



PREDYKCJA GENOMOWEJ WARTOŚCI HODOWLANEJ

Dawid Słomian



1. Model predykcji
2. Rutynowa ocena genomowej wartości hodowlanej w Polsce
3. Ocena genomowej wartości hodowlanej na świecie

Ocena konwencjonalna vs. ocena genomowa

- Ocena **konwencjonalna** → wydajności + spokrewnienie
 - Niedokładna dla **młodych buhajów** – mała liczba córek / brak córek
- Ocena **genomowa** → genotypy SNP
 - Przyspieszenie oceny + wzrost dokładności oceny wartości hodowlanej dla **młodych buhajów**

Genotypy SNP

[Header]

BSGT Version	3.2.32
Processing Date	11/24/2008 10:14 AM
Content	BovineSNP50_A.bpm
Num SNPs	54001
Total SNPs	54001
Num Samples	32
Total Samples	1054

[Data]

SNP Name	Sample ID	GC Score	SNP Index	Allele1 - AB	Allele2 - AB	Chr	Position	GT Score
ARS-BFGL-BAC-10172	4408169492_K	0.883	1B	B			14 4736993	0.849
ARS-BFGL-BAC-1020	4408169492_K	0.899	2B	B			14 6339014	0.8626
ARS-BFGL-BAC-10245	4408169492_K	0.6582	3B	B			14 30073020	0.71
ARS-BFGL-BAC-10345	4408169492_K	0.9092	4A	B			14 4497877	0.8721
ARS-BFGL-BAC-10365	4408169492_K	0.8021	5B	B			14 25140301	0.833
ARS-BFGL-BAC-10375	4408169492_K	0.8858	6A	B			14 4983527	0.8513
ARS-BFGL-BAC-10591	4408169492_K	0.867	7A	B			14 15446975	0.8363
ARS-BFGL-BAC-10793	4408169492_K	0.8722	8B	B			14 27452258	0.8403
ARS-BFGL-BAC-10867	4408169492_K	0.9316	9A	B			14 32700054	0.8949
ARS-BFGL-BAC-10919	4408169492_K	0.7805	10A	B			14 29520816	0.778
ARS-BFGL-BAC-10952	4408169492_K	0.9314	11B	B			10 19315327	0.8947
ARS-BFGL-BAC-10960	4408169492_K	0.6543	12B	B			10 21056606	0.7079
ARS-BFGL-BAC-10975	4408169492_K	0.8622	13A	B			10 21682679	0.8358
ARS-BFGL-BAC-10986	4408169492_K	0.8687	14A	B			10 25897020	0.8376
ARS-BFGL-BAC-10993	4408169492_K	0.8146	15A	B			10 80403647	0.7993
ARS-BFGL-BAC-11000	4408169492_K	0.9135	16A	A			10 81191638	0.8762

Predykcja genomowej wartości hodowlanej

① estymacja efektów SNP → **SNP-BLUP**

Model predykcji efektów SNP – zbiór referencyjny

- Ocena efektów SNP (σ **zbiór referencyjny / treningowy**)
- Osobniki posiadające wartości hodowlane **i** genotypy SNP
- SNP-BLUP

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_q\mathbf{q} + \mathbf{Z}_a\mathbf{a} + \mathbf{e}$$

Model predykcji efektów SNP – zbiór referencyjny

- Ocena efektów SNP (♂ zbiór referencyjny)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_q \mathbf{q} + \mathbf{Z}_a \mathbf{a} + \mathbf{e}$$

— $\mathbf{y}$ DRP - wartości hodowlane po deregresji

— $\boldsymbol{\beta}$ efekty stałe

— $\mathbf{q}$ efekt SNP $\mathbf{q} \sim N(\mathbf{0}, \mathbf{G})$

— $\mathbf{G} = \begin{bmatrix} 1 & & \text{"0"} \\ & \dots & \\ \text{"0"} & & 1 \end{bmatrix} \frac{(1-w)\sigma_a^2}{N_{SNP}}$

— $\mathbf{Z}_q \in \{-1, 0, 1\}$

Model predykcji efektów SNP – zbiór referencyjny

- Ocena efektów SNP (♂ zbiór referencyjny)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_q\mathbf{q} + \mathbf{Z}_a\mathbf{a} + \mathbf{e}$$

— $\mathbf{a}$ resztowy efekt poligeniczny $\mathbf{a} \sim N(\mathbf{0}, \mathbf{A})$

$$\text{— } \mathbf{A} = \begin{bmatrix} a_{ii} & a_{ij} & a_{ij} \\ a_{ji} & a_{ii} & a_{ij} \\ a_{ji} & a_{ji} & a_{ii} \end{bmatrix} w \sigma_a^2$$

— $w \in \{ 0, 0.05, 0.10, \text{ itp. } \}$

— $\mathbf{Z}_a \in \{ 0, 1 \}$

Model predykcji efektów SNP – zbiór referencyjny

- Ocena efektów SNP (σ zbiór referencyjny)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_q\mathbf{q} + \mathbf{Z}_a\mathbf{a} + \mathbf{e}$$

– $\mathbf{e}$ błąd $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$

$$\text{– } \mathbf{R} = \begin{bmatrix} \frac{1}{EDC_i} & & \text{"0"} \\ & \frac{1}{EDC_i} & \\ \text{"0"} & & \frac{1}{EDC_i} \end{bmatrix} \sigma_e^2$$

- Ocena efektów **SNP** (♂ zbiór referencyjny)

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z_q & X^T R^{-1} Z_a \\ Z_q^T R^{-1} X & Z_q^T R^{-1} Z_q + G^{-1} & Z_q^T R^{-1} Z_a \\ Z_a^T R^{-1} X & Z_a^T R^{-1} Z_q & Z_a^T R^{-1} Z_a + A^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ q \\ a \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z_q^T R^{-1} y \\ Z_a^T R^{-1} y \end{bmatrix}$$

- iteration on data with residual update
- Legarra i Misztal, 2008

Predykcja genomowej wartości hodowlanej

① estymacja efektów SNP → **G-BLUP**

Model predykcji efektów SNP – zbiór referencyjny

- G-BLUP

$$y = X\beta + Zg + e$$

- y DRP
- β efekty stałe
- g DGV $g \sim N(\mathbf{0}, G)$ → efekt osobnika
- $G = \frac{ZZ^T}{2 \sigma_{i=1}^{N_{SNP}} (p_i \cdot q_i)}$ → VanRaden, 2008
- Z macierz wystąpień
- e błąd $e \sim N(\mathbf{0}, R)$

Obliczanie bezpośredniej wartości hodowlanej

- Ocena efektów **osobników** (σ zbiór referencyjny)

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ g \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

- Obliczenie efektów SNP $\rightarrow$ Liu et al. 2014

$$\hat{q} = \frac{Z A^{-1} \hat{g}}{N_{SNP} + Z A^{-1} Z^T}$$

Predykcja genomowej wartości hodowlanej

②

Obliczanie bezpośredniej wartości hodowlanej – zbiór testowy

- Obliczanie **bezpośredniej** wartości hodowlanej
- Źródło informacji → SNP
- **DGV**
- Direct Genomic Value
- **Zbiór testowy** → Wszystkie ♂ i ♀ posiadające genotypy SNP

$$DGV = X\hat{\beta} + Z_q\hat{q} + e$$

Obliczanie bezpośredniej wartości hodowlanej

- **Aproksymacja dokładności** genomowej wartości hodowlanej

$$y = b_0 + b_1 d + e$$

- y DRP
- b_0 wyraz wolny
- b_1 współczynnik regresji
- d DGV
- e błąd

$R^2 \rightarrow$ dokładność DGV

Jednakowa dokładność dla każdego osobnika

Obliczanie **genomowej** wartości hodowlanej

- Obliczanie **kombinowanej** wartości hodowlanej
- Źródła informacji → SNP i EBV
- **GEBV**
- Genomically enhanced breeding value
 - ♂ zbiór referencyjny

$$GEBV = \begin{bmatrix} r_{DGV} & r_{EBV} \end{bmatrix} \begin{bmatrix} r_{DGV} & r_{DGV}r_{EBV} \\ r_{DGV}r_{EBV} & r_{EBV} \end{bmatrix}^{-1} \begin{bmatrix} DGV \\ EBV \end{bmatrix}$$

- ♂ i ♀ zbiór testowy

$$GEBV = \begin{bmatrix} r_{DGV} & r_{PI} \end{bmatrix} \begin{bmatrix} r_{DGV} & r_{DGV}r_{PI} \\ r_{DGV}r_{PI} & r_{PI} \end{bmatrix}^{-1} \begin{bmatrix} DGV \\ PI \end{bmatrix}$$

Rutynowa ocena genomowej wartości hodowlanej w Polsce

Wyniki oceny rutynowej

Rutynowa ocena genomowa → Instytut Zootechniki

Lp.	Buhaj	ML[kg]	ML[kg] powt.	Tł[kg]	Tł[kg] powt.	% Tł	% Tł powt.	BI[kg]	BI[kg] powt.	% BI	% BI powt.	IP[kg]	LA[kg]	LA[kg] powt.	% LA	% LA powt.
1	ALTAKADERO US3142181280 840M003142181280	2153 MACE	0.88	33.7 MACE	0.87	-0.55 MACE	0.87	56.8 MACE	0.84	-0.14 MACE	0.84	147.3	90.3 EBV	0.62	-0.01 EBV	0.62
2	CICERO DE000358665633 DEUM000358665633	2081 MACE	0.86	54.1 MACE	0.86	-0.31 MACE	0.86	50.5 MACE	0.82	-0.18 MACE	0.82	155.2	71.4 EBV	0.59	0.01 EBV	0.59
3	BRODIE CA11857459 CANM000011857459	1939 MACE	0.98	31.7 MACE	0.98	-0.49 MACE	0.98	43.9 MACE	0.98	-0.21 MACE	0.98	119.4	66.6 EBV	0.97	-0.25 EBV	0.97
4	MELVIEW UK560989202462 GBRM560989202462	1935 MACE	0.84	18.5 MACE	0.85	-0.63 MACE	0.85	51.3 MACE	0.80	-0.13 MACE	0.80	121.1	71.9 EBV	0.54	-0.04 EBV	0.54
5	MANOSO DE001267420197 DEUM001267420197	1887 MACE	0.86	50.0 MACE	0.86	-0.28 MACE	0.86	51.1 MACE	0.82	-0.11 MACE	0.82	152.3	83.2 EBV	0.61	-0.09 EBV	0.61
6	M - LECHÉ US3010353594 840M003010353594	1874 MACE	0.93	41.9 MACE	0.92	-0.36 MACE	0.92	35.0 MACE	0.91	-0.28 MACE	0.91	112.0	82.5 EBV	0.82	0.01 EBV	0.82
7	STEP RED DE000356650139 DEUM000356650139	1848 MACE	0.96	23.6 MACE	0.95	-0.54 MACE	0.95	45.9 MACE	0.95	-0.16 MACE	0.95	115.4	85.0 EBV	0.92	-0.05 EBV	0.92
8	CINEMA DK003372306258 DNKM000000256828	1840 MACE	0.94	50.3 MACE	0.94	-0.26 MACE	0.94	47.9 MACE	0.93	-0.13 MACE	0.93	146.1	84.9 EBV	0.89	-0.06 EBV	0.89
9	TUCKER US3125066425 840M003125066425	1792 MACE	0.91	8.8 MACE	0.90	-0.68 MACE	0.90	29.5 MACE	0.89	-0.31 MACE	0.89	67.8	85.8 EBV	0.85	0.02 EBV	0.85
10	CHICO US74700496 USAM000074700496	1791 MACE	0.89	44.7 MACE	0.89	-0.30 MACE	0.89	42.2 MACE	0.86	-0.18 MACE	0.86	129.2	85.8 EBV	0.74	-0.02 EBV	0.74
11	LAIDLAW US3014365477 840M003014365477	1790 MACE	0.86	40.3 MACE	0.86	-0.35 MACE	0.86	47.5 MACE	0.82	-0.12 MACE	0.82	135.3	62.8 EBV	0.53	0.02 EBV	0.53
12	BROKAT MLEK PL005391463182 POLM005391463182	1744 GMACE	0.74	30.9 GMACE	0.69	-0.43 GMACE	0.69	37.1 GMACE	0.72	-0.21 GMACE	0.72	105.2	73.1 GPI	0.72	-0.10 GPI	0.72
13	MERRYGUY US3138483358 840M003138483358	1742 MACE	0.86	58.5 MACE	0.87	-0.13 MACE	0.87	47.2 MACE	0.82	-0.11 MACE	0.82	152.9	54.3 EBV	0.50	-0.06 EBV	0.50
14	SUPREM DE000360385508 DEUM000360385508	1736 MACF	0.84	53.0 MACF	0.85	-0.19 MACF	0.85	53.8 MACF	0.80	-0.03 MACF	0.80	160.7	- -	- -	- -	- -

Wyniki oceny rutynowej

Rutynowa ocena genomowa → Instytut Zootechniki

Ocena wartości hodowlanej buhajów

[Strona główna](#)[Wyniki oceny](#)[Metody oceny](#)[Równania konwersji](#)[Publikacje](#)[Linki](#)

Buhaj **BROKAT MLEK** - wyniki oceny wartości hodowlanej - ocena 2022/3 (grudzień 2022, data publikacji oceny: 07.12.2022)

BUHAJ

Nazwa	BROKAT MLEK
Numer	PL005391463182
Numer międzynarodowy	POLM005391463182
Data urodzenia	2019-05-31
Rasa odmiana	HO
Ojciec	BRAVENESS
Numer ojca	US003014562366
Numer międzynarodowy ojca	840M003014562366



INDEKS I PODINDEKSY

Indeks PF	135
Podindeks produkcyjny	126
Podindeks pokroju ogólny	124
Podindeks ramy ciała	129
Podindeks siły mleczności	115
Podindeks nóg i racic	114
Podindeks wymienia	116
Podindeks płodności	103

WARTOŚCI HODOWLANE

Cechy produkcyjne

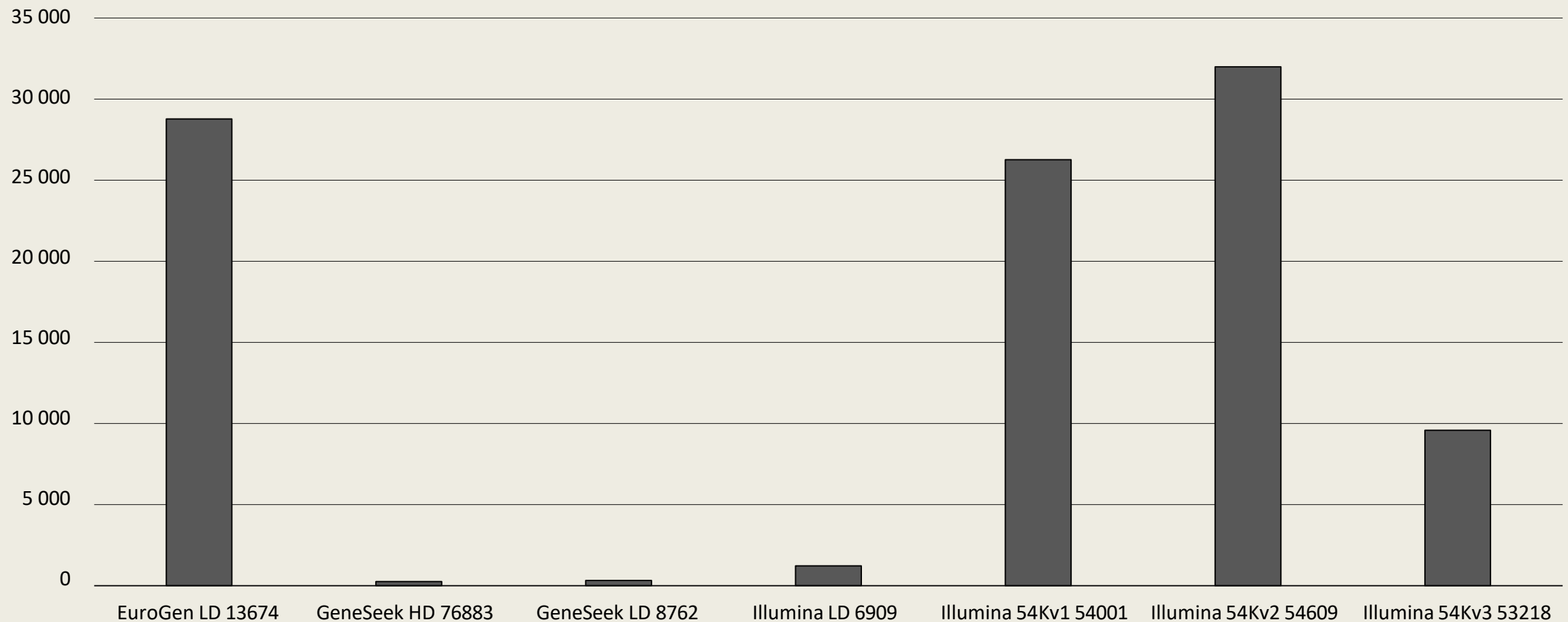
	Typ oceny	Wartość hodowlana	Powtarzalność	Liczba córek	Liczba obór
Wydajność mleka [kg]	GMACE	1744	0.74	0	0
Wydajność tłuszczu [kg]	GMACE	30.9	0.69	0	0
% tłuszczu	GMACE	-0.43	0.69	0	0
Wydajność białka [kg]	GMACE	37.1	0.72	0	0
% białka	GMACE	-0.21	0.72	0	0
Indeks produkcyjny [kg]	GMACE	105.2			
Wydajność laktozy [kg]	GPI	73.1	0.72	0	0
% laktoza	GPI	-0.10	0.72	0	0

Genomowa ocena rutynowa - serwer

- Baza danych genotypów w IZ
- Dedykowany serwer obliczeniowy IZ
- Oprogramowanie
 - bash shell
 - Fortran
 - SAS, Python, R
 - Findhap

Genomowa ocena rutynowa – baza danych SNP

Liczba zgenotypowanych osobników 04/2021
120 552



Genomowa ocena rutynowa – baza danych SNP

- Łatwy, szybki eksport/import danych dla populacji referencyjnej
- Łatwy, szybki eksport/import danych dla młodych osobników
- Rutynowy backup
- Automatyczna imputacja: wszystkie chipy → 50K
- Statystyki podsumowujące jakość genotypowania i imputacji
- Weryfikacja pochodzenia → SNP + standardy ICAR
- Porównywanie genotypów

Genomowa ocena rutynowa – nowe cechy

- wydajność laktozy,
- body conformation score
- temperament
- szybkość oddawania mleka
- łatwość wycieleń (direct & maternal)
- Indeks poronień (direct & maternal)

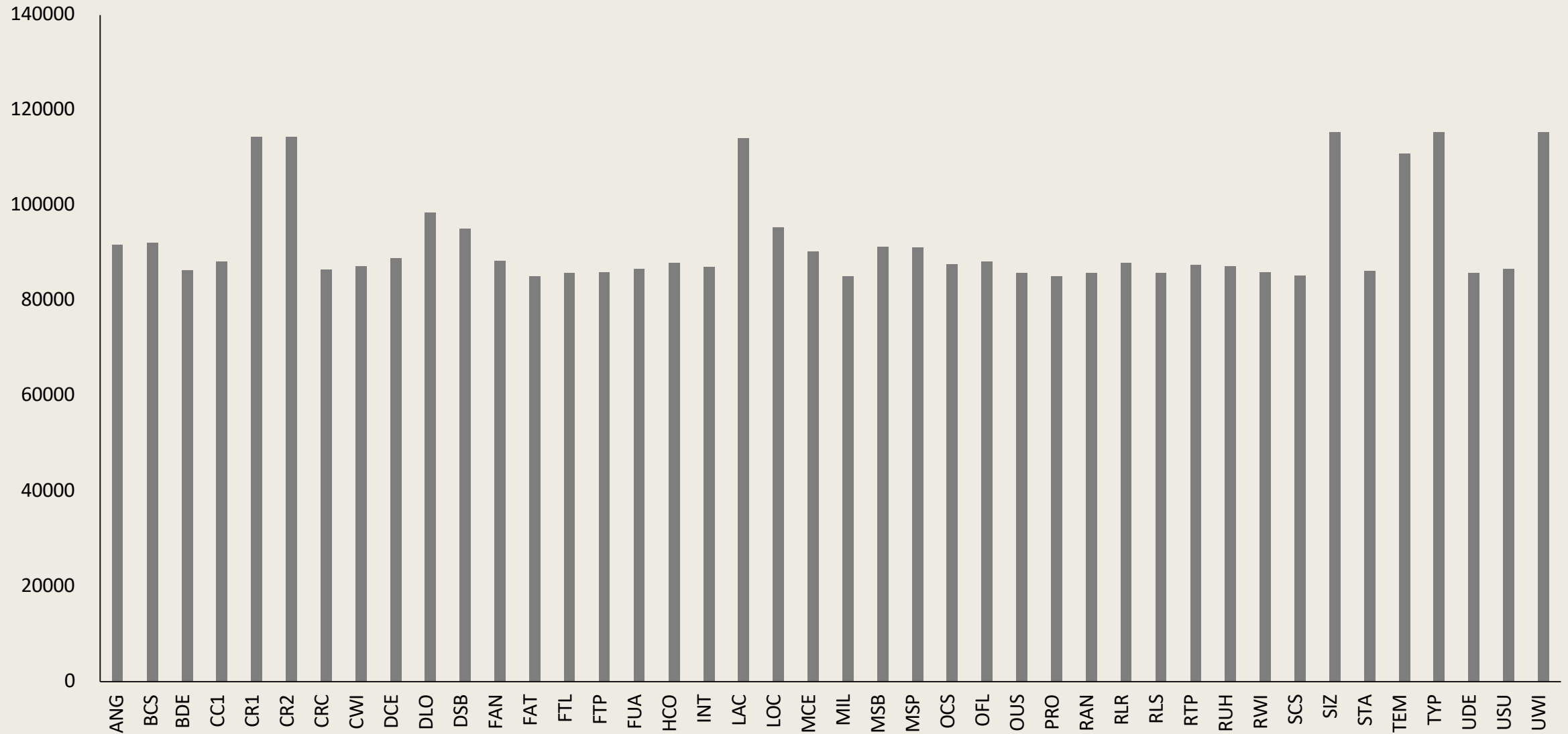
Genomowa ocena rutynowa – populacja referencyjna

Liczba osobników referencyjnych 04/2021



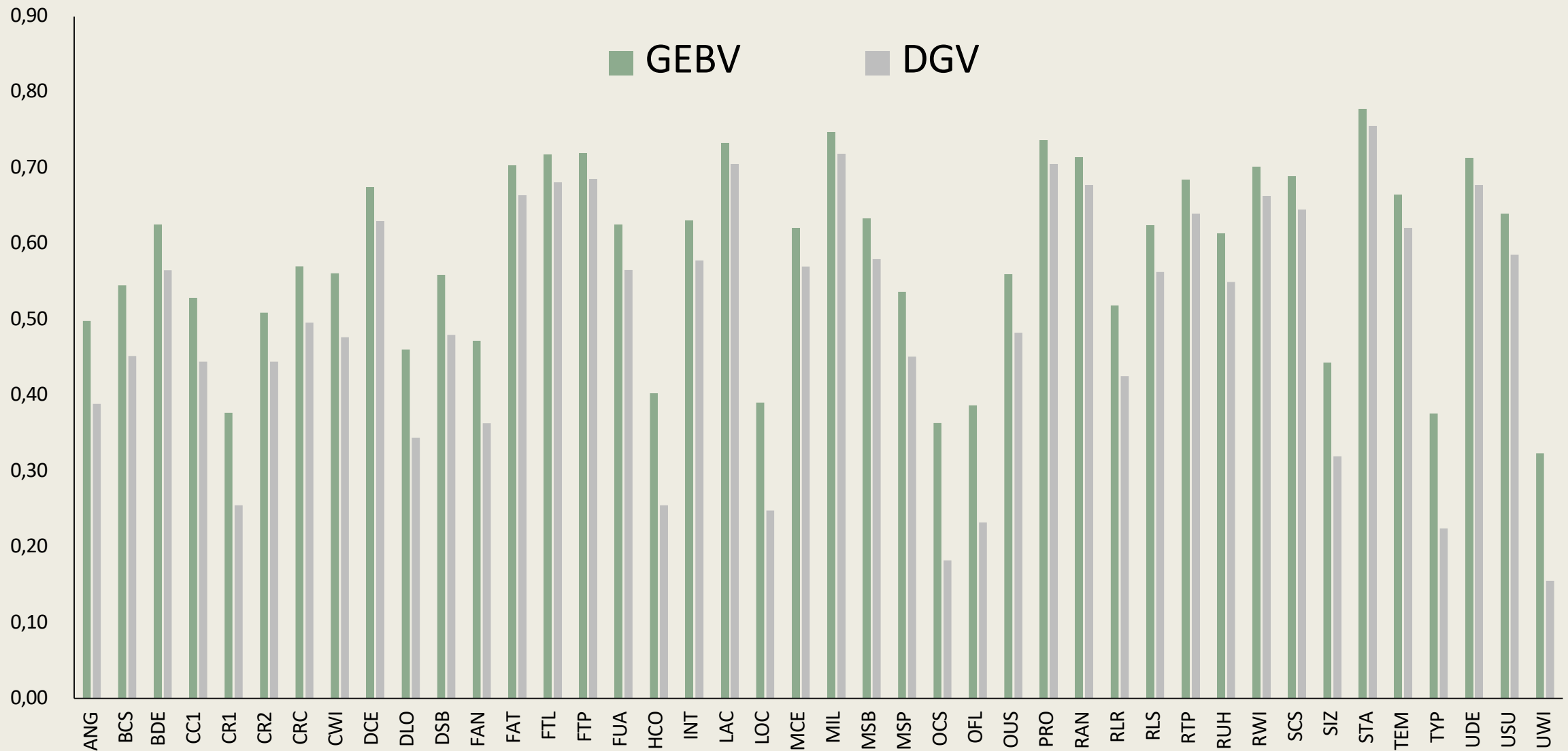
Genomowa ocena rutynowa – populacja testowa

Liczba osobników testowych 04/2021



Genomowa ocena rutynowa – populacja testowa

Dokładność predyktorów 04/2021



Genomowa ocena rutynowa – obliczenia rutynowe

- Ocena rutynowa wszystkich buhajów
 - 6 razy w roku
 - Przygotowanie i weryfikacja danych
 - Obliczanie efektów SNP
 - Obliczanie DGV, GEBV, dokładności
 - Przygotowanie danych do Interbull
 - Przygotowanie plików wynikowych dla hodowców

Genomowa ocena rutynowa – obliczenia rutynowe

- Oceny dodatkowe dla młodych buhajów
 - „na życzenie”
 - Przygotowanie i weryfikacja danych
 - ~~○ Obliczanie efektów SNP~~
 - Obliczanie DGV, GEBV, dokładności
 - ~~○ Przygotowanie danych do Interbull~~
 - Przygotowanie plików wynikowych dla hodowców

Genomowa ocena rutynowa – obliczenia nierutynowe

- Modyfikacje modelu oceny
- Walidacja Interbull
- Selekcja genomowa dla nowych cech
- Badania naukowe

Genomowa ocena rutynowa

CPU:

- Estymacja efektów SNP → ~ 21 min. – 3 godz. / cecha
- Pełna ocena → do 10 godz. / cecha

Ocena genomowej wartości hodowlanej na świecie

Selekcja genomowa na świecie - kraje

- Australia
- Belgia
- Czechy
- Kanada
- kraje skandynawskie
- Francja
- Niemcy + Austria
- Wielka Brytania
- Węgry
- Irlandia
- Włochy
- Japonia
- Nowa Zelandia
- Norwegia
- Polska
- Słowenia
- Hiszpania
- Szwajcaria
- USA
- Holandia

Selekcja genomowa na świecie - rasy

- Holsztyńsko-Fryzyjska
- Jersey
- Brown swiss
- Red
- Montbéliarde
- Normande
- Simmental
- Guernsey