A}_{22})^{-1} - \mathbf{A}_{22}^{-1} \quad (22)$$

Kde:

$\mathbf{w}$  je váha matice příbuznosti s hodnotou obvykle 0,01-0,2,

$\mathbf{A}_{22}$  je blok matice příbuznosti pro genotypovaná zvířata.

Problémem metod ssGBLUP a GBLUP, zvláště při větším počtu genotypovaných jedinců, je potřeba invertování matice  $\mathbf{G}$ . Oproti tomu matici příbuznosti  $\mathbf{A}$  je jednoduché vytvořit již invertovanou podle Hendersona (1976).

Hlavním očekávaným přínosem genomiky je zvýšení spolehlivosti předpovědi PH. Hayes *et al.* (2009) uvádějí, že spolehlivost GEPH u mladých býků bez potomstva se pohybuje mezi 0,2 a 0,67.

## 4 Materiál a metody

### 4.1 Vstupní soubory

#### 4.1.1 Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi

Vstupní soubor s plemennými hodnotami a jejich spolehlivostmi, přepočtenými metodou MACE na podmínky České republiky, poskytla Českomoravská společnost chovatelů, a. s. Po načtení souboru a úpravě měřítek plemenných hodnot a spolehlivostí byl celý soubor statisticky vyhodnocen. Pro další výpočty byli použiti pouze ti býci, jejichž plemenná hodnota se nacházela méně jak tři směrodatné odchylky nad, nebo pod průměrem celého souboru a kteří dosahovali spolehlivosti předpovědi PH alespoň 0,6, tak, aby do souboru vstupovali pouze býci s dostatečným počtem dcer v KU.

Kvalita ogenotypování každého jedince byla hodnocena podle počtu lokusů se známým genotypem na jedince. Býk byl vyloučen ze souboru v případě, že byl znám genotyp býka v méně jak 90 % lokusů.

U procentického podílu tuku a bílkovin bylo ještě třeba předpovědět PH jedinců pro tyto vlastnosti, protože tyto PH nebyly v souboru obsaženy. PH byly přepočítány s použitím známých PH pro kg mléka a PH pro kg tuku a bílkovin dle následujícího vztahu (23) (Plemdat, 2010):

$$PH\%složky = \frac{100 * PHkg složky - xproc * PHkg mléka}{PHkg mléka + xmlk} \quad (23)$$

Kde:

**xproc** je průměrné % hodnocené složky v populaci,

**xmlk** je průměrná užitkovost kg mléka v populaci.

Hodnoty **xproc** (bílkoviny – 3,28 % a tuk 3,83 %) a **xmlk** (8170 kg), které byly použity pro dopočet PH procentického zastoupení bílkovin a tuku poskytl Plemdat.

#### 4.1.2 Výběr SNP pro předpověď regresních koeficientů

Býci byli genotypováni s použitím čipu Illumina BovineSNP50 v2 BeadChip (Illumina, 2012) pro 54 609 bodových mutací. Soubor s genotypy poskytl Svaz chovatelů

holštýnského skotu, o. s. Celý soubor genotypů byl podroben kontrole z hlediska kvality genotypování a zdvojení údajů.

Kontrola byla provedena i pro jednotlivé SNP, kde byl sledován počet jedinců se známým genotypem v každém SNP a zastoupení méně četné alely (MAF). SNP byl vyřazen z předpovědi regresních koeficientů, pokud genotyp v tomto lokusu chyběl u více než 10 % jedinců, nebo v případě, že MAF byla nižší než 5 %. V případě vysoké četnosti chybějícího genotypu, nebo v případě nízké MAF by toto mohlo vést k nepřesně předpovězenému regresnímu koeficientu. V případě, že byl neznámý genotyp u jedince v SNP, který ale prošel výběrem, byl tento genotyp nahrazen průměrným genotypem vypočteným z genotypů ostatních býků v tomto lokusu.

## 4.2 Předpověď regresních koeficientů

Samotná předpověď regresních koeficientů jednotlivých SNP byla prováděna ve dvou kolech. V prvním kole byly předpovězeny regresní koeficienty jednotlivých SNP, které následně byly použity pro předpověď PGH býků, kteří svým DRP vstupovali do výpočtu. Byla vypočítána absolutní hodnota rozdílu mezi DRP a PGH každého býka a byla provedena selekce býků pro druhé kolo předpovědi regresních koeficientů. Ze souboru byli vyřazeni býci, kteří měli absolutní hodnotu rozdílu mezi DRP a PGH vyšší než průměrná absolutní hodnota u všech býků plus dvě směrodatné odchylky. Vysoký rozdíl mezi DRP a PGH může být způsoben chybou ve vstupním souboru, nebo vysokou chybou předpovědi PH. Vyřazování býků s vysokou absolutní hodnotou rozdílu mezi DRP a PGH by mělo vést k zpřesnění předpovědi regresních koeficientů.

Ve druhém kole byly předpovězeny a vyhodnoceny regresní koeficienty jednotlivých SNP. Vedle vlastních regresních koeficientů byl vyhodnocován rovněž genetický rozptyl jednotlivých lokusů a heterozygotnost v těchto lokusech.

Pro předpověď regresních koeficientů byla použita hřebenová regrese, metoda RR-BLUP (ridge regression best linear unbiased prediction). Při této metodě se na diagonálu soustavy rovnic přičítají hodnoty, v tomto případě se jedná o poměr rozptylů násobený počtem vyhodnocovaných SNP. Přičtené hodnoty naruší závislost jednotlivých řádků v soustavě. Jako závisle proměnná slouží DRP vypočtená podle následujícího vztahu (24):

$$\mathbf{DRP} = \frac{\mathbf{PH}}{r^2} \quad (24)$$

Kde  $r^2$  je spolehlivost předpovědi PH.

Jedná se o váženou analýzu, DRP jednotlivých býků je vážena pomocí EDC, býci s vyšší spolehlivostí předpovědi PH tak k předpovědi regresních koeficientů přispívají více.

Předpověď regresních koeficientů byla provedena podle následujícího smíšeného modelu (25):

$$\mathbf{DRP} = \mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{v} + \mathbf{e} \quad (25)$$

Soustava rovnic pro tento model (26) upravená podle Szyda *et al.* (2011):

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{T} \\ \mathbf{T}'\mathbf{W}\mathbf{X} & \mathbf{T}'\mathbf{W}\mathbf{T} + \mathbf{f}\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{DRP} \\ \mathbf{T}'\mathbf{W}\mathbf{DRP} \end{bmatrix} \quad (26)$$

Kde:

$\mathbf{DRP}$  je vektor odregresovaných PH býků,

$\mathbf{X}$  je incidenční matice (vektor jednotek) přiřazující býky k fixnímu efektu – společné konstantě,

$\mathbf{T}$  je incidenční matice přiřazující DRP k jednotlivým regresním koeficientům, jednotlivé prvky v matici představují počty druhých alel v n-tém SNP, m-tého jedince,

$\mathbf{W}$  je diagonální matice vah obsahující EDC býků,

$\mathbf{I}$  je jednotková matice,

$\mathbf{f}$  je poměr rozptylů  $\mathbf{k}$  násobený počtem SNP  $\mathbf{k} = (\mathbf{4} - \mathbf{h}^2) / \mathbf{h}^2$ ,

$\hat{\boldsymbol{\beta}}$  je odhadovaný fixní efekt – společná konstanta,

$\hat{\mathbf{v}}$  je vektor předpovídaných regresních koeficientů,

$\mathbf{e}$  je náhodná chyba.

Koeficient dědivosti  $\mathbf{h}^2$  byl použit 0,25 u doживosti (Příbyl *et al.*, 2012) a 0,5 u procentického zastoupení bílkovin a tuku (Bouška *et al.*, 2006).

#### 4.2.1 Řešení soustavy rovnic pomocí Preconditioned conjugate gradient

Sestavenou soustavu rovnic nebylo z důvodu její velikosti možno řešit přímo s využitím inverze levé strany soustavy, ale musel být použit iterativní postup řešení, při kterém inverze není potřeba. Zvolen byl Preconditioned conjugate gradient (PCG) který je podle literatury (Legara *et Misztal*, 2008; Lidauer *et al.*, 1999) v řešení rovnic smíšených modelů rychlejší než Jacobiho iterace, nebo Gauss-Seidelova iterace. Vlastní algoritmus byl upraven podle Dale Van Vleck *et Dwyer* (1985), Tsuruta *et al.* (2001) a Lidauer *et al.* (1999) (27):

Vstupní vektory:

$$\begin{aligned} \mathbf{v}_0 &= \mathbf{0} \\ \mathbf{M} &= \frac{\mathbf{1}}{\text{diag}(\text{LHS})} \\ \mathbf{r}_0 &= \text{LHS} - \text{RHS} * \mathbf{v}_0 \\ \mathbf{p}_0 &= \mathbf{M} \otimes \mathbf{r}_0 \\ \mathbf{w}_0 &= \text{LHS} * \mathbf{p}_0 \end{aligned}$$

Iterativní opakování:

$$\begin{aligned} \mathbf{w}_n &= \text{LHS} * \mathbf{p}_{n-1} \\ \lambda_n &= \mathbf{p}'_{n-1} * \mathbf{w}_n \\ \alpha_n &= (\mathbf{p}'_{n-1} * \mathbf{r}_{n-1}) / \lambda_n \\ \mathbf{v}_n &= \mathbf{v}_{n-1} + \alpha_n * \mathbf{p}_{n-1} \\ \mathbf{r}_n &= \mathbf{r}_{n-1} - \alpha_n * \mathbf{w}_n \\ \beta_n &= (-\mathbf{r}'_n * \mathbf{w}_n) / \lambda_n \\ \mathbf{p}_n &= \mathbf{r}_n + \beta_n * \mathbf{p}_{n-1} \end{aligned} \tag{27}$$

Kde:

**LHS** je levá strana soustavy rovnic,

**RHS** je pravá strana soustavy rovnic,

$\mathbf{v}_n$  je vektor řešení v n-té iteraci,

**M** je vektor obsahující invertované diagonální prvky LHS,

$\mathbf{r}_n$  je vektor residuí v n-té iteraci,  
 $\mathbf{w}_n$  je pracovní vektor v n-té iteraci,  
 $\mathbf{p}_n$  je „search-direction“ vektor v n-té iteraci,  
 $\alpha_n, \beta_n$  a  $\lambda_n$  jsou „step-size“ skaláry v n-té iteraci,  
 $\otimes$  je symbol pro přímé násobení.

Podle Tsuruta *et al.* (2001) byl vektor residuí  $\mathbf{r}_n$  počítán každé 50. kolo podle vztahu  $\mathbf{r}_n = \mathbf{RHS} - \mathbf{LHS} * \mathbf{v}_n$  z důvodu zpřesnění předpovědi, protože jinak je vektor  $\mathbf{r}_n$  pouze aproximován a s každým kolem dochází k určitým chybám.

Iterativní řešení bylo ukončeno v případě, že nejvyšší absolutní hodnota prvku ve vektoru  $\mathbf{r}_n$  byla nižší než  $1 \cdot 10^{-8}$ .

Na základě řešení soustavy rovnic (26) byly vytvořeny grafy regresních koeficientů na jednotlivých chromozomech, byly lokalizovány lokusy s vysokými absolutními hodnotami regresních koeficientů a tyto výsledky byly porovnány s dostupnou vědeckou literaturou. Vyhodnocovány byly absolutní hodnoty regresních koeficientů, jelikož v souboru s SNP není patrné, která alela je původní a která mutovaná. Při předpovědi GEPH s využitím regresních koeficientů na tom, která alela je původní nezáleží.

### 4.3 Výpočet genetických rozptylů a heterozygotností v jednotlivých lokusech

Genetický rozptyl jednotlivých lokusů byl počítán podle následujícího vztahu (28) (Meuwissen *et al.*, 2001):

$$\sigma_i^2 = 2p_i(1 - p_i) \hat{v}_i^2 \quad (28)$$

Kde:

$\sigma_i^2$  je genetický rozptyl i-tého lokusu

$p_i$  je četnost jedné z alel i-tého lokusu

$\hat{v}_i^2$  je druhá mocnina předpovězeného regresního koeficientu i-tého lokusu

Heterozygotnost v daném lokusu byla počítána na základě stejného vztahu bez násobení druhou mocninou předpovězeného regresního koeficientu (29) (Meuwissen *et al.*, 2001):

$$\mathbf{Het} = 2\mathbf{p}_i (1 - \mathbf{p}_i) \quad (29)$$

#### 4.4 Předpověď PGH a GEPH

Po předpovězení regresních koeficientů jednotlivých SNP byla zpětně předpovězena PGH a GEPH. Tento krok sloužil jako kontrola přesnosti předpovědi regresních koeficientů. Předpovězené PGH a GEPH byly porovnávány se vstupními DRP a PH na základě korelace mezi jednotlivými proměnnými, na základě jejich průměrů, směrodatných odchylek a na základě absolutní hodnoty rozdílu mezi DRP a PH. Na vysokou přesnost by poukazoval vysoký koeficient korelace a nízké hodnoty absolutních rozdílů mezi DRP a PGH a mezi PH a GEPH.

Předpověď PGH byla provedena podle následujícího vztahu (29):

$$\mathbf{PGH} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{T}\hat{\boldsymbol{v}} \quad (29)$$

Kde:

$\mathbf{X}$  je incidenční matice (vektor jedniček) přiřazující býky k fixnímu efektu – společné konstantě,

$\mathbf{T}$  je incidenční matice přiřazující DRP k jednotlivým regresním koeficientům,

$\hat{\boldsymbol{\beta}}$  je odhadnutá společná konstanta,

$\hat{\boldsymbol{v}}$  je vektor předpovězených regresních koeficientů.

PGH byla následně sloučena s rodokmenovou PH. Sloučením se zahrne do předpovědi GEPH i polygenní efekt a proměnlivost, který nebyla vysvětlena regresními koeficienty. Sloučení PGH a rodokmenové PH bylo provedeno podle následujícího indexu (30):



$$\mathbf{GEPH} = \boldsymbol{\gamma}\mathbf{PGH} + \boldsymbol{\delta}\mathbf{PH} \quad (30)$$

Kde:

$\boldsymbol{\gamma}$  je váha PGH,

$\boldsymbol{\delta}$  je váha rodokmenové PH.

Váhy  $\boldsymbol{\gamma}$  a  $\boldsymbol{\delta}$  lze stanovit na základě spolehlivosti předpovědi PGH a PH (Szyda *et al.*, 2011). V této práci byly hodnoty  $\boldsymbol{\gamma}$  (0,8) a  $\boldsymbol{\delta}$  (0,2) zvoleny přímo.

Pro řešení celého postupu, popsaného v této metodice byl napsán vlastní program v prostředí SAS, který je vyžádání k dispozici u autora.

## 5 Výsledky

### 5.1 Dojivost

#### 5.1.1 Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi

Soubor s plemennými hodnotami pro dojivost přepočtenými Interbullem obsahoval PH 131 312 jedinců, které byly přepočteny na podmínky ČR. Průměrná PH v celém souboru činila -46,7 kg se směrodatnou odchylkou 578,4 kg. Nejnižší PH činila -2 563,7 kg a nejvyšší 2 180,3 kg. Ze souboru byli odstraněni jedinci, kteří se nacházeli mimo rozpětí PH -1 781,9 kg až 1 688,5 kg, a následně jedinci, kteří měli spolehlivost předpovědi PH nižší než 0,6. Tato hlediska splňovalo 122 976 jedinců, z čehož bylo genotypováno a zároveň mělo maximálně 10 % chybějících genotypů napříč lokusy 3 268 býků, ostatní jedinci byli ze souboru vyřazeni. U 2 904 jedinců zároveň byla známá jejich PH, která zároveň splňovala výše uvedené požadavky, a tyto jedinci mohli tak vstoupit do prvního kola předpovědi regresních koeficientů svými DRP. Do druhého kola předpovědi vstupovalo 2 767 býků. Parametry jedinců vstupujících do prvního a druhého kola předpovědi regresních koeficientů jsou uvedeny v Tab. 1.

**Tab. 1.** Plemenné hodnoty a odregresované plemenné hodnoty býků vstupujících do předpovědi regresních koeficientů pro dojivost

	1. Kolo		2. Kolo	
	$\mu$	$\sigma$	$\mu$	$\sigma$
PH (kg)	111,3	509,8	121,7	483,5
DRP (kg)	113,5	647,4	127,6	600,3
$r^2$	0,811	0,108	0,817	0,107
n	2 904 býků		2 767 býků	

PH – plemenná hodnota, DRP – odregresovaná plemenná hodnota,  $r^2$  – spolehlivost předpovědi plemenné hodnoty, n – počet jedinců

#### 5.1.2 Výběr SNP pro předpověď regresních koeficientů

Každý jedinec byl genotypován v celkem 54 609 lokusech, avšak některé lokusy vykazovaly vysoké procento chybějícího, nebo neznámého genotypu a nízkou četnost jedné z alel. V prvním kole předpovědi před odstraněním takovýchto lokusů byl genotyp známý

v průměru u 97,2 % SNP napříč všemi lokusy a jedinci a v souboru se nalézaly SNP, které měly genotyp známý u méně jak 1 % býků. Průměrný genotyp činil 1,081. Po odstranění lokusů s vysokým procentem neznámého genotypu a nízkou četností jedné z alel stoupl podíl lokusů se známým genotypem na 99,8 % a průměrný genotyp klesl na 1,058. 16 187 lokusů mělo známý genotyp u úplně všech jedinců, kteří vstupovali do předpovědi. Celkově požadavky v prvním kole předpovědi splňovalo 40 892 SNP, 13 717 SNP bylo z předpovědi vyřazeno.

Ve druhém kole předpovědi byl genotyp známý u 97,3 % všech SNP a jeho hodnota dosahovala v průměru shodně, jako v prvním kole 1,081. Po odstranění některých SNP podle shodných pravidel jako v prvním kole byl genotyp známý opět u 99,8 % všech SNP a průměrný genotyp poklesl na 1,059. Požadavky na MAF plnilo 42 798 SNP. Z celkového souboru 54 609 SNP mělo maximálně 10 % chybějících genotypů 51 796 SNP. V souboru se nalézaly SNP (celkem 16 476), které měly genotyp známý u všech jedinců. Ve druhém kole výběrem prošlo 40 890 SNP, což je 74,9 %, 13 719 SNP bylo vyřazeno. Nejvyšší procento vyřazených SNP bylo na chromozomu X, kde bylo vyřazeno 38,3 % SNP. Celkové počty SNP a počty SNP na každém chromozomu, pro které byl předpovězen regresní koeficient, jsou uvedeny v Tab. 2.

**Tab. 2.** Délky chromozomů, celkové počty SNP a počty použitých SNP ve druhém kole předpovědi na každém chromozomu

Chrom.	Mb	Celkem SNP	Použito – dojivost	Použito - % bílkovin	Použito - % tuku
1	158,3	3 431	2 600	2 604	2 601
2	137,1	2 829	2 071	2 071	2 072
3	121,4	2 550	1 933	1 932	1 932
4	120,8	2 570	1 935	1 935	1 936
5	121,2	2 271	1 690	1 689	1 690
6	119,5	2 575	1 986	1 989	1 989
7	112,6	2 352	1 706	1 715	1 713
8	113,4	2 430	1 808	1 811	1 808
9	105,7	2 095	1 558	1 557	1 558
10	104,3	2 206	1 662	1 663	1 664
11	107,3	2 295	1 742	1 740	1 741
12	91,2	1 773	1 316	1 319	1 316
13	84,2	1 850	1 385	1 386	1 386
14	84,6	1 831	1 385	1 385	1 385
15	85,3	1 762	1 324	1 324	1 322
16	81,7	1 726	1 247	1 247	1 245
17	75,2	1 600	1 231	1 232	1 232
18	66,0	1 376	1 053	1 053	1 052
19	64,1	1 420	1 084	1 085	1 085
20	72,0	1 568	1 193	1 194	1 194
21	71,6	1 483	1 082	1 083	1 083
22	61,4	1 324	982	981	981
23	52,5	1 092	842	840	841
24	62,7	1 312	964	965	963
25	42,9	1 004	787	788	788
26	51,7	1 116	832	836	836
27	45,4	981	756	757	756
28	46,3	981	764	766	767
29	51,5	1 087	841	843	843
X	148,8	1 170	722	726	722
Neznámý	9,5	548	409	409	410
<b>Celkem</b>	<b>2 670,4</b>	<b>54 608*</b>	<b>40 890</b>	<b>40 925</b>	<b>40 911</b>

\* navíc jedno SNP na chromozomu Y

Chrom. – chromozom, Mb – délka chromozomu v megabázích podle Bos Taurus UMD 3.1

Primary Assembly (<http://www.ncbi.nlm.nih.gov>)

### 5.1.3 Předpověď regresních koeficientů

Tab. 3. uvádí průměrné absolutní hodnoty regresních koeficientů a jejich směrodatné odchylky. Průměrná absolutní hodnota regresního koeficientu činila 0,506 kg a její směrodatná odchylka 0,409 kg. Nejvyšší průměrné absolutní hodnoty regresních koeficientů měly SNP na chromozomu X (0,751 kg,  $\sigma = 0,594$  kg), na chromozomu 14 (0,616 kg,  $\sigma = 0,678$  kg) a na chromozomu 5 (0,595 kg,  $\sigma = 0,463$ ). Skutečné hodnoty regresních koeficientů se pohybovaly v rozmezí -7,472 kg až 6,170 kg se směrodatnou odchylkou 0,650 kg a průměrem 0,004 kg.

**Tab. 3.** Průměrné absolutní hodnoty předpovězených regresních koeficientů pro doživost a jejich směrodatné odchylky

Chrom.	$\mu_{ \hat{\nu} }$ (kg)	$\sigma_{ \hat{\nu} }$ (kg)
1	0,444	0,345
2	0,446	0,341
3	0,487	0,371
4	0,426	0,343
5	0,595	0,463
6	0,485	0,374
7	0,479	0,364
8	0,456	0,346
9	0,485	0,375
10	0,447	0,339
11	0,525	0,402
12	0,462	0,359
13	0,491	0,397
14	0,616	0,678
15	0,488	0,377
16	0,496	0,392
17	0,503	0,384
18	0,570	0,449
19	0,600	0,462
20	0,528	0,406
21	0,472	0,375
22	0,538	0,417
23	0,543	0,427
24	0,548	0,413
25	0,505	0,394
26	0,554	0,430
27	0,522	0,392
28	0,581	0,456
29	0,533	0,394
X	0,751	0,594
Neznámý	0,502	0,425
Průměr	0,519	0,409

Chrom. – chromozom,  $\mu_{|\hat{\nu}|}$  - průměrná absolutní hodnota předpovězeného regresního koeficientu,  $\sigma_{|\hat{\nu}|}$  - směrodatná odchylka absolutních hodnot předpovězených regresních koeficientů

SNP byly seřazeny podle absolutních hodnot předpovězených regresních koeficientů. Ze 100 SNP, které dosahovaly nejvyšší absolutní hodnoty regresního koeficientu, bylo 32 lokalizováno na chromozomu 14, 14 na chromozomu X, 12 bylo na chromozomu 5 a sedm těchto SNP se nacházelo na chromozomu 18. Na ostatních chromozomech se nacházelo vždy méně jak pět SNP, které by se svou absolutní hodnotou regresního koeficientu nacházely v souboru 100 nejvýznamnějších SNP. Na chromozomu 14 se nacházelo 77,4 % SNP s absolutní hodnotou regresního koeficientu vyšší než 3 kg a 19,8 % těchto SNP se nacházelo na chromozomu X. Z SNP s absolutní hodnotou předpovězeného regresního koeficientu vyšší jak 2 kg se nacházelo 19,8 % na chromozomu 14, 13,5 % na chromozomu X, 9,9 % na chromozomu 5 a 6,8 % na chromozomu 18.

Tab. 4 uvádí 40 SNP s nejvyšší absolutní hodnotou regresních koeficientů včetně chromozomu, na kterém se nachází a jeho pozici.

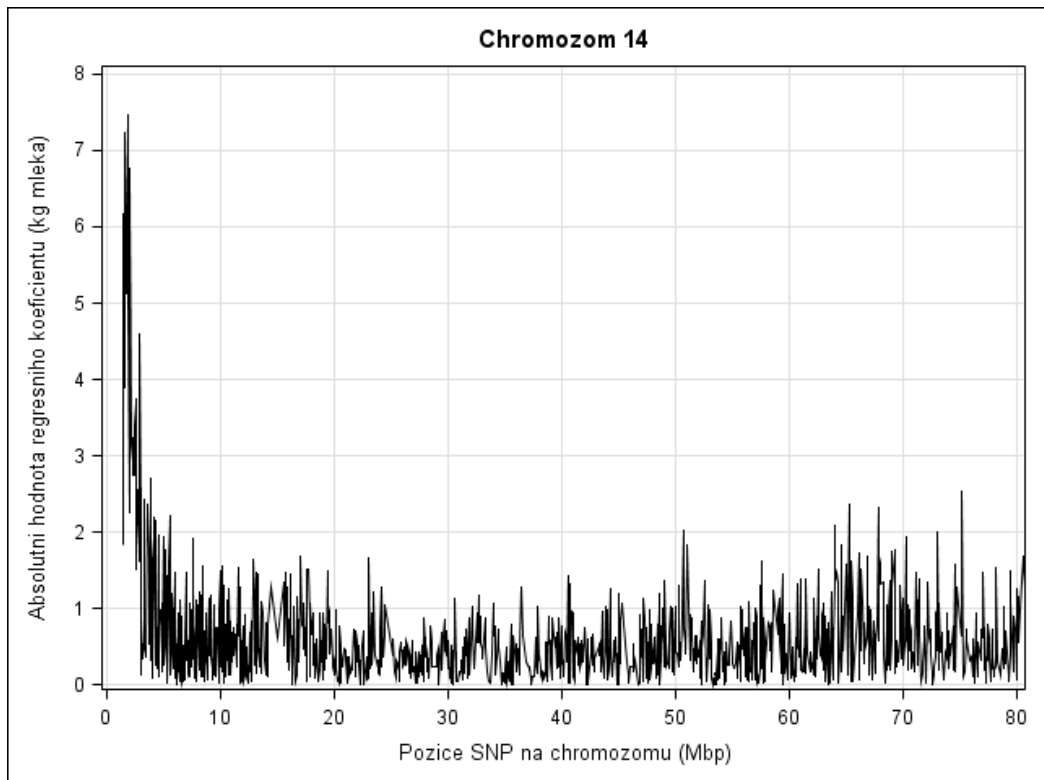
Nejvyšší absolutní hodnotu předpovězeného regresního koeficientu mělo SNP ARSBFGLNGS4939 na chromozomu 14 na pozici 1 801 116 bp uvnitř genu DGAT1, který se podle databáze Bos Taurus UMD 3.1 (<http://www.ncbi.nlm.nih.gov>), nachází na pozici 1 795 425-1 804 838 bp (<http://www.ncbi.nlm.nih.gov>). Další 21 SNP z Tab. 4 se nachází v blízkosti tohoto genu. Graf 1. ukazuje absolutní hodnoty regresních koeficientů na chromozomu 14. Na chromozomu 6 a 20, na kterých by se měly nacházet QTL pro dojivost, jsou nejvyšší absolutní hodnoty regresních koeficientu 2,574 kg (SNP BTA92518nors) a 2,723 kg (ARSBFGLNGS89521).

**Tab. 4.** SNP s nejvyššími absolutními hodnotami předpovězených regresních koeficientů pro dojivost

SNP	Chr.	Pozice (bp)	$ \hat{v} $ (kg)	$\sigma^2$	Het.
ARSBFGLNGS4939	14	1 801 116	7,472	21,925	0,393
ARSBFGLNGS57820	14	1 651 311	7,243	20,601	0,393
ARSBFGLNGS107379	14	2 054 457	6,763	19,723	0,431
Hapmap30383BTC005848	14	1 489 496	6,170	18,123	0,476
ARSBFGLNGS34135	14	1 675 278	5,180	13,256	0,494
ARSBFGLNGS94706	14	1 696 470	5,126	13,008	0,495
UAIFASA6878	14	2 002 873	5,090	12,938	0,499
Hapmap52798ss46526455	14	1 923 292	5,063	12,808	0,500
ARSBFGLNGS18858	14	2 909 929	4,597	10,049	0,475
BTA112878nors	X	17 249 271	4,185	5,932	0,339
BTA34956nors	14	1 514 056	3,898	7,427	0,489
Hapmap30646BTC002054	14	2 553 525	3,752	6,809	0,484
Hapmap30086BTC002066	14	2 524 432	3,704	6,831	0,498
UAIFASA8997	14	2 194 228	3,460	4,200	0,351
Hapmap25486BTC072553	0	0	3,438	5,078	0,430
Hapmap30374BTC002159	14	2 468 020	3,408	5,743	0,494
Hapmap50520BTA97804	X	19 953 258	3,341	5,524	0,495
ARSBFGLNGS18365	14	2 117 455	3,240	4,485	0,427
BTA35941nors	14	2 276 443	3,224	5,092	0,490
Hapmap25384BTC001997	14	2 217 163	3,150	4,956	0,500
Hapmap51205BTA30369	X	102 997 412	3,146	3,970	0,401
ARSBFGLNGS6633	19	37 217 989	3,142	4,429	0,449
Hapmap24715BTC001973	14	2 239 085	2,918	4,257	0,500
Hapmap34380BES4Contig2931272	5	92 955 993	2,890	3,768	0,451
ARSBFGLNGS40456	X	13 991 979	2,863	3,372	0,411
ARSBFGLNGS103844	X	18 596 782	2,822	2,600	0,326
Hapmap54224rs29017378	13	6 637 085	2,777	3,839	0,498
Hapmap7840BTA54134	22	31 649 896	2,767	3,505	0,458
BTA111275nors	26	9 195 089	2,762	3,348	0,439
ARSBFGLNGS26520	14	2 386 688	2,754	3,390	0,447
ARSBFGLNGS101653	14	2 319 504	2,752	2,872	0,379
ARSBFGLNGS89521	20	3 085 726	2,723	2,579	0,348
ARSBFGLNGS33248	14	3 885 798	2,702	3,273	0,448
ARSBFGLNGS899	28	33 328 952	2,702	3,589	0,492
Hapmap32415BTA74556	5	93 871 154	2,694	3,581	0,493
BTA03959nors	18	62 221 442	2,625	3,390	0,492
ARSBFGLNGS88771	28	33 359 220	2,620	3,424	0,499
ARSBFGLNGS117254	5	112 083 946	2,613	3,338	0,489
BTA74585nors	5	94 496 854	2,579	3,262	0,490
BTA92518nors	6	92 242 383	2,574	3,311	0,500

SNP – označení jednonukleotidového polymorfismu, Chr. – chromozom,  $|\hat{v}|$  - absolutní hodnota předpovězeného regresního koeficientu;  $\sigma^2$  - genetický rozptyl SNP, Het. – heterozygotnost





**Graf 1.** Absolutní hodnoty regresních koeficientů pro dojivost na chromozomu 14

#### 5.1.4 Genetický rozptyl lokusů

Průměrný genetický rozptyl jednoho lokusu byl 0,170 se směrodatnou odchylkou 0,384. Heterozygotnost v průměru dosahovala hodnoty 0,372 se směrodatnou odchylkou 0,118. Průměrný genetický rozptyl lokusu a průměrnou heterozygotnost jednoho lokusu na každém chromozomu ukazuje Tab. 5. Největší genetický rozptyl byl zjištěn u SNP umístěných na chromozomu 14 (0,360,  $\sigma = 1,424$ ) a na chromozomu X (0,368,  $\sigma = 0,603$ ). Rozdíly v heterozygotnosti lokusů na jednotlivých chromozomech byly minimální.

**Tab. 5.** Genetické rozptyly lokusů pro dojivost na jednotlivých chromozomech

Chrom.	$\mu_{\sigma^2}$	$\sigma_{\sigma^2}$	$\mu_{\text{Het}}$	$\sigma_{\text{Het}}$
1	0,126	0,208	0,364	0,121
2	0,127	0,204	0,377	0,115
3	0,149	0,239	0,377	0,115
4	0,121	0,224	0,369	0,119
5	0,229	0,389	0,371	0,120
6	0,150	0,249	0,372	0,118
7	0,142	0,220	0,370	0,118
8	0,130	0,201	0,368	0,119
9	0,147	0,245	0,363	0,122
10	0,123	0,193	0,370	0,118
11	0,176	0,286	0,372	0,116
12	0,137	0,225	0,367	0,121
13	0,161	0,303	0,376	0,120
14	0,360	1,424	0,386	0,112
15	0,156	0,250	0,373	0,118
16	0,161	0,274	0,372	0,118
17	0,162	0,251	0,371	0,117
18	0,218	0,369	0,382	0,114
19	0,229	0,367	0,373	0,117
20	0,173	0,283	0,368	0,119
21	0,147	0,238	0,378	0,117
22	0,185	0,304	0,369	0,120
23	0,197	0,325	0,374	0,119
24	0,183	0,287	0,371	0,118
25	0,170	0,280	0,382	0,116
26	0,196	0,340	0,366	0,120
27	0,170	0,269	0,371	0,119
28	0,222	0,386	0,379	0,116
29	0,177	0,265	0,376	0,119
X	0,368	0,603	0,373	0,118
Neznámý	0,175	0,374	0,376	0,112
Průměr	0,179	0,325	0,373	0,118

Chrom. – chromozom,  $\mu_{\sigma^2}$  – průměrný genetický rozptyl,  $\sigma_{\sigma^2}$  – směrodatná odchylka genetického rozptylu,  $\mu_{\text{Het}}$  – průměrná heterozygotnost,  $\sigma_{\text{Het}}$  – směrodatná odchylka heterozygotnosti

### 5.1.5 Předpověď PGH a GEPH

Pro kontrolu přesnosti předpovědi regresních koeficientů byla předpovězená PGH a GEPH jedinců vstupujících do předpovědi (2 904 býků) s použitím regresních koeficientů a ty byly následně porovnány se vstupními DRP a PH. Po prvním kole předpovědi byla korelace mezi vstupními DRP a předpovězenými PGH u všech býků 0,950 a korelace mezi PH a GEPH činila 0,968. Průměrná absolutní hodnota rozdílu mezi DRP býka a jeho PGH byla 185,1 kg. Po odstranění býků s vysokou absolutní hodnotou rozdílu mezi jejich DRP a PGH byly hodnoty korelačních koeficientů shodné jako před odstraněním, avšak průměrná absolutní hodnota rozdílu mezi vstupní DRP a PGH klesla na 158,6 kg. Vyřazení býků, kteří neplnili podmínky pro to, aby mohli vstoupit do druhého kola předpovědi regresních koeficientů, mírně vzrostla průměrná spolehlivost předpovědi PH z 0,811 na 0,817.

Po druhém kole předpovědi regresních koeficientů korelace mírně vzrostla na 0,951 mezi DRP a PGH a na 0,969 mezi PH a GEPH u celého souboru býků. Průměrná absolutní hodnota rozdílu mezi DRP a PGH byla 159,9 kg. Pro porovnání u souboru býků se spolehlivostí předpovědi PH 0,99 činily korelace 0,998 a 0,999. Průměrná absolutní hodnota rozdílu byla pouze 27,3 kg. Detailní výsledky ukazuje Tab. 6.

**Tab. 6.** Parametry býků vstupujících do druhého kola předpovědi regresních koeficientů pro doživost a výsledky předpovědi PGH a GEPH

	$r^2 \geq 0,6$		$r^2 \geq 0,85$		$r^2 \geq 0,9$		$r^2 \geq 0,95$		$r^2 = 0,99$	
	$\mu$ (kg)	$\sigma$ (kg)	$\mu$ (kg)	$\sigma$ (kg)	$\mu$ (kg)	$\sigma$ (kg)	$\mu$ (kg)	$\sigma$ (kg)	$\mu$ (kg)	$\sigma$ (kg)
PH	121,7	483,4	274,4	463,0	264,1	457,2	330,9	448,9	468,5	403,3
DRP	127,6	600,2	298,0	506,1	279,1	486,6	337,9	460,0	473,3	407,4
PGH	150,7	478,7	296,1	423,6	283,7	417,6	321,5	429,8	450,8	393,1
GEPH	144,9	475,9	291,7	429,4	279,8	424,0	323,4	433,2	454,4	395,0
rozd.	159,9	134,5	104,2	94,9	85,7	79,2	44,7	40,2	27,3	23,3
$r_{\text{DRP} \times \text{PGH}}$	0,951		0,970		0,978		0,994		0,998	
$r_{\text{PH} \times \text{GEPH}}$	0,969		0,982		0,987		0,996		0,999	
n	2 767		1 253		870		349		106	

$r^2$  - spolehlivost předpovědi PH – plemenná hodnota, DRP – odregresovaná plemenná hodnota, PGH – přímá genetická hodnota, GEPH – genomická plemenná hodnota, |rozd. | - absolutní hodnota rozdílu mezi DRP a PGH,  $r_{\text{DRP} \times \text{PGH}}$  – korelace mezi DRP a PGH,  $r_{\text{PH} \times \text{GEPH}}$  – korelace mezi PH a GEPH, n – počet jedinců

## 5.2 Procentický obsah bílkovin

### 5.2.1 Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi

PH pro procentický obsah bílkovin bylo potřeba nejdříve dopočítat z PH pro kg bílkovin. Soubor byl shodný jako v případě dojivosti a obsahoval PH pro 131 312 jedinců. Po přepočtu PH bylo u souboru provedeno statistické hodnocení a odstranění jedinců podle stejných kritérií, jako tomu bylo u dojivosti. Průměrná PH v celém souboru byla 0,005 % se směrodatnou odchylkou 0,285 %. Ze souboru byli tedy odstraněni jedinci s PH mimo rozmezí -0,850 až 0,860 %. Podmínky splňovalo 131 219 jedinců a z toho 2 995 jedinců bylo ogenotypováno, nemělo více jak 10 % SNP s neznámým genotypem a mohlo tak svým genotypem a DRP vstoupit do prvního kola předpovědi regresních koeficientů, do druhého kola vstupovalo 2 856 býků. PH, DRP a spolehlivosti předpovědi PH býků vstupujících do prvního a druhého kola předpovědi regresních koeficientů ukazuje Tab. 7.

**Tab. 7.** Plemenné hodnoty a odregresované plemenné hodnoty býků vstupujících do předpovědi regresních koeficientů pro procentický obsah bílkovin

	1. kolo		2. kolo	
	$\mu$	$\sigma$	$\mu$	$\sigma$
PH (%)	0,020	0,112	0,016	0,106
DRP (%)	0,025	0,129	0,019	0,119
$r^2$	0,894	0,067	0,898	0,066
n	2 995 býků		2 856 býků	

PH – plemenná hodnota, DRP – odregresovaná plemenná hodnota,  $r^2$  – spolehlivost předpovědi plemenné hodnoty, n – počet jedinců

### 5.2.2 Výběr SNP pro předpověď regresních koeficientů

Stejně jako u dojivosti byly vybrány SNP, které neměly četnost neznámého genotypu vyšší jak 10 % a hodnotu MAF nižší jak 5 %. V prvním kole předpovědi regresních koeficientů byl genotyp známý u 97,2 % lokusů před výběrem a u 99,8 % po výběru. Průměrný genotyp napříč lokusy a býky činil 1,081 před výběrem a 1,059 po výběru. Nastavené kritéria splňovalo v prvním kole 40 943 SNP, tj 75,0 % původního souboru. Ve druhém kole byly procenta známých genotypů a průměrné genotypy před a po výběru shodné s prvním kolem. MAF 5 % a vyšší mělo 42 836 SNP, genotyp známý alespoň u 90 %

býků mělo 51 795 SNP. V druhém kole byly regresní koeficienty předpovězeny pro 40 925 SNP (74,9 % původního souboru). Největší počet vyřazených SNP ze souboru byl opět na chromozomu X, na kterém nastavená kritéria nesplňovalo 38,0 % SNP. Tab. 2 ukazuje počty SNP na každém chromozomu, které prošly výběrem a byl pro ně předpovězen regresní koeficient.

### 5.2.3 Předpověď regresních koeficientů

U procentického zastoupení bílkovin vychází průměrná absolutní hodnota regresního koeficientu  $1,206 \times 10^{-4} \%$  se směrodatnou odchylkou  $1,008 \times 10^{-4} \%$ , skutečná průměrná hodnota regresního koeficientu byla  $4,492 \times 10^{-7} \%$  ( $\sigma = 1,572 \times 10^{-4} \%$ ). Nejvyšší průměrnou absolutní hodnotu měly SNP na chromozomu 14 ( $1,639 \times 10^{-4} \%$ ,  $\sigma = 1,762 \times 10^{-4} \%$ ), chromozomu X ( $1,800 \times 10^{-4} \%$ ,  $\sigma = 1,426 \times 10^{-4} \%$ ) a chromozomu 20 ( $1,500 \times 10^{-4} \%$ ,  $\sigma = 1,343 \times 10^{-4} \%$ ). Průměry předpovězených regresních koeficientů a jejich směrodatných odchylek na každém chromozomu ukazuje Tab. 8.

**Tab. 8.** Průměrné absolutní hodnoty předpovězených regresních koeficientů pro procentický obsah bílkovin a jejich směrodatné odchylky

Chrom.	$\mu_{ \hat{\nu} } \times 10^4 \%$	$\sigma_{ \hat{\nu} } \times 10^4 \%$
1	0,973	0,761
2	1,101	0,847
3	1,222	1,151
4	1,019	0,809
5	1,326	1,051
6	1,308	1,070
7	1,123	0,861
8	1,042	0,839
9	1,083	0,871
10	1,136	0,920
11	1,253	0,958
12	1,183	0,944
13	1,140	0,853
14	1,639	1,762
15	1,199	0,951
16	1,174	0,924
17	1,165	0,876
18	1,233	0,975
19	1,392	1,035
20	1,500	1,343
21	1,165	0,892
22	1,167	0,888
23	1,398	1,090
24	1,075	0,841
25	1,248	0,990
26	1,260	0,981
27	1,125	0,859
28	1,230	0,957
29	1,273	1,002
X	1,800	1,426
Neznámý	1,160	0,951
Průměr	1,229	0,990

Chrom. – chromozom,  $\mu_{|\hat{\nu}|}$  - průměrná absolutní hodnota předpovězeného regresního koeficientu,  $\sigma_{|\hat{\nu}|}$  - směrodatná odchylka absolutních hodnot předpovězených regresních koeficientů

Stejně jako u dojivosti byly SNP seřazeny podle absolutních hodnot regresních koeficientů. 40 SNP s nejvyšší absolutní hodnotou regresních koeficientů ukazuje Tab. 9. 22 SNP z těchto 40 SNP je na chromozomu 14, osm na chromozomu 20, šest na chromozomu 3, dva na chromozomu X a po jednom SNP na chromozomech 5 a 6. Nejvyšší absolutní hodnotu regresního koeficientu měl, stejně jako u dojivosti, SNP ARSBFGLNGS4939 na chromozomu 14 s hodnotou  $19,090 \times 10^{-4} \%$ . Dalších 17 SNP z Tab. 9. se nachází v blízkosti tohoto SNP, lokalizovaného uvnitř genu DGAT1. Graf 3. ukazuje absolutní hodnoty regresních koeficientů na chromozomu 14.

Na chromozomu 3 měl nejvyšší absolutní hodnotu předpovězeného regresního koeficientu ( $11,670 \times 10^{-4} \%$ ) SNP ARSBFGLNGS64215, který se nachází uvnitř genu EFNA1 a poblíže tohoto genu jsou umístěny další 4 SNP z Tab. 9. Absolutní hodnoty regresních koeficientů na tomto chromozomu ukazuje Graf 2.

Nejvyšší absolutní hodnotu regresního koeficientu na chromozomu 20 měl SNP ARSBFGLNGS118998 ( $10,130 \times 10^{-4} \%$ ) uvnitř genu GHR. Ostatní SNP na chromozomu 20 z Tab. 9. se nachází v blízkosti tohoto genu. Absolutní hodnoty regresních koeficientů na chromozomu 20 ukazuje Graf 4.

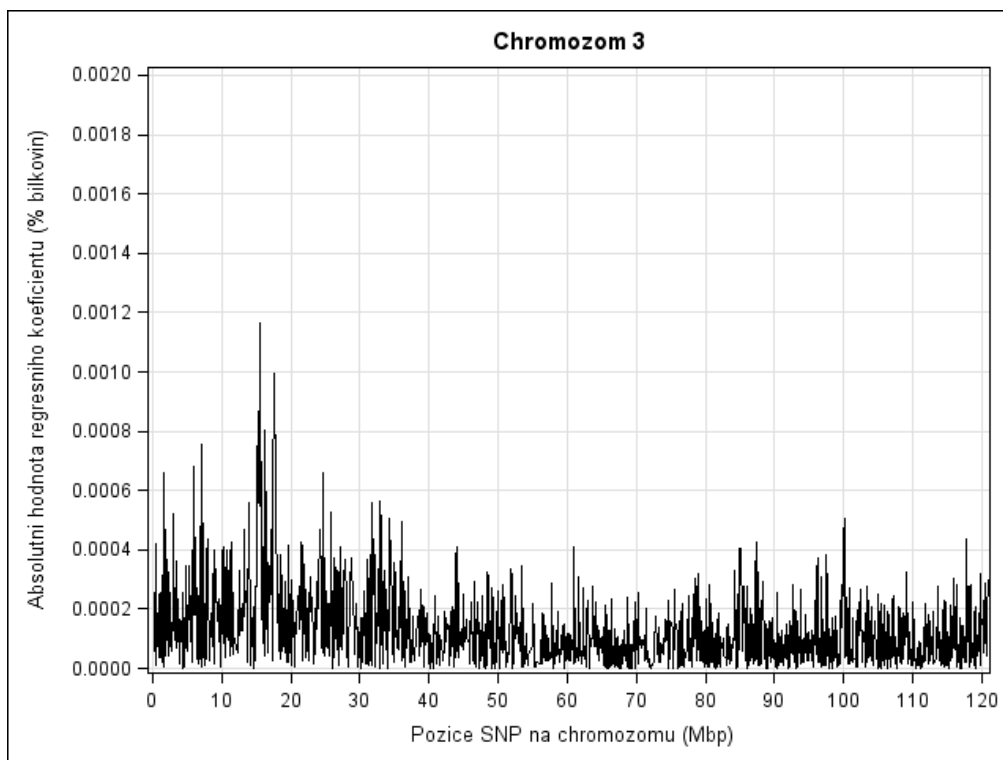
Ze 100 SNP s nejvyšší absolutní hodnotou předpovězeného regresního koeficientu bylo 39 na chromozomu 14, 18 na chromozomu 20, 11 na chromozomu 3, devět na chromozomu 6 a osm na chromozomu X.

**Tab. 9.** SNP s nejvyššími absolutními hodnotami předpovězených regresních koeficientů pro procentický obsah bílkovin

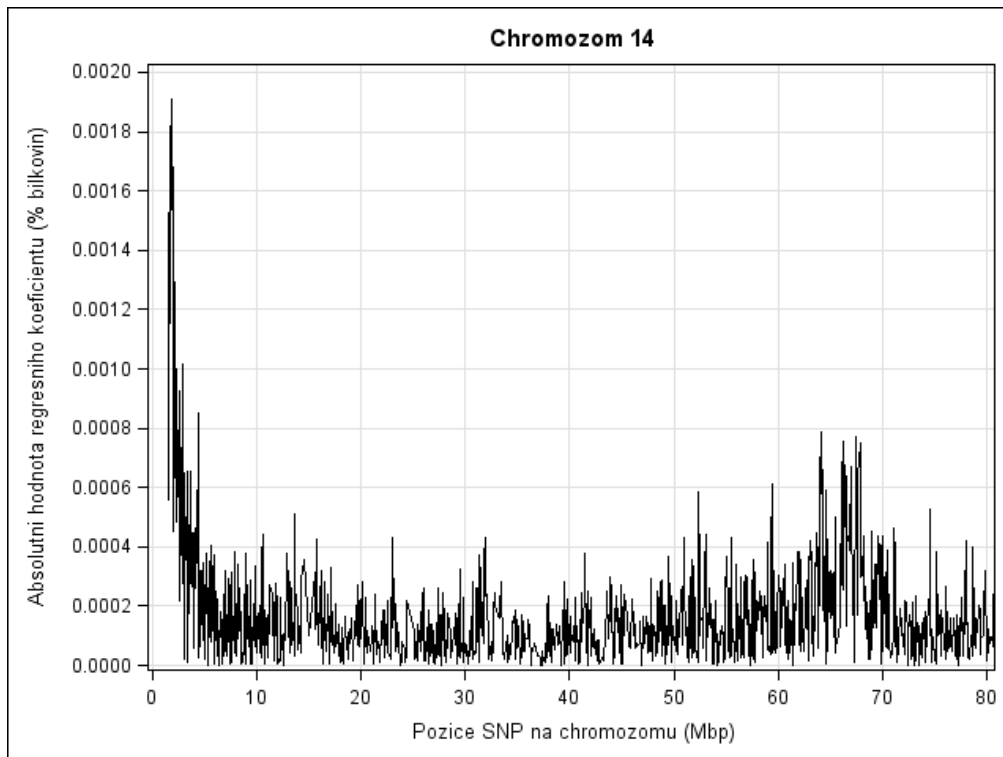
SNP	Chr.	Pozice (bp)	$ \hat{v}  \times 10^4$ %	$\sigma^2 \times 10^7$	Het.
ARSBFGLNGS4939	14	1 801 116	19,090	14,386	0,395
ARSBFGLNGS57820	14	1 651 311	18,190	13,074	0,395
ARSBFGLNGS107379	14	2 054 457	16,820	12,207	0,431
Hapmap30383BTC005848	14	1 489 496	15,250	11,094	0,477
Hapmap52798ss46526455	14	1 923 292	13,910	9,666	0,500
UAIFASA6878	14	2 002 873	12,850	8,244	0,499
ARSBFGLNGS64215	3	15 525 599	11,670	4,056	0,298
ARSBFGLNGS94706	14	1 696 470	11,650	6,726	0,496
ARSBFGLNGS34135	14	1 675 278	11,530	6,570	0,494
ARSBFGLNGS18858	14	2 909 929	10,140	4,908	0,477
ARSBFGLNGS118998	20	32 030 332	10,130	4,223	0,412
Hapmap24715BTC001973	14	2 239 085	9,980	4,980	0,500
BTA121572nors	3	17 681 439	9,960	3,960	0,399
Hapmap25384BTC001997	14	2 217 163	9,510	4,520	0,500
Hapmap30086BTC002066	14	2 524 432	9,280	4,283	0,497
BTA112878nors	X	17 249 271	9,090	2,741	0,332
ARSBFGLNGS18365	14	2 117 455	9,070	3,492	0,424
ARSBFGLNGS26909	20	33 551 316	8,870	2,376	0,424
BTB00777678	20	34 110 282	8,810	3,574	0,460
Hapmap39811BTA122745	20	33 311 925	8,690	3,251	0,430
ARSBFGLNGS13586	3	15 380 518	8,660	3,558	0,474
BTA84181nors	20	31 552 475	8,550	3,317	0,454
ARSBFGLNGS100480	14	4 364 952	8,520	3,490	0,481
BTA34956nors	14	1 514 056	8,330	3,388	0,488
ARSBFGLNGS17229	5	118 375 557	8,180	2,857	0,427
ARSBFGLNGS107582	3	16 243 393	8,050	3,198	0,493
Hapmap30646BTC002054	14	2 553 525	7,890	3,011	0,484
Hapmap49965BTA34335	14	64 151 328	7,880	2,517	0,405
ARSBFGLNGS24670	14	67 443 766	7,700	1,970	0,332
Hapmap47975BTA50226	20	32 293 167	7,660	2,909	0,496
UAIFASA1818	X	17 973 050	7,610	2,909	0,326
ARSBFGLNGS112786	3	7 147 596	7,580	2,870	0,500
Hapmap39823BTA35254	14	66 256 529	7,570	2,253	0,393
Hapmap50669BTA35344	14	67 831 515	7,520	2,827	0,500
Hapmap42643BTA69885	3	15 498 530	7,460	1,581	0,284
BTB01842107	20	37 374 576	7,350	1,523	0,282
Hapmap29888BTC003509	14	2 803 998	7,350	2,545	0,471
ARSBFGLNGS34049	20	35 249 040	7,310	2,433	0,455
Hapmap38371BTA105598	6	87 715 723	7,040	2,336	0,471
ARSBFGLNGS4929	14	63 972 246	7,000	1,616	0,330

SNP – označení jednonukleotidového polymorfismu, Chr. – chromozom,  $|\hat{v}|$  - absolutní hodnota předpovězeného regresního koeficientu;  $\sigma^2$  - genetický rozptyl SNP, Het. - heterozygotnost

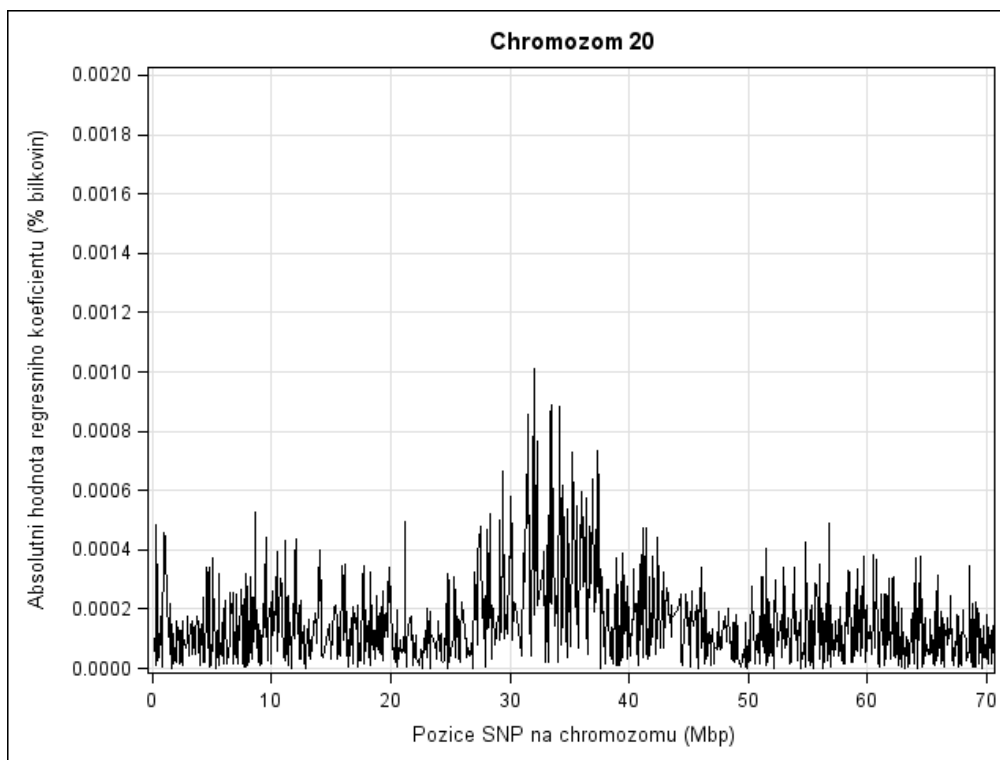




**Graf 2.** Absolutní hodnoty regresních koeficientů pro procentický obsah bílkovin na chromozomu 3



**Graf. 3** Absolutní hodnoty regresních koeficientů pro procentický obsah bílkovin na chromozomu 14



**Graf 4.** Absolutní hodnoty regresních koeficientů pro procentický obsah bílkovin na chromozomu 20

#### 5.2.4 Genetický rozptyl lokusů

Průměrná heterozygotnost v SNP použitých ve druhém kole předpovědi regresních koeficientů pro procentický obsah bílkovin v mléce činila 0,372 a směrodatná odchylka 0,118.

Genetický rozptyl vysvětlený jedním SNP činil v průměru  $99,64 \times 10^{-10}$  a směrodatná odchylka genetického rozptylu  $2,41 \times 10^{-8}$ . Nejvyšší genetické rozptyly měly SNP na chromozomu 14 a X, kde činily průměrné genetické rozptyly jednoho lokusu  $2,47 \times 10^{-8}$  ( $\sigma = 90,337 \times 10^{-9}$ ) a  $2,096 \times 10^{-8}$  ( $\sigma = 32,676 \times 10^{-9}$ ). Genetické rozptyly a heterozygotnosti na všech chromozomech jsou vedeny v Tab. 10. Rozptyly 40 SNP s nejvyšší hodnotou předpovězených regresních koeficientů jsou uvedeny v Tab. 9.

**Tab. 10.** Genetické rozptyly lokusů pro procentický obsah bílkovin na jednotlivých chromozomech

Chrom.	$\mu_{\sigma^2} \times 10^9$	$\sigma_{\sigma^2} \times 10^9$	$\mu_{\text{Het}}$	$\sigma_{\text{Het}}$
1	6,013	9,957	0,364	0,121
2	7,707	11,940	0,377	0,115
3	11,392	27,206	0,377	0,115
4	6,835	11,463	0,369	0,119
5	11,602	20,792	0,371	0,120
6	11,484	20,928	0,372	0,118
7	7,984	12,760	0,369	0,119
8	7,040	12,645	0,368	0,119
9	7,719	13,724	0,364	0,122
10	8,575	16,159	0,370	0,118
11	9,949	14,989	0,372	0,115
12	9,085	15,435	0,367	0,121
13	8,241	12,334	0,376	0,120
14	24,720	90,337	0,387	0,112
15	9,690	15,947	0,373	0,118
16	8,721	14,529	0,372	0,118
17	8,646	13,633	0,371	0,117
18	10,250	17,941	0,382	0,114
19	12,024	18,429	0,373	0,117
20	15,784	34,435	0,368	0,119
21	8,757	14,649	0,378	0,117
22	8,688	14,413	0,369	0,120
23	12,722	20,813	0,375	0,118
24	7,344	12,077	0,372	0,118
25	10,402	17,055	0,382	0,116
26	10,103	16,781	0,365	0,121
27	7,934	12,172	0,371	0,119
28	9,941	16,436	0,379	0,116
29	10,542	16,842	0,376	0,120
X	20,956	32,676	0,372	0,119
Neznámý	8,852	15,844	0,376	0,112
Průměr	10,313	19,205	0,373	0,118

Chrom. – chromozom,  $\mu_{\sigma^2}$  – průměrný genetický rozptyl,  $\sigma_{\sigma^2}$  – směrodatná odchylka genetického rozptylu,  $\mu_{\text{Het}}$  – průměrná heterozygotnost,  $\sigma_{\text{Het}}$  – směrodatná odchylka heterozygotnosti

### 5.2.5 Předpověď PGH a GEPH

Po první kole předpovědi regresních koeficientů činily korelace mezi předpovězenými PGH a DRP 0,964 a mezi PH a GEPH 0,979. Průměrná absolutní hodnota rozdílu mezi vstupními DRP a předpovězenými PHG byla 0,026 %. Kvůli vysoké absolutní hodnotě rozdílu mezi vstupní DRP a PGH bylo ze souboru vyřazeno 139 býků. Po odstranění těchto býků stouply na 0,974 mezi DRP a PGH a na 0,984 mezi PH a GEPH. Průměrná absolutní hodnota rozdílu klesla na 0,022 %.

Ve druhém kole mírně stouply korelace mezi DRP a PGH (0,975) i mezi PH a GEPH (0,985) u souboru všech býků, průměrná absolutní hodnota rozdílu mezi DRP a PGH zůstala stejná jako v prvním kole předpovědi po odstranění býků s vysokou chybou mezi DRP a PGH. U býků s nejspolehlivěji předpovězenou PH ( $r^2 = 0,99$ ) byly korelace mezi DRP a PGH a mezi PH a GEPH 1,000. Průměrná absolutní hodnota rozdílu mezi PGH a DRP činila 0,003 %. Detailnější výsledky ukazuje Tab. 11.

**Tab. 11.** Parametry býků vstupujících do druhého kola předpovědi regresních koeficientů pro procentický obsah bílkovin a výsledky předpovědi PGH a GEPH

	$r \geq 0,6$		$r \geq 0,85$		$r \geq 0,9$		$r \geq 0,95$		$r = 0,99$	
	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)
PH	0,016	0,106	0,010	0,106	0,001	0,104	0,001	0,101	0,024	0,103
DRP	0,019	0,119	0,012	0,115	0,001	0,110	0,000	0,104	0,025	0,104
PGH	0,018	0,103	0,010	0,101	0,002	0,100	0,002	0,098	0,026	0,102
GEPH	0,018	0,103	0,010	0,102	0,001	0,101	0,001	0,099	0,025	0,103
rozd.	0,022	0,020	0,018	0,017	0,014	0,013	0,009	0,008	0,003	0,003
$r_{\text{DRP} \times \text{PGH}}$	0,975		0,982		0,989		0,994		1,000	
$r_{\text{PH} \times \text{GEPH}}$	0,985		0,990		0,994		0,997		1,000	
n	2 856		2 026		1 496		872		197	

$r^2$  - spolehlivost předpovědi PH – plemenná hodnota, DRP – odregresovaná plemenná hodnota, PGH – přímá genetická hodnota, GEPH – genomická plemenná hodnota, |rozd. | - absolutní hodnota rozdílu mezi DRP a PGH,  $r_{\text{DRP} \times \text{PGH}}$  – korelace mezi DRP a PGH,  $r_{\text{PH} \times \text{GEPH}}$  – korelace mezi PH a GEPH, n – počet jedinců

## 5.3 Procentický obsah tuku

### 5.3.1 Úprava vstupního souboru s PH a výběr jedinců vstupujících do předpovědi

U této vlastnosti bylo stejně jako u procentického zastoupení bílkovin potřeba dopočítat PH z PH pro kg tuku a PH pro doživost. Průměrná PH pro procentický obsah tuku u celého souboru činila 0,052 % a její směrodatná odchylka 0,229 %. Ze souboru byli odstraněni jedinci s PH mimo rozmezí -0,635 % až 0,739 %. Po odstranění extrémních PH obsahoval soubor 130 631 jedinců.

Do předpovědi regresních koeficientů vstupovalo svým DRP v prvním kole 2 966 jedinců, kteří plnili nastavená kritéria na minimální spolehlivost předpovědi PH 0,6 a maximální počet neznámých genotypů napříč lokusy 10 %. Do druhého kola předpovědi vstupovalo 2 833 býků. Parametry býků vstupujících do prvního a druhého kola předpovědi ukazuje Tab. 12.

**Tab. 12.** Plemenné hodnoty a odregresované plemenné hodnoty býků vstupujících do předpovědi regresních koeficientů pro procentický obsah tuku

	1. Kolo		2. Kolo	
	$\mu$	$\sigma$	$\mu$	$\sigma$
PH (%)	0,054	0,242	0,044	0,230
DRP (%)	0,066	0,275	0,053	0,259
$r^2$	0,894	0,067	0,898	0,066
n	2 966 býků		2 833 býků	

PH – plemenná hodnota, DRP – odregresovaná plemenná hodnota,  $r^2$  – spolehlivost předpovědi plemenné hodnoty, n – počet jedinců

### 5.3.2 Výběr SNP pro předpověď regresních koeficientů

Kritéria pro výběr SNP byla shodná s předchozími vyhodnocovanými vlastnostmi. Procenta lokusů se známým genotypem a průměrné genotypy v obou kolech jsou shodné s výsledky u procentického zastoupení bílkovin. V prvním kole byly předpovězeny regresní koeficienty pro 40 924 SNP, ve druhém pro 40 911, což činí zaokrouhleně 74,9 % původního souboru. Nastavená kritéria na minimální hodnotu MAF plnilo 42 820 lokusů a kritérium na maximálně 10 % chybějících genotypů napříč býky plnilo 51 794 SNP. Nejvyšší procento

vyřazených SNP bylo opět na chromozomu X (38,3 %). Počty SNP na každém chromozomu, pro které byl předpovězen regresní koeficient, jsou uvedeny v Tab. 2.

### 5.3.3 Předpověď regresních koeficientů

Průměr absolutních hodnot předpovězených regresních koeficientů pro procentický obsah tuku v mléce vychází  $2,708 \times 10^{-4} \%$ , směrodatná odchylka  $2,575 \times 10^{-4} \%$ . Nejvyšší absolutní hodnoty regresních koeficientů měly SNP na chromozomu 14 s průměrnou hodnotou  $4,256 \times 10^{-4} \%$  ( $\sigma = 7,848 \times 10^{-4} \%$ ) a na chromozomu X ( $3,906 \times 10^{-4} \%$ ,  $\sigma = 3,274 \times 10^{-4} \%$ ). Výsledky pro všechny chromozomy jsou uvedeny v Tab. 13.

**Tab. 13.** Průměrné absolutní hodnoty předpovězených regresních koeficientů pro procentický obsah tuku a jejich směrodatné odchylky

Chrom.	$\mu_{ \hat{\beta} } \times 10^4 \%$	$\sigma_{ \hat{\beta} } \times 10^4 \%$
1	2,271	1,777
2	2,407	1,929
3	2,680	2,108
4	2,287	1,807
5	3,425	2,935
6	2,545	1,958
7	2,456	1,942
8	2,485	1,977
9	2,496	1,924
10	2,407	1,959
11	2,595	2,042
12	2,941	2,364
13	2,557	1,962
14	4,256	7,848
15	2,734	2,144
16	2,559	2,050
17	2,583	2,008
18	2,919	2,366
19	3,096	2,415
20	2,635	2,228
21	2,539	1,916
22	2,785	2,087
23	2,978	2,375
24	2,611	2,016
25	2,605	2,097
26	2,798	2,095
27	2,832	2,214
28	2,913	2,272
29	2,704	2,186
X	3,906	3,274
Neznámý	2,686	2,427
Průměr	2,764	2,345

Chrom. – chromozom,  $\mu_{|\hat{\beta}|}$  - průměrná absolutní hodnota předpovězeného regresního koeficientu,  $\sigma_{|\hat{\beta}|}$  - směrodatná odchylka absolutních hodnot předpovězených regresních koeficientů

Tab. 14. obsahuje SNP s nejvyšší absolutní hodnotou předpovězených regresních koeficientů. Nejvyšší absolutní hodnotu předpovězeného regresního koeficientu ( $98,070 \times 10^{-4} \%$ ) měl SNP ARSBFGLNGS4939 uvnitř genu DGAT1, stejně jako u předchozích vlastností. Poblíž tohoto genu je umístěno dalších 35 SNP z Tab. 14., dva SNP jsou lokalizovány na chromozomu 5, jeden na chromozomu X a jeden s neznámou pozicí.

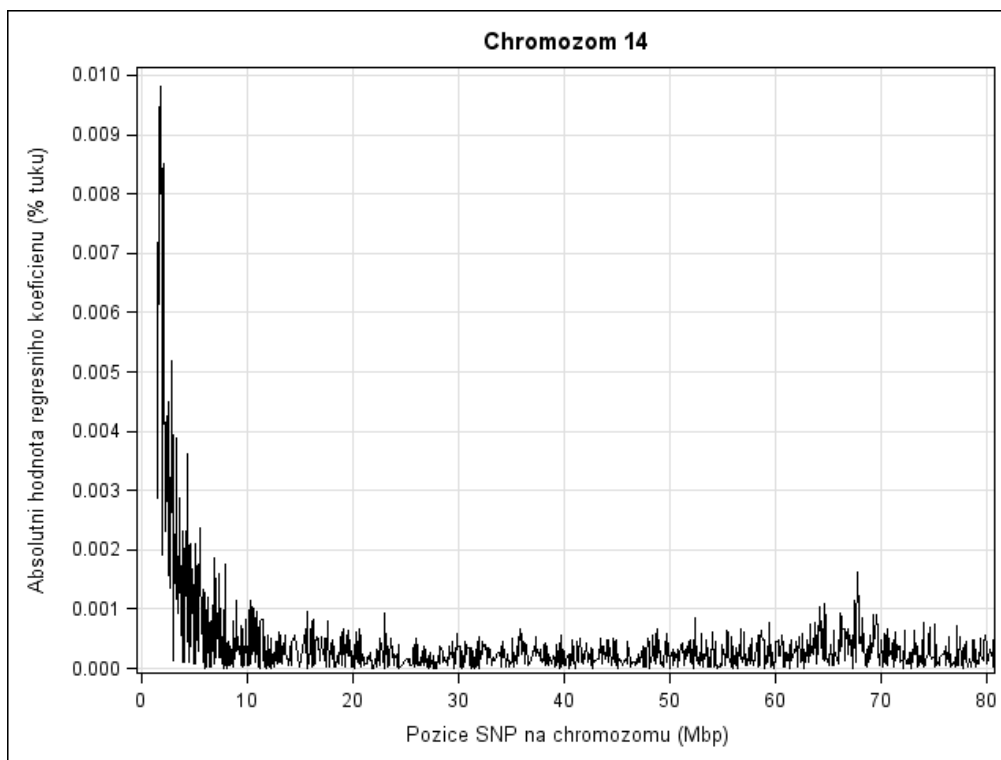
Ze 100 SNP s nejvyšší absolutní hodnotou předpovězeného regresního koeficientu je 67 na chromozomu 14, 18 na chromozomu 5, sedm na chromozomu X, dva na chromozomu 20, ostatní chromozomy mají jeden, nebo žádný. Absolutní hodnoty regresních koeficientů ukazuje Graf 5.



**Tab. 14.** SNP s nejvyššími absolutními hodnotami předpovězených regresních koeficientů pro procentický obsah tuku

SNP	Chr.	Pozice (bp)	$ \hat{v}  \times 10^4$ %	$\sigma^2 \times 10^7$	Het.
ARSBFGLNGS4939	14	1 801 116	98,070	373,393	0,388
ARSBFGLNGS57820	14	1 651 311	94,640	349,258	0,390
ARSBFGLNGS107379	14	2 054 457	85,080	309,498	0,428
Hapmap30383BTC005848	14	1 489 496	71,920	246,037	0,476
Hapmap52798ss46526455	14	1 923 292	70,810	250,593	0,500
UAIFASA6878	14	2 002 873	69,150	238,635	0,499
ARSBFGLNGS94706	14	1 696 470	61,590	187,579	0,494
ARSBFGLNGS34135	14	1 675 278	61,340	185,436	0,493
ARSBFGLNGS18858	14	2 909 929	51,870	127,844	0,475
ARSBFGLNGS18365	14	2 117 455	47,580	96,190	0,425
Hapmap30086BTC002066	14	2 524 432	44,940	100,563	0,498
Hapmap30646BTC002054	14	2 553 525	42,650	88,295	0,485
Hapmap25384BTC001997	14	2 217 163	41,580	86,358	0,499
BTA35941nors	14	2 276 443	41,040	82,379	0,489
Hapmap30374BTC002159	14	2 468 020	40,440	80,858	0,494
Hapmap24715BTC001973	14	2 239 085	39,980	79,890	0,500
Hapmap36620SCAFFOLD500187571	14	3 297 177	38,850	75,094	0,498
BTA34956nors	14	1 514 056	37,910	70,305	0,489
ARSBFGLNGS59769	14	3 006 509	37,590	69,637	0,493
ARSBFGLNGS100480	14	4 364 952	36,310	63,220	0,480
UAIFASA8997	14	2 194 228	34,300	41,307	0,351
Hapmap29888BTC003509	14	2 803 998	34,130	55,065	0,473
ARSBFGLNGS26520	14	2 386 688	33,240	49,557	0,449
ARSBFGLNGS103064	14	2 754 909	30,910	47,771	0,500
Hapmap30381BTC005750	14	1 463 676	28,830	35,835	0,431
ARSBFGLNGS74378	14	3 640 788	28,670	33,700	0,410
Hapmap25486BTC072553	0	0	28,350	34,611	0,431
ARSBFGLNGS56327	14	4 336 714	27,610	34,709	0,455
ARSBFGLNGS85419	14	3 243 897	27,230	33,668	0,454
Hapmap24717BTC002824	14	2 951 045	26,760	31,781	0,444
Hapmap32415BTA74556	5	93 871 154	26,510	34,612	0,492
ARSBFGLNGS22866	14	2 826 632	26,230	33,851	0,492
ARSBFGLNGS3122	14	2 721 633	25,750	32,529	0,491
Hapmap53294rs29016908	5	94 645 698	24,200	21,139	0,361
Hapmap29758BTC003619	14	2 674 264	23,780	27,859	0,493
ARSBFGLNGS115947	14	5 494 654	23,570	26,563	0,478
Hapmap50520BTA97804	X	19 953 258	23,360	26,965	0,494
ARSBFGLNGS101653	14	2 319 504	23,090	20,343	0,382
UAIFASA9288	14	3 956 956	23,060	21,304	0,401
Hapmap26072BTC065132	14	4 149 375	23,030	24,208	0,456

SNP – označení jednonukleotidového polymorfismu, Chr. – chromozom,  $|\hat{v}|$  – absolutní hodnota předpovězeného regresního koeficientu;  $\sigma^2$  – genetický rozptyl SNP, Het. – heterozygotnost



**Graf 5.** Absolutní hodnoty regresních koeficientu pro procentický obsah tuku na chromozomu 14

### 5.3.4 Genetický rozptyl lokusů

Průměrný genetický rozptyl pro procentický obsah tuku činil  $57,268 \times 10^{-9}$  ( $\sigma = 428,694 \times 10^{-9}$ ). Souhrnné výsledky genetického rozptylu lokusů obsahuje Tab. 15. Nejvyšší průměrný genetický rozptyl měly SNP na chromozomu 14 ( $352,457 \times 10^{-9}$ ,  $\sigma = 2261,935 \times 10^{-9}$ ). Průměrné genetické rozptyly na ostatních chromozomech byly výrazně nižší. Tab. 14. obsahuje genetické rozptyly nejvýznamnějších SNP.

**Tab. 15.** Genetické rozptyly lokusů pro procentický obsah tuku na jednotlivých chromozomech

Chrom.	$\mu_{\sigma^2} \times 10^9$	$\sigma_{\sigma^2} \times 10^9$	$\mu_{\text{Het}}$	$\sigma_{\text{Het}}$
1	32,850	58,697	0,364	0,121
2	38,529	64,091	0,377	0,115
3	46,690	75,372	0,377	0,115
4	33,883	59,038	0,369	0,119
5	81,768	178,992	0,371	0,120
6	41,266	66,500	0,372	0,118
7	39,548	66,187	0,369	0,119
8	40,838	69,079	0,368	0,119
9	39,829	63,178	0,364	0,122
10	38,599	68,982	0,370	0,118
11	44,027	75,077	0,372	0,115
12	56,462	98,780	0,367	0,121
13	40,600	63,263	0,376	0,120
14	352,457	2261,935	0,387	0,112
15	49,091	79,267	0,373	0,118
16	42,718	72,727	0,372	0,118
17	43,252	77,092	0,371	0,117
18	58,871	99,901	0,382	0,114
19	62,570	100,834	0,373	0,117
20	47,087	92,565	0,368	0,119
21	40,895	62,432	0,378	0,117
22	48,789	77,082	0,369	0,120
23	58,986	95,916	0,375	0,118
24	42,750	68,586	0,372	0,118
25	45,924	76,101	0,382	0,116
26	48,022	73,775	0,365	0,121
27	52,306	91,596	0,371	0,119
28	55,561	96,356	0,379	0,116
29	50,335	87,581	0,376	0,120
X	104,484	205,270	0,372	0,119
Nezn.	52,680	181,917	0,376	0,112
Průměr	59,086	158,328	0,373	0,118

Chrom. – chromozom,  $\mu_{\sigma^2}$  – průměrný genetický rozptyl,  $\sigma_{\sigma^2}$  – směrodatná odchylka genetického rozptylu,  $\mu_{\text{Het}}$  – průměrná heterozygotnost,  $\sigma_{\text{Het}}$  – směrodatná odchylka heterozygotnosti

### 5.3.5 Předpověď PGH a GEPH

Po prvním kole předpovědi činila korelace mezi vstupními DRP býků a jejich předpovězenými PGH 0,963. Korelace mezi PH a GEPH činila 0,979. Průměrná absolutní hodnota rozdílu mezi vstupními DPR a PGH byla 0,059 %. Po odstranění býků, kteří nespĺňovali nastavenou maximální absolutní hodnotu rozdílu mezi vstupní DRP a PGH, stouply korelace na hodnoty 0,973 a 0,983. Průměrná absolutní hodnota rozdílu klesla na 0,051 %.

Ve druhém kole předpovědi korelace pouze mírně vzrostly na hodnoty 0,974 (mezi DRP a PGH) a 0,984 (mezi PH a GEPH). Průměrná absolutní hodnota rozdílu mezi DRP a PGH klesla na 0,050 %. U souboru býků s nejspolehlivěji předpovězenou PH ( $r^2 = 0,99$ ) dosahovaly obě korelace hodnot 1,000, průměrná absolutní hodnota rozdílu činila pouze 0,009 %. Další výsledky jsou uvedeny v Tab. 16.

**Tab. 16.** Parametry býků vstupujících do druhého kola předpovědi regresních koeficientů pro procentický obsah tuku a výsledky předpovědi PGH a GEPH

	$r \geq 0,6$		$r \geq 0,85$		$r \geq 0,9$		$r \geq 0,95$		$r = 0,99$	
	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)	$\mu$ (%)	$\sigma$ (%)
PH	0,044	0,230	0,026	0,238	0,001	0,233	-0,003	0,226	-0,007	0,220
DRP	0,053	0,259	0,030	0,258	0,001	0,246	-0,003	0,223	-0,007	0,221
PGH	0,047	0,220	0,027	0,222	0,002	0,219	-0,001	0,216	-0,002	0,216
GEPH	0,047	0,221	0,027	0,225	0,002	0,221	-0,002	0,218	-0,003	0,217
rozd.	0,050	0,045	0,041	0,040	0,032	0,031	0,022	0,021	0,009	0,009
$r_{DRP \times PGH}$	0,974		0,982		0,989		0,994		1,000	
$r_{PH \times GEPH}$	0,984		0,990		0,994		0,997		1,000	
n	2 833		2 026		1 487		872		196	

$r^2$  - spolehlivost předpovědi PH – plemenná hodnota, DRP – odregresovaná plemenná hodnota, PGH – přímá genetická hodnota, GEPH – genomická plemenná hodnota, |rozd. | - absolutní hodnota rozdílu mezi DRP a PGH,  $r_{DRP \times PGH}$  – korelace mezi DRP a PGH,  $r_{PH \times GEPH}$  – korelace mezi PH a GEPH, n – počet jedinců

## 6 Diskuse

Byly předpovězeny regresní koeficienty bodových mutací u vlastností dojivost, procentický obsah tuku a procentický obsah bílkovin. Výsledky této práce nejsou přímo porovnatelné s výsledky ostatních autorů, protože porovnatelné publikace, které by uváděly substituční efekty jednotlivých SNP, se nepodařilo dohledat. Většina prací, které se zabývají lokusy ovlivňujícími kvantitativní vlastnosti, řeší spíše statistickou průkaznost existence QTL a jeho pozici (Khatkar *et al.*, 2004; Meuwissen *et al.*, 2004; Daetwyler *et al.*, 2008), případně substituční efekty, avšak při podstatně nižší hustotě markerů (Boichard *et al.*, 2003; Cole *et al.*, 2009; Banos *et al.*, 2008; Signorelli *et al.*, 2009). Substituční efekty v těchto pracích tak vycházejí podstatně vyšší a práce většinou neobsahují genetické rozptyly lokusů. Genetické rozptyly jsou více vypovídající, protože SNP může mít sice vysoký substituční efekt, avšak nízkou heterozygotnost a může tak vysvětlovat menší část genetické proměnlivosti než jiný lokus s menším substitučním efektem.

Do předpovědi regresních koeficientů pro dojivost vstupoval menší počet býků než do předpovědi regresních koeficientů pro procentický obsah bílkovin a tuku. Tento fakt bude s nejvyšší pravděpodobností způsoben vyšším koeficientem dědivosti u procentického zastoupení tuku a bílkovin a tím pádem spolehlivěji předpovězenými PH, jak ukazují i výsledky.

U procentického zastoupení bílkovin a tuku bylo rovněž nižší procento (4,64 % a 4,48 %) býků vyřazených po prvním kole předpovědi regresních koeficientů ve srovnání s dojivostí (4,75 % vyřazených), což lze opět vysvětlit vyšším koeficientem dědivosti u procentického zastoupení složek. Díky vyššímu koeficientu dědivosti je v průměru přesnější předpověď PH, ze kterých se dopočítávaly DRP vstupující do předpovědi, a zároveň vyšší dědivost způsobí spolehlivěji předpovězené regresní koeficienty, stejně jako je tomu u předpovědi PH. Spolehlivěji předpovězené regresní koeficienty v důsledku zpřesní předpověď PGH, potažmo GEPH.

Počty SNP, pro které byl předpovídán regresní koeficient, se u jednotlivých vlastností příliš neliší, což je dáno tím, že do předpovědi vstupuje stále stejný soubor s SNP a většina býků vstupovala do předpovědi regresních koeficientů u všech vlastností. Procento neznámých genotypů před provedením výběru v prvním i druhém kole bylo u všech vlastností prakticky shodné. Výsledky ukazují, že většina vyřazených SNP byla vyřazená z důvodu nedostatečné hodnoty MAF, kterou u všech vlastností splňovalo v průměru pouze

42 818 SNP. Illumina (2012) uvádí, že počet lokusů s  $MAF > 0,05$  u holštýnského skotu je 42 849. Vyřazení lokusů s nízkou hodnotou MAF mělo u všech vlastností v obou kolech za následek pokles průměrného genotypu. Szyda *et al.* (2011) použili 46 267 lokusů, avšak při minimální MAF 0,01. Wiggans *et al.* (2009) ve své studii uvádějí, že 6 572 SNP je monomorfních, 3 649 má  $MAF < 0,02$  a 40 847 SNP je možné použít pro předpověď. Jiang *et al.* (2010) uvádějí, že 11 008 SNP má  $MAF < 0,03$ .

Nejvyšší procenta vyřazených lokusů byla u všech vlastností na chromozomu X, což je pravděpodobně dáno tím, že lokusy na nehomologních úsecích nemohou být heterozygotní a je tak větší pravděpodobnost, že bude nižší hodnota MAF.

U dojivosti vycházela nejvyšší absolutní hodnota regresního koeficientu 7,472 kg. Szyda *et al.* (2011) předpovídali regresní koeficienty 46 267 SNP pro dojivost a uvádějí nejvyšší hodnotu regresního koeficientu 3,67 kg. Tento rozdíl může být způsoben vyšším počtem použitých SNP a vyšším počtem býků vstupujících do předpovědi v této diplomové práci a tím pádem i vyšší přesností výsledků. V neposlední řadě je potřeba říci, že se lišili i býci vstupující do předpovědi.

Nejvyšší absolutní hodnoty regresního koeficientu měl SNP na chromozomu 14 uvnitř genu DGAT1 (diacylglycerol O-acyltransferasa 1) a další SNP v blízkosti tohoto genu. Grisart *et al.* (2002) našli mutaci uvnitř tohoto genu, která má za důsledek záměnu lysinu (K) alaninem (A) a zjistili velikost substitučního efektu pro dojivost  $-158 \pm 24,5$  kg (DRP počítáno z poloviny PH). Signorelli *et al.* (2009) publikovali substituční efekt v tomto genu - 2,4 kg/den. Mach *et al.* (2012) uvádějí denní dojivosti 24,8 kg, 20,8 kg a 22,6 kg u genotypů AA, KA a KK. Vysoké regresní koeficienty uvnitř tohoto genu odpovídají dříve publikovaným QTL na chromozomu 14 (Khatkar *et al.*, 2004; Boichard *et al.*, 2003). Jiang *et al.* (2010) publikoval 40 statisticky významných SNP pro dojivost ( $P < 0,05$ ), 11 z nich je shodných s lokusy v Tab. 4. Vysoké hodnoty regresních koeficientů pro dojivost byly zjištěny i na chromozomu X, což však může být způsobeno nižší hustotou SNP ve srovnání s ostatními chromozomy. Na ostatních chromozomech vycházely regresní koeficienty nižší.

Genetické rozptyly odpovídají výši regresního koeficientu a heterozygotnosti konkrétního SNP a tím pádem SNP s nejvyššími hodnotami substitučních efektů mají zároveň i nejvyšší hodnoty genetických rozptylů, protože mezi lokusy nebyly výrazné odchylky v heterozygotnosti. Pokud se však sečtou genetické rozptyly všech lokusů činí celkový genetický rozptyl u dojivosti 6931, což je několikanásobně méně, než lze očekávat. Obdobná

situace je u zbývajících vlastností. Tento fakt je pravděpodobně způsoben tím, že vedle genetických rozptylů jednotlivých lokusů existují i interlokusové genetické kovariance.

U procentického zastoupení bílkovin byl SNP s nejvyšší absolutní hodnotou předpovězeného regresního koeficientu shodný s doživostí. Regresní koeficient tohoto SNP pro procentický obsah bílkovin činil  $19,090 \times 10^{-4}$  %. Grisart *et al.* (2002) uvádějí substituční efekt  $0,04 \pm 0,006$  % (DRP odvozeno od poloviny PH). Mach *et al.* (2012) uvádějí průměrné procentický obsah bílkovin 4,16 % (AA), 4,34 % (KA) a 4,61 % (KK).

Na chromozomu 20 měly vysoké absolutní hodnoty předpovězených regresních koeficientů SNP uvnitř a v blízkosti genu pro receptor růstového hormonu (GHR). Signorelli *et al.* (2009) publikovali substituční efekty několika SNP v rámci tohoto genu, statisticky průkazné byly dva SNP na hladině významnosti 0,07, se substitučním efektem 0,1 %.

Vysoké hodnoty předpovězených regresních koeficientů pro procentický obsah bílkovin měly rovněž lokusy na chromozomu 3, uvnitř a v blízkosti genu EFNA1, u kterého se však v literatuře nepovedlo dohledat spojitost s procentickým zastoupením bílkovin. 7 SNP v Tab. 9. je shodných s lokusy, které publikoval Jiang *et al.* (2010).

U procentického zastoupení tuku byl SNP s nejvyšší hodnotou předpovězeného regresního koeficientu shodný s předchozími vlastnostmi. U této vlastnosti byl však regresní koeficient SNP uvnitř genu DGAT1 v poměru k průměrné hodnotě regresního koeficientu napříč všemi lokusy podstatně vyšší (36,2 krát) než u předchozích vlastností. Grisart *et al.* (2002) zdůvodňují tento fakt tím, že se jedná o gen pro tvorbu enzymu, který přímo zasahuje do syntézy triglyceridů a uvádějí velikost substitučního efektu  $0,17 \pm 0,012$  % (DRP počítáno z poloviny PH). Signorelli *et al.* (2009) uvádějí substituční efekt v tomto lokusu dokonce 0,55 %. Vysoké regresní koeficienty měly dále SNP na chromozomu 5, které se nenacházely uvnitř žádného genu, nejbliže se nacházel gen MGST1.

Při předpovědi PGH a GEPH byly u všech vlastností zjištěny vysoké korelace se vstupními DRP a PH. U doživosti vycházejí korelace nižší než u procentického zastoupení tuku a bílkovin, což je dáno nižším koeficientem dědivosti u této vlastnosti. S rostoucí spolehlivostí předpovědi PH byků rostou i korelace mezi DRP a PGH a mezi PH a GEPH. Absolutní hodnota rozdílu mezi DRP a předpovězenou PGH s rostoucí spolehlivostí předpovědi PH byků klesá. Směrodatné odchylky PGH a GEPH byly nižší než směrodatné odchylky PH a PGH.

Vysoké korelace a nízké hodnoty absolutního rozdílu mezi vstupními DRP a výstupními PGH, především u byků s vysokou spolehlivostí předpovědi PH, poukazuje na skutečnost, že regresní koeficienty, které jsou použity pro zpětnou předpověď PGH

jsou předpovězeny s vysokou přesností. Vzhledem k tomu, že v populacích skotu probíhá intenzivní šlechtění, dochází i ke změnám regresních koeficientů v čase. Při použití metody RR-BLUP pro předpověď GEPH je proto podle Schaeffera (2006) potřeba regresní koeficienty nově předpovídat minimálně jednou za dva roky.



## 7 Závěr

Byly předpovězeny regresní koeficienty SNP pro doživost, procentický obsah bílkovin a tuku. Nejvyšší absolutní hodnotu regresního koeficientu měl u všech tří vlastností SNP uvnitř genu DGAT1 na chromozomu 14, který je v odborných publikacích uváděn jako QTL pro tyto vlastnosti. U procentického zastoupení bílkovin byl zjištěn vysoký regresní koeficient i u SNP uvnitř genu GHR na chromozomu 20.

Vysoké korelace mezi vstupními DRP a předpovězenými přímými PGH, stejně jako vysoké korelace mezi PH a GEPH býků vstupujících do předpovědi, svědčí o vysoké přesnosti předpovědi regresních koeficientů a možnosti jejich použití pro předpověď GEPH mladých zvířat bez vlastní užítkovosti nebo potomstva, která bude přesnější než předpověď rodokmenové PH. Vědecká hypotéza, stanovená na počátku této práce, byla tedy potvrzena a cíl splněn.

## 8 Seznam literatury

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., Lawlor, T. J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal Dairy Science*. 93 (2). 743-752.
- Arranz, J. J., Coppieters, W., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mezer, C., Riquet, J., Simon, P., Vamanshoven, P., Wagenaar, D., Georges, M. 1998. A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. *Animal Genetics*. 29 (2). 107-115.
- Banos, G., Woolliams, J. A., Woodward, B. W., Forbes, A. B., Coffey, M. P. 2008. Impact of Single Nucleotide Polymorphisms in Leptin, Leptin Receptor, Growth Hormone Receptor, and Diacylglycerol Acyltransferase (DGAT1) Gene Loci on Milk Production, Feed, and Body Energy Traits of UK Dairy Cows. *Journal of Dairy Science*. 91 (8). 3190-3200.
- Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M. Y., Leveziel, H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution*. 35 (1). 77-101.
- Bouška, J., Doležal, O., Jílek, F., Kudrna, V., Kvapilík, J., Příbyl, J., Rajmon, R., Sedmíková, M., Skřivanová, V., Šlosárková, S., Tyrolová, Y., Vacek, M., Žižlavský, J. 2006. Chov dojeného skotu. Profi Press, s. r. o. Praha. 186 s. ISBN: 8086726169.
- Calus, M. P. L., Veerkamp, R. F. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics*. 124 (6). 362-368.
- Christensen, O. F., Lund, M. S. 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42 (2).
- Cole, J. B., VanRaden, P. M., O'Connell, J. R., Van Tassell, C. P., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., Wiggans, G. R. 2009. Distribution and location of genetic effects for dairy traits. *Journal of Dairy Science*. 92 (6). 2931-2946.

- Daetwyler, H. D., Schenkel, F. S., Sargolzael, M., Robinson, J. A. B. 2008. A Genome Scan to Detect Quantitative Trait Loci for Economically Important Traits In Holstein Cattle Using Two Methods and Dense Single Nucleotide Polymorphism Map. *Journal of Dairy Science*. 91 (8). 3225-3236.
- Dale Van Vleck, L., Dwyer, D. J. 1985. Successive Overrelaxation, Block Iteration, and Method of Conjugate Gradients for Solving Equations for Multiple Trait Evaluation of Sires. *Journal of Dairy Science*. 68 (3). 760-767.
- Ducrocq, V., Liu, Z. 2009. Combining Genomic and Classical Information in National BLUP Evaluations. *Interbull Bulletin*. 40. 172-177.
- Fikse, W. F., Baanos, G. 2001. Weighting Factors of Sire Daughter Information in International Genetic Evaluations. *Journal of Dairy Science*. 84 (7). 1759-1767.
- Forni, S., Aguilar, I., Misztal, I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution*, 43 (1).
- Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X., Womack, J. E., Hoeschele, I. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*. 139 (2). 907-920.
- Goddard, M. 1985. A method of comparing sires evaluated in different countries. *Livestock Production Science*. 13 (4). 321-331.
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. Snell, R. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DFAT1 gene with major effect on milk yield and composition. *Genome Research*. 12 (2). 222-231.
- Hayes, B. J., Browman, P. J., Chamberlain, A. J., Goddard, M. E. 2009. Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 92 (2). 433-443

- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 31 (2). 423-447.
- Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*. 32 (1). 69-83.
- Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., Da, Y., Wiggans, G. R., VanRaden, P. M., Lewin, H. A. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological Genomics*. 1 (3). 165-175.
- Illumina. BovineSNP50 Genotyping BeadChip [online]. Illumina. 2012. [cit. 2013-8-8]. Dostupné z <[http://res.illumina.com/documents/products/datasheets/datasheetbovine\\_snp50.pdf](http://res.illumina.com/documents/products/datasheets/datasheetbovine_snp50.pdf)>.
- Illumina. Infinium HD Assay. [online]. Illumina. 2014. [cit. 2014-1-2]. Dostupné z <[http://www.illumina.com/technology/infinium\\_hd\\_assay.ilmn](http://www.illumina.com/technology/infinium_hd_assay.ilmn)>.
- Interbull. 1990. Recommended procedures for international use of sire proofs. *Interbull bulletin*. 4.
- Jairath, L., Dekkers, J. C. M., Schaeffer, L. R., Liu, Z., Burnside, E. B., Kolstad., B. 1998. Genetic Evaluation for Herd Life in Canada. *Journal of Dairy Science*. 81 (2). 550-562.
- Jakubec, V., Bezdíček, J., Louda, F. 2010. Selekce – Inbríding - Hybridizace. *Agrovýzkum Rapotín, s. r. o. Rapotín*. 382 s. ISBN: 978-80-87144-22-0.
- Jakubec, V., Říha, J., Golda, J., Majzlík, I. 1999. Odhad plemenné hodnoty hospodářských zvířat. *Asociace chovatelů masných plemen. VÚCHS Rapotín, Česká zemědělská univerzita. Rapotín*. 177 s.
- Jiang, L., Jianfeng, L., Dongxiao, S., Peipei, M., Xiandong, D., Ying, Y., Zhang, Q. 2010. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. *PLOS ONE*. 5.

- Jiménez-Montero, J. A., González, O., Alenda, R. 2013. Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle. *Journal of Dairy Science*. 96 (1). 625-634.
- Kashi, X., Hallerman, E., Soller, M. 1990. Marker-assisted selection of candidate bulls for progeny testing programmes. *Animal production*. 51 (1). 63-74.
- Khatkar, M. S., Thomson, P. C., Tammen, I., Raadsma, H. W. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genetics Selection Evolution*. 36 (2). 163-190.
- Legara, A., Misztal, I. 2008. Technical note: Computing Strategies in Genome-Wide Selection. *Journal of Dairy Science*. 91 (1). 360-366.
- Lidauer, M., Strandén, I., Mäntysaari, E. A., Pösö, J., Kettunen, A. 1999. Solving Large Test-Day Models by Iteration on Data and Preconditioned Conjugate Gradient. *Journal of Dairy Science*. 82 (12). 2788-2796.
- Liu, Z., Reinhardt, F., Bünger, A., Reents, R. 2004. Derivation and calculation of approximate reliabilities and daughter deviations of random regression test-day model for genetic evaluation of dairy cattle. *Journal of Dairy Science*. 87 (6). 1896-1907.
- Liu, Z., Reinhardt, F., Reents, R. 2004. A multi-trait MACE model for international bull comparison based on daughter yield deviations. *Interbull bulletin*. 32. 46-52.
- Mach, N., Blum, Y., Bannik, A., Causeur, D., Houee-Bigot, M., Lagarrigue, S., Smits, M. A. 2012. Pleiotropic effects of polymorphism of the gene diacylglycerol-O-transferase 1 (DGAT1) in the mammary gland tissue of dairy cows. *Journal of Dairy Science*. 95 (9). 4989-5000.
- Meuwissen, T. H. E., Goddard, M. E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics Selection Evolution*. 36 (3). 261-279.
- Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157 (4). 1819-1829.

- Misztal, I., Legarra, A., Aguilar, I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*. 92 (9). 4648-4655.
- Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. CAB international. Wallingford. p. 344. ISBN: 0851990002.
- Pimentel, E. C. G., Simianer, H., König, S. 2011. Effective Number of Genes and Accuracy of Genomic Evaluations. *Interbull Bulletin*. 44. 122-125.
- Plante, Y., Gibson, J. P., Nadesalingam, J., Mehrabani-Yeganeh, H., Lefebvre, S., Vandervoort, G., Jansen, G. B. 2001. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *Journal of Dairy Science*. 84 (6). 1516-1524.
- Plemdat. 2010. Popis modelu pro odhady PH mléčné užitkovosti [online]. Benešov. Plemdat, s.r.o. [cit. 2013-3-11]. Dostupné z <[http://www.plemdat.cz/cz/pages/Popis\\_mleko.pdf](http://www.plemdat.cz/cz/pages/Popis_mleko.pdf)>.
- Příbyl, J. 1995. Využití markerů při selekci hospodářských zvířat. *Živočišná výroba*. 40 (8). 375-382
- Příbyl, J., Haman, J., Kott, T., Příbylová, J., Šimečková, M., Vostrý, L., Zavadilová, L., Čermák, V., Růžička, Z., Šplíchal, J., Verner, M., Motyčka, J., Vondrášek, L. 2012. Single-step prediction of genomic breeding value in a small dairy cattle population with strong import of foreign genes. *Czech Journal of Animal Science*. 57 (4). 151-159.
- Ron, M., Heyen, D. W., Weller, J. I., Band, M., Feldmesser, E., Pasternak, H. D. Y., Wiggans, G. R., VanRaden. P. M., Ezra, E., Lewin, H. A. 1998. Detection and analysis of a locus affecting milk concentration in the US and Israeli dairy cattle populations. In: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, Vol. 26. University of New-England. Armindale. 422-425.
- Ron, M., Kliger, D., Feldmesser, E., Seroussi, E., Ezra, E., Weller, J. I., 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by daughter design. *Genetics*. 159 (2). 727-735.

- Ruane, J., Colleau, J. J. 1994. The value of marker assisted selection when QTL is marked. In: 45th Annual meeting EAAP.
- Schaeffer, L. R. 1994. Multiple-Country Comparison of Dairy Sires. *Journal of Dairy Science*. 77 (9). 2671-2678.
- Schaeffer, L. R. 2001. Multiple trait international bull comparisons. *Livestock Production Science*. 69 (2). 145-153.
- Schaeffer, L. R. 2006. Strategy for applying genom-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 123 (4). 218-223.
- Signorelli, F., Orru, L., Napolitano, F., De Matteis, G., Scata, M. C., Catilo, G., Marchitelli, C., Moioli, B. 2009. Exploring polymorphisms and effects on milk traits on the DGAT1, SCD1 and GHR genes in four cattle breeds. *Livestock Science*. 125. 74-79.
- Sigurdsson, A., Banos, G. 1995. Dependent variables in International sire evaluation. *Acta Agriculturae Scandinavica*. 45 (4). 209-217.
- Spelman, R. J., Coppieters, W., Karim, L., van Arendok, J. A. M., Bovenhuis, H. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in Dutch Holstein-Friesian population. *Genetics*. 144 (4). 1799-1808.
- Sullivan, P. G., Wilton, J. W. 2001. Multiple-trait MACE with a variable number of traits per country. *Interbull Bulletin*. 27. 68-72.
- Szyda J., Zarnecki, A., Suchocki, T., Kaminski, S. 2011. Fitting and validating the genomic evaluation model to Polish Holstein-Friesian cattle. *Journal of Applied Genetics*. 52 (3). 363-366.
- Szyda, J., Ptak, E., Komisarek, J., Zarnecki A. 2008. Practical plication of daughter yield deviations in cattle breeding. *Journal of Applied Genetics*. 49 (2). 183-191.
- Tsuruta, S., Misztal, I., Strandén, I. 2001. Use of preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. *Journal of Animal Science*. 79 (5). 1166-1172.

- Van Vleck, L. D., Dwyer, D. J. 1985. Successive Overrelaxation, Block Iteration, and Method of Conjugate Gradients for Solving equations for Multiple Trait Evaluation of Sires. *Journal of Dairy Science*. 68 (3). 760-767.
- VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*. 91 (11). 4414-4423.
- VanRaden, P. M., Sullivan, P. G. 2010. International genomic evaluation methods for dairy cattle. *Genetics Selection Evolution*. 42 (7).
- VanRaden, P. M., Wiggans, G. R. 1991. Derivation, calculation, and use of national model information. *Journal of Dairy Science*. 74 (8). 2737-2746.
- Weigel, K. A., Lin, S. W. 1999. Development of International Conversion Equations Using Robust Regression Methodology. *Journal of Dairy Science*. 82 (9). 2023-2029.
- Wiggans, G. R., Sonstegard, T. S., VanRaden, P. M., Matukumalli, L. K., Schnabel, R. D., Taylor, J. F., Schenkel, F. S., Van Tassell, C. P. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *Journal of Dairy Science*. 97 (7). 3431-3436.
- Wilmink, J. B. M., Meijering, A., Engel, B. 1986. Conversion of breeding values for milk from foreign populations. *Livestock Production Science*. 14 (3). 223-229.
- Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfistergenskow, M., Witte, L. A., Grigola, F. E., Uimari, P., Thaller, G., Bishop, M. D. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in large outbred pedigree. *Genetics*. 149 (4). 1959-1973.



## 9 Seznam použitých zkratek a symbolů

BLUP – best linear unbiased prediction – nejlepší lineární nevychýlená předpověď

DNA – deoxyribonukleová kyselina

DRP – deregressed proof – odregresovaná plemenná hodnota

DYD – daughter yield deviation – odchylka v užítkovosti dcer

EDC – effective daughter contribution – efektivní počet dcer

ERC – effective record contribution – efektivní počet vlastních užítkovostí

GEPH – genomická plemenná hodnota

MACE - multiple across country evaluation

MAF – minor allele frequency – četnost méně četné alely

MAS – marker assisted selection – selekce s využitím markerů

MT-MACE – multi trait multiple across country evaluation – víceznaková metoda MACE

PGH – přímá genetická hodnota

PH – plemenná hodnota

QTL – quantitative trait loci – lokus ovlivňující kvantitativní vlastnost

RR-BLUP – ridge regression best linear unbiased prediction – hřebenová regrese – BLUP s  
přičtením hodnot na diagonálu soustavy

SNP – single nucleotide polymorphism – bodová (jednonukleotidová) mutace v DNA

ST-MACE – single trait multiple across country evaluation- jednoznaková metoda MACE

YD – yield deviation – odchylka ve vlastní užítkovosti