



Übung 9 vom 04.12.2015

Genomische Zuchtwerte mit SNP-BLUP schätzen

Birgit Gredler-Grandl

Aufgabe – Genomische Zuchtwertschätzung SNP-BLUP

- Genomische Zuchtwertschätzung mit der Methode SNP-BLUP
- Referenzpopulation 325 Stiere mit Zuchtwerten für Eiweiss-%
- Genotypisierung für 10 SNPs
- Bei 31 Kälber stehen Selektionsentscheide an
- Kälber sind für 10 SNPs genotypisiert
- Für die Kälber sollen genomischen Zuchtwerte für Eiweiss-% geschätzt werden

Aufgabe – Genomische Zuchtwertschätzung

SNP-BLUP

- 1. SNP-Effekte schätzen

$$y = \mathbf{1}_n \mu + Xg + e \quad \Rightarrow \quad \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' X \\ X' \mathbf{1}_n & X' X + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' y \\ X' y \end{bmatrix}$$

- 2. Direkt genomische Zuchtwerte für die Kälber schätzen

$$DGZW = \sum_i^n X_i \hat{g}_i$$

SNP-BLUP in R

- Pfad setzen:
- `setwd("/Pfad-zu-den-Daten/")`
- Parameter setzen:
- `nmarkers<-10`
- `nrecords<-325`
- `lamda<-10`
- Daten einlesen:
- X-Matrix Referenzpopulation (Genotypen 325 Stiere)
- `x<-matrix(scan("xmatrix.txt"),ncol=nmarkers,byrow=T)`
- **Read 3250 items**
- Phenotypen Referenzpopulation:
- `y<-matrix(scan("yvec.txt"), byrow=T)`

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

SNP-BLUP in R

- `> dim(x)`
- `[1] 325 10`
- `> head(x)`
- `[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]`
- `[1,] 0 0 0 0 0 0 0 1 2 0 2`
- `[2,] 1 0 0 1 1 1 1 2 1 0 1`
- `[3,] 1 0 0 1 0 0 0 1 1 1 1`
- `[4,] 1 1 1 1 0 1 1 2 1 1 1`
- `[5,] 0 1 1 1 1 1 1 2 1 0 1`
- `[6,] 1 0 0 1 1 1 1 2 1 0 1`

SNP-BLUP in R

- `> dim(y)`
- `[1] 325 1`
- `> head(y)`
- `[,1]`
- `[1,] 0.186`
- `[2,] 1.226`
- `[3,] 0.864`
- `[4,] 1.228`
- `[5,] 0.446`
- `[6,] 1.125`

SNP-BLUP in R

- Einlesen der Kälberdaten:
- X-Matrix der 31 Kälber
- `> xkalb <- matrix(scan("x_kalb.dat"),ncol=nmarkers, byrow=T)`
- **Read 310 items**
- `> dim(xkalb)`
- [1] 31 10
- `> head(xkalb)`
- | | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] | [,8] | [,9] | [,10] |
|------|------|------|------|------|------|------|------|------|------|-------|
| [1,] | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 0 | 1 |
| [2,] | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 0 | 1 |
| [3,] | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 0 | 1 |
| [4,] | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 0 | 1 |
| [5,] | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 2 |
| [6,] | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 0 | 1 |

SNP-BLUP in R

- Vektor mit „1“:
- `ones<-array(1,c(nrecords))`
- `> dim(ones)`
- [1] 325
- `> head(ones)`
- [1] 1 1 1 1 1 1
- Einheitsmatrix I:
- `> ident_matrix <- diag(nmarkers)`
- `> dim(ident_matrix)`
- [1] 10 10
- `> head(ident_matrix)`
- [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
- [1,] 1 0 0 0 0 0 0 0 0 0
- [2,] 0 1 0 0 0 0 0 0 0 0

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Dimension der Koeffizientenmatrix?
- Dimensionen der einzelnen Submatrizen:
- $\mathbf{1}'\mathbf{1}$: 1×325 * $325 \times 1 \rightarrow 1 \times 1$
- $\mathbf{1}'\mathbf{X}$: 1×325 * $325 \times 10 \rightarrow 1 \times 10$
- $\mathbf{X}'\mathbf{1}$: 10×325 * $325 \times 1 \rightarrow 10 \times 1$
- $\mathbf{X}'\mathbf{X}$: 10×325 * $325 \times 10 \rightarrow 10 \times 10$
- $\mathbf{I} * \lambda$: 10×10
- $\mathbf{X}'\mathbf{X} + \mathbf{I} * \lambda$: 10×10

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Dimension der Koeffizientenmatrix?
- Dimensionen der einzelnen Submatrizen:
 - $\mathbf{1}'\mathbf{1}$: 1×325 * 325×1 → 1×1
 - $\mathbf{1}'\mathbf{X}$: 1×325 * 325×10 → 1×10
 - $\mathbf{X}'\mathbf{1}$: 10×325 * 325×1 → 10×1
 - $\mathbf{X}'\mathbf{X}$: 10×325 * 325×10 → 10×10
 - $\mathbf{I} * \lambda$: 10×10
 - $\mathbf{X}'\mathbf{X} + \mathbf{I} * \lambda$: 10×10
- Dimension der Koeffizientenmatrix:
 - 11×11

<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Dimension der Koeffizientenmatrix festlegen:
- `coeff <- array(0,c(nmarkers+1, nmarkers+1))`
- `> dim(coeff)`
- [1] 11 11
- `> head(coeff)`
- [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
- [1,] 0 0 0 0 0 0 0 0 0 0 0
- [2,] 0 0 0 0 0 0 0 0 0 0 0
- [3,] 0 0 0 0 0 0 0 0 0 0 0
- [4,] 0 0 0 0 0 0 0 0 0 0 0
- [5,] 0 0 0 0 0 0 0 0 0 0 0
- [6,] 0 0 0 0 0 0 0 0 0 0 0

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Koeffizientenmatrix befüllen:
 - `coeff[1:1,1:1] <- t(ones)%*%ones`
 - `coeff[1:1,2:(nmarkers+1)] <- t(ones)%*%x`
 - `coeff[2:(nmarkers+1),1:1] <- t(x)%*%ones`
 - `> head(coeff)`
 -
 -
 -
 -
 -
 -
 -
- | | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] | [,8] | [,9] | [,10] | [,11] |
|------|------|------|------|------|------|------|------|------|------|-------|-------|
| [1,] | 325 | 291 | 93 | 93 | 368 | 301 | 297 | 501 | 282 | 123 | 282 |
| [2,] | 291 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [3,] | 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [4,] | 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [5,] | 368 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [6,] | 301 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Koeffizientenmatrix befüllen:
- `xx <- t(x)%*%x`
- `i10 <- lamda*ident_matrix`
- `> head(i10)`
- [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
- [1,] 10 0 0 0 0 0 0 0 0 0 0
- [2,] 0 10 0 0 0 0 0 0 0 0 0
- [3,] 0 0 10 0 0 0 0 0 0 0 0
- [4,] 0 0 0 10 0 0 0 0 0 0 0
- [5,] 0 0 0 0 10 0 0 0 0 0 0
- [6,] 0 0 0 0 0 10 0 0 0 0 0
- `xx1 <- xx + i10`

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Koeffizientenmatrix befüllen:
- `coeff[2:(nmarkers+1),2:(nmarkers+1)] <- xx1`
- `> coeff`
- [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
- [1,] 325 291 93 93 368 301 297 501 282 123 282
- [2,] 291 453 98 98 456 366 374 461 126 154 126
- [3,] 93 98 103 93 145 112 122 156 41 74 41
- [4,] 93 98 93 103 145 112 122 156 41 74 41
- [5,] 368 456 145 145 592 470 465 591 154 205 154
- [6,] 301 366 112 112 470 441 412 476 132 124 132
- [7,] 297 374 122 122 465 412 433 478 129 120 129
- [8,] 501 461 156 156 591 476 478 863 411 197 411
- [9,] 282 126 41 41 154 132 129 411 420 41 410
- [10,] 123 154 74 74 205 124 120 197 41 169 41
- [11,] 282 126 41 41 154 132 129 411 410 41 420

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Right hand side (RHS)
- Dimension RHS festlegen
- `rhs <- array(0,(nmarkers+1))`
- `> dim(rhs)`
- `[1] 11`
- `> rhs`
- `[1] 0 0 0 0 0 0 0 0 0 0 0`
- `rhs[1] <- t(ones)%*%y`
- `> rhs`
- `[1] 144.254 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000`
- `rhs[2:(nmarkers+1)] <- t(x)%*%y`
- `> rhs`
- `[1] 144.254 198.967 71.389 71.389 245.536 201.584 206.512 240.299 42.972
94.266 42.972`

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Gleichungssystem lösen → Mittelwert und SNP-Effekte schätzen
- `solution_vec <- solve(coeff,rhs)`
- `> solution_vec`
- [1] -0.13327471 0.08134269 0.04162717 0.04162717 0.02160162 0.06642844
0.26182118 0.07407547 -0.02160162 0.20748910 -0.02160162

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Direkt genomischen Zuchtwert berechnen für die 31 Kälber
- `g_hat <- solution_vec[2:(nmarkers+1)]`
- `dgzw <- xkalb%*%g_hat`
- `> dim(dgzw)`
- `[1] 31 1`
- `> head(dgzw)`
- `[,1]`
- `[1,] 0.61939598`
- `[2,] 0.53614163`
- `[3,] 0.53614163`
- `[4,] 0.53614163`
- `[5,] -0.01233102`
- `[6,] 0.61939598`

$$DGZW = \sum_i^n X_i \hat{g}_i$$

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Direkt genomischen Zuchtwert berechnen für die 31 Kälber

- `g_hat <- solution_vec[2:(nmarkers+1)]`

- `dgzw <- xkalb%*%g_hat`

- `> dim(dgzw)`

- `[1] 31 1`

- `> head(dgzw)`

- `[,1]`

- `[1,] 0.61939598`

- `[2,] 0.53614163`

- `[3,] 0.53614163`

- `[4,] 0.53614163`

- `[5,] -0.01233102`

- `[6,] 0.61939598`

$$DGZW = \sum_i^n X_i \hat{g}_i$$

- `> summary(dgzw)`

V1

Min. :-0.01233

1st Qu. : 0.43872

Median : 0.53614

Mean : 0.56309

3rd Qu. : 0.72120

Max. : 1.16561

SNP-BLUP in R

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- Nach 4 Jahren stehen für die Kälber traditionelle Zuchtwerte aus der Nachkommenprüfung zur Verfügung.
- Wie gut ist die Übereinstimmung zwischen DGZW und dem traditionellen Zuchtwert?
- `> head(ykalb)`
 - [1,] 1
 - [1,] -0.555
 - [2,] -0.556
 - [3,] -0.556
 - [4,] -0.562
- `> cor(dgzw,ykalb)`
 - [1,] 1
 - [1,] 0.2601208



Genomische Selektion

Kursunterlagen Genomic Selection, Ben Hayes, 2015

Heutige Vorlesung

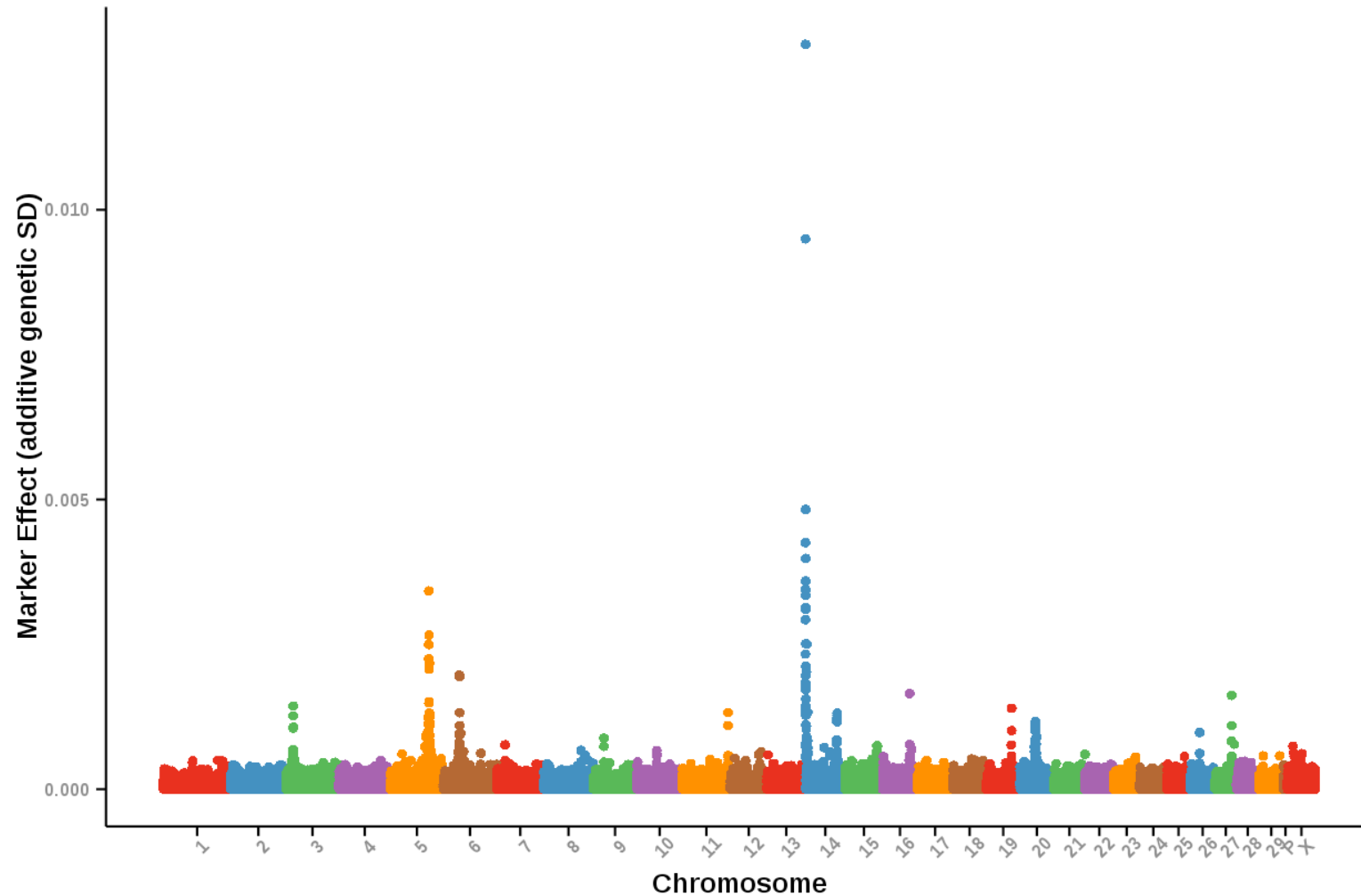
- Bayes'sche Methoden zur Schätzung genomischer Zuchtwerte
- Einflussfaktoren auf die Sicherheit der genomischen Zuchtwertschätzung
- Genomische Selektion im Zuchtprogramm

Bayes'sche Methoden zur Schätzung von SNP-Effekten

- Annahmen in SNP-BLUP oder G-BLUP:
 - Alle SNP-Effekte haben kleine Effekte (non-zero)
 - Folgen einer Normalverteilung
 - Erklären alle den gleichen Anteil an genetischer Varianz
- → Marker können keine moderaten/grossen Effekte aufweisen, aber auch nicht 0 sein (kein Effekt)
- Diese Annahmen aber nicht zutreffend für alle Merkmale
→ unterschiedliche genetische Architektur
- z.B. Marker in hohem LD mit QTL mit grossem Effekt auf ein Merkmal
- z.B. kein QTL in Region → SNP Effekte sollen 0 sein

Genetische Architektur von Merkmalen

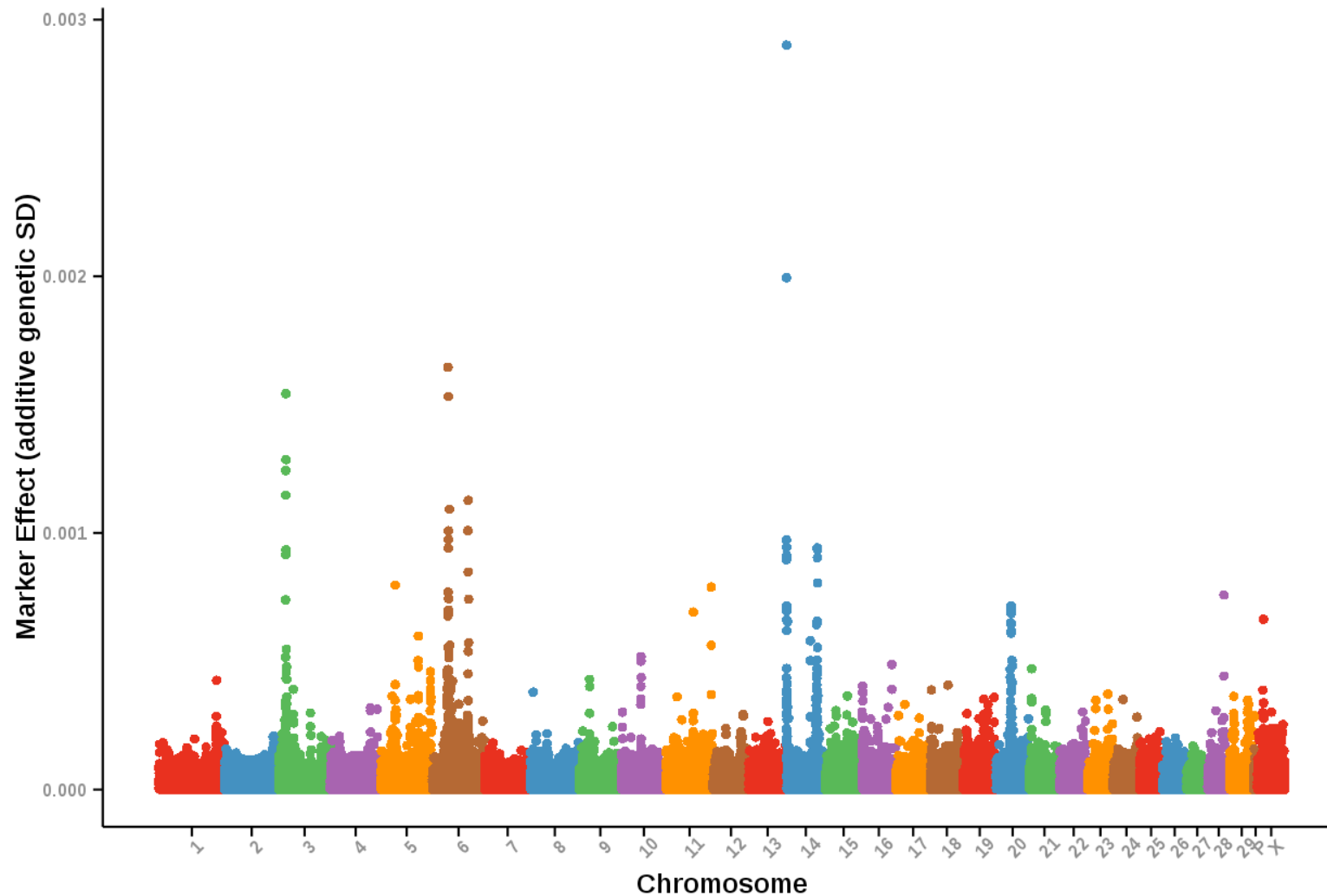
Distribution of marker effects for HO Fat_Percent (1512 run)



https://www.cdcb.us/Report_Data/Marker_Effects/marker_effects.cfm?reed=HO&Trait=Fat_Percent

Genetische Architektur von Merkmalen

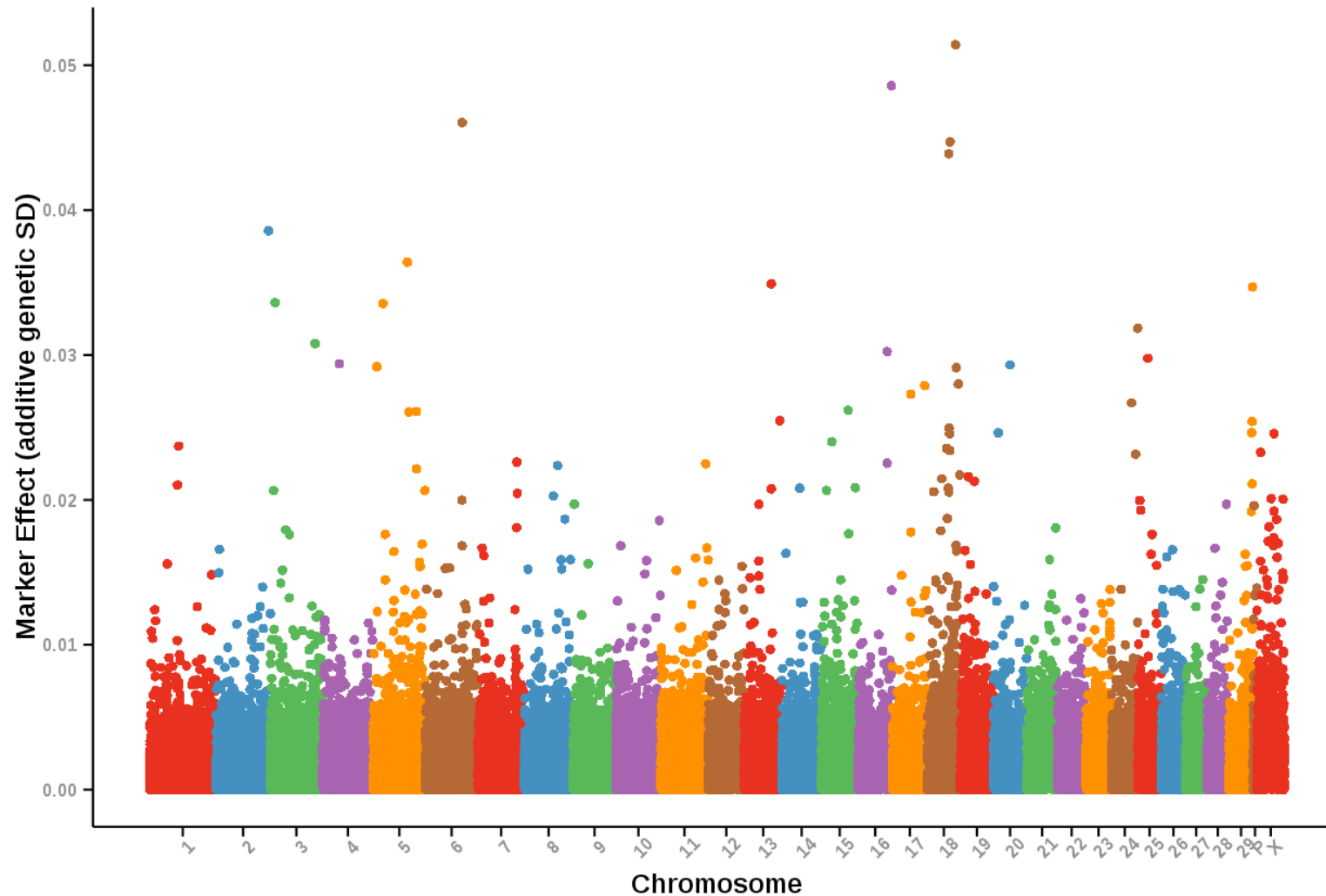
Distribution of marker effects for HO Pro_Percent (1512 run)



https://www.cdcb.us/Report_Data/Marker_Effects/marker_effects.cfm?reed=HO&Trait=Pro_Percent

Genetische Architektur von Merkmalen

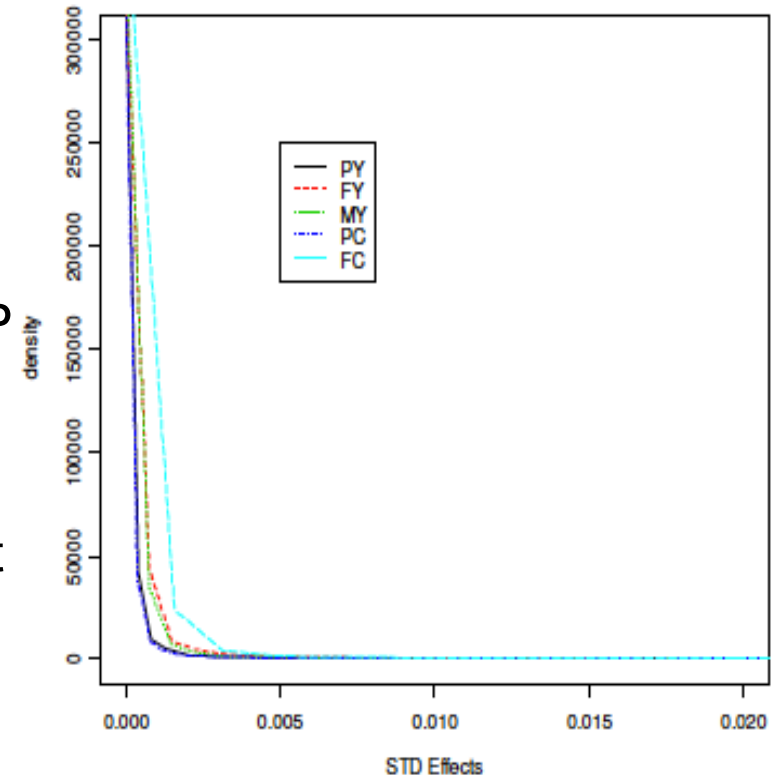
Distribution of marker effects for HO Cow_Conc_Rate (1512 run)



https://www.cdcb.us/Report_Data/Marker_Effects/marker_effects.cfm?Breed=HO&Trait=Cow_Conc_Rate

Bayes'sche Methoden zur Schätzung von SNP-Effekten

- Bayes'sche Methoden erlauben a priori Wissen über SNP-Effekte miteinzubeziehen
- BayesA:
 - t-Verteilung, grössere Chance dass SNP moderate Effekte aufweisen
- BayesB:
 - Viele SNP mit keinem Effekt, wenige mit grossen Effekten
- ... Viele mehr ... Unter starker Entwicklung



→ VO Angewandte Statistische Methoden in den Nutztierwissenschaften

Methodenvergleich an echten Daten (Hayes, 2009)

- 1500 Holstein Stiere (Australien)
- Gentoypisiert für 56000 SNP
- Aufteilung der Daten in:
 - Referenz: Stiere geboren <2003
 - Validierung: Stiere geboren \geq 2003
- Sicherheit genomische Zuchtwertschätzung
 - Korrelation zwischen DGZW und traditionellem Zuchtwert basierend auf Töchterleistungen für die Validierungstiere



Methodenvergleich an echten Daten (Hayes, 2009)

Table 3 MEBV- Correlation between predicted MEBV and ABV in the validation data set (Bulls proven in years 2005, 2006, 2007)

<i>Method</i>	<i>Protein kg</i>	<i>Fat kg</i>	<i>Protein %</i>	<i>Fat %</i>
Bayes B	0.55	0.51	0.68	0.73
Bayes A	0.53	0.48	0.66	0.70
BLUP	0.60	0.48	0.66	0.64

Bayes'sche Methoden zur Schätzung von SNP-Effekten

- Bayes'sche Methoden können von Vorteil sein wenn:
 - QTL mit moderaten/grossen Effekten auf das Merkmal
 - Sehr grossen Anzahl SNP (z.B. High Density Chip), → einige SNP Effekte werden auf 0 gesetzt

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- Simulationsstudien haben sehr hohe Sicherheit der genomischen Zuchtwertschätzung gezeigt (>0.8)
- Ergebnisse mit Praxisdaten liegen je nach Merkmal oft deutlich darunter
- Was sind die Einflussfaktoren auf die Sicherheit der genomischen Zuchtwertschätzung?

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- Einflussfaktoren
 - Linkage Disequilibrium zwischen Marker und QTL
 - Grösse der Referenzpopulation
 - Heritabilität des Merkmals
 - Anzahl an QTL mit Einfluss auf das Merkmal
 - Verwandtschaft zwischen Referenzpopulation und Selektionskandidaten
 - ...

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- **Linkage Disequilibrium (LD)** zwischen Marker und QTL
 - Je niedriger das LD, desto mehr SNPs sind notwendig damit zumindest 1 SNP in LD mit jedem QTL ist
 - Meuwissen (2009, GSE, 41:35) zeigte in Simulationen, dass die Anzahl folgendermassen bestimmt werden kann:
 - $10 * N_e * L$
 - N_e ... Effektive Populationsgrösse
 - L ... Länge des Genoms in Morgan
 - Holstein: $N_e \sim 100$, $L=30$ Morgan
 - $10 * 100 * 30 = 30000$ Marker
 - \rightarrow 50K SNP Chip

N_e = Anzahl von Individuen in einer Idealpopulation, die der Anzahl Individuen in einer realen Population mit unterschiedlichem Geschlechterverhältnis entspricht.

$$N_e = 4 \frac{N_m \times N_w}{N_m + N_w}$$

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- Höhere SNP-Dichte – z.B. Vergleich 50k – HD
- Ertl et al. (2014, JDS 97:487-496)
- Fleckvieh in Bayern/Österreich

Trait	50K	HD
Milk yield	0.398	0.414
Fat yield	0.419	0.427
Protein yield	0.378	0.392
SCS	0.548	0.559
Milkability	0.458	0.481
Muscling	0.480	0.501
Udder	0.297	0.313
Feet and legs	0.315	0.330
Stature	0.366	0.379
Average	0.407	0.422

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- Höhere SNP-Dichte – z.B. Vergleich 50k – HD
- Erbe et al. (2012, JDS 95:4114-4129)
- Holstein/Jersey Australien

Reference	Validation	Milk yield			Fat yield			Protein yield		
		50K	800K	TRANS	50K	800K	TRANS	50K	800K	TRANS
Holstein	Holstein	0.62	0.63	0.63	0.64	0.65	0.63	0.55	0.57	0.56
	Jersey	0.27	0.24	0.40	0.12	0.21	0.12	-0.05	0.05	0.21
Jersey	Holstein	0.19	0.03	0.15	0.29	0.29	0.18	0.13	0.10	0.12
	Jersey	0.49	0.48	0.53	0.48	0.46	0.47	0.42	0.41	0.43
Combined	Holstein	0.61	0.62	0.62	0.65	0.66	0.64	0.56	0.57	0.57
	Jersey	0.45	0.51	0.57	0.50	0.49	0.45	0.43	0.46	0.53

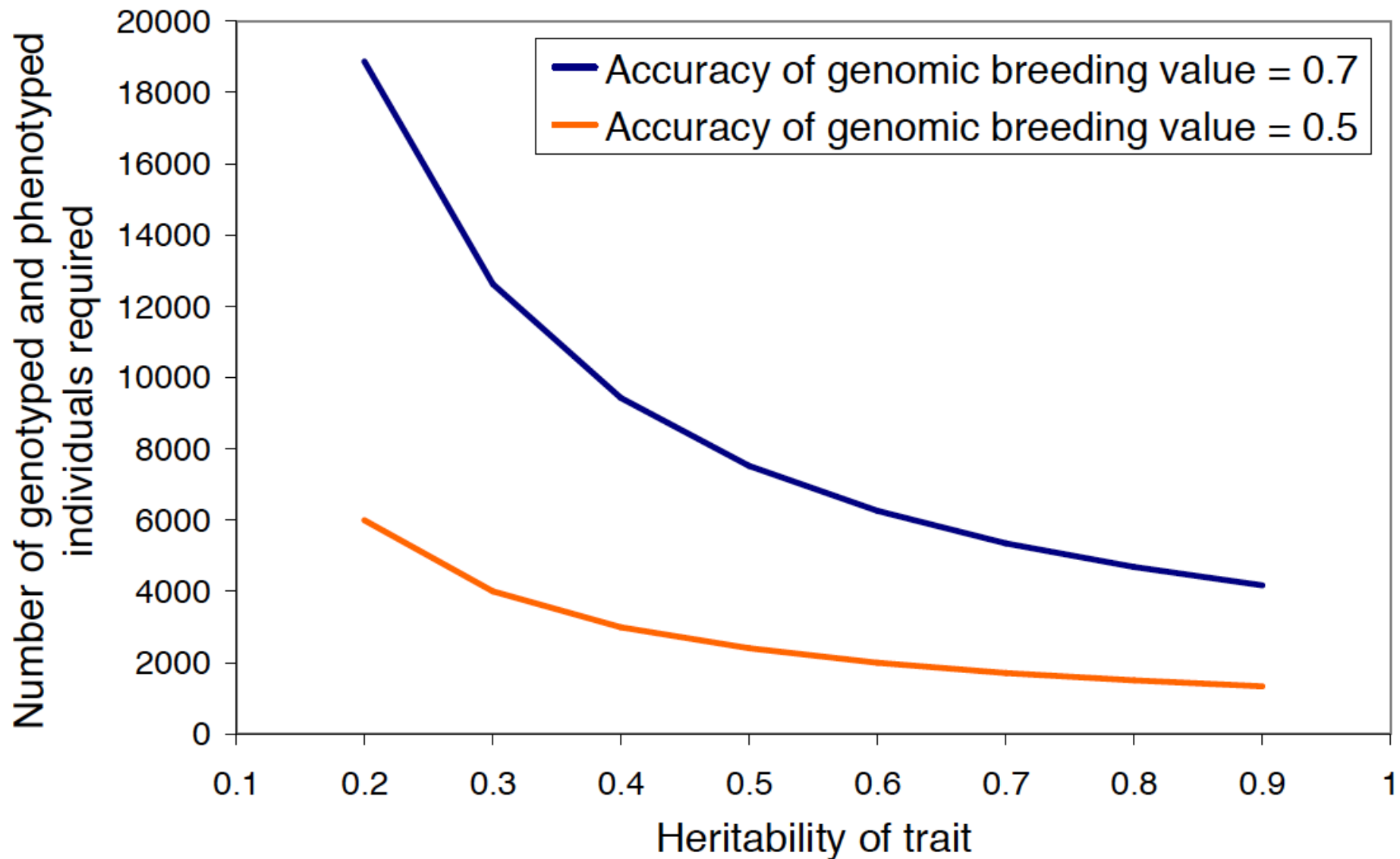
Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

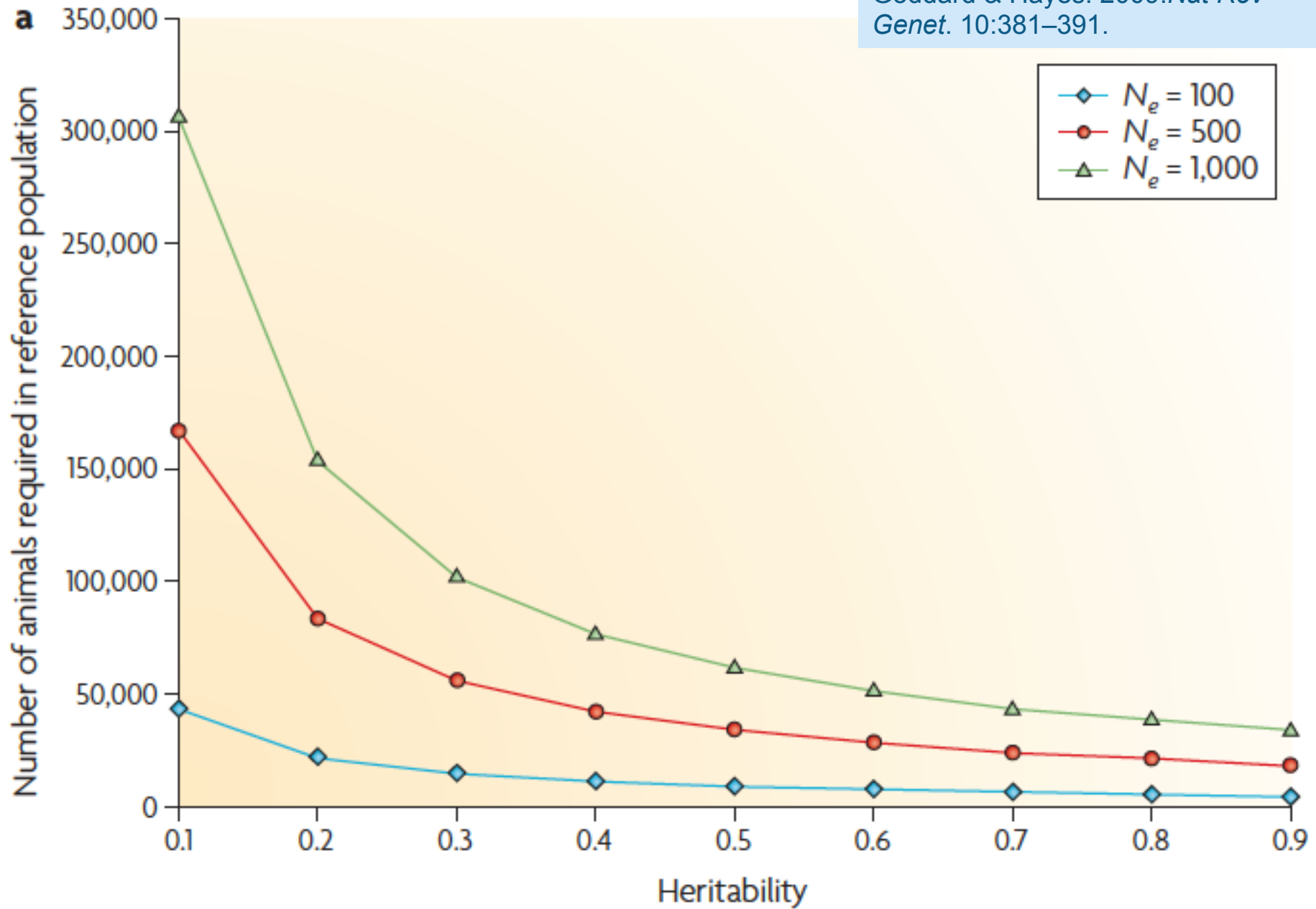
- Grösse **Referenzpopulation**, **Heritabilität** und **Anzahl QTL** mit Einfluss auf das Merkmal
- Daetwyler et al, (2008) und Goddard (2008)

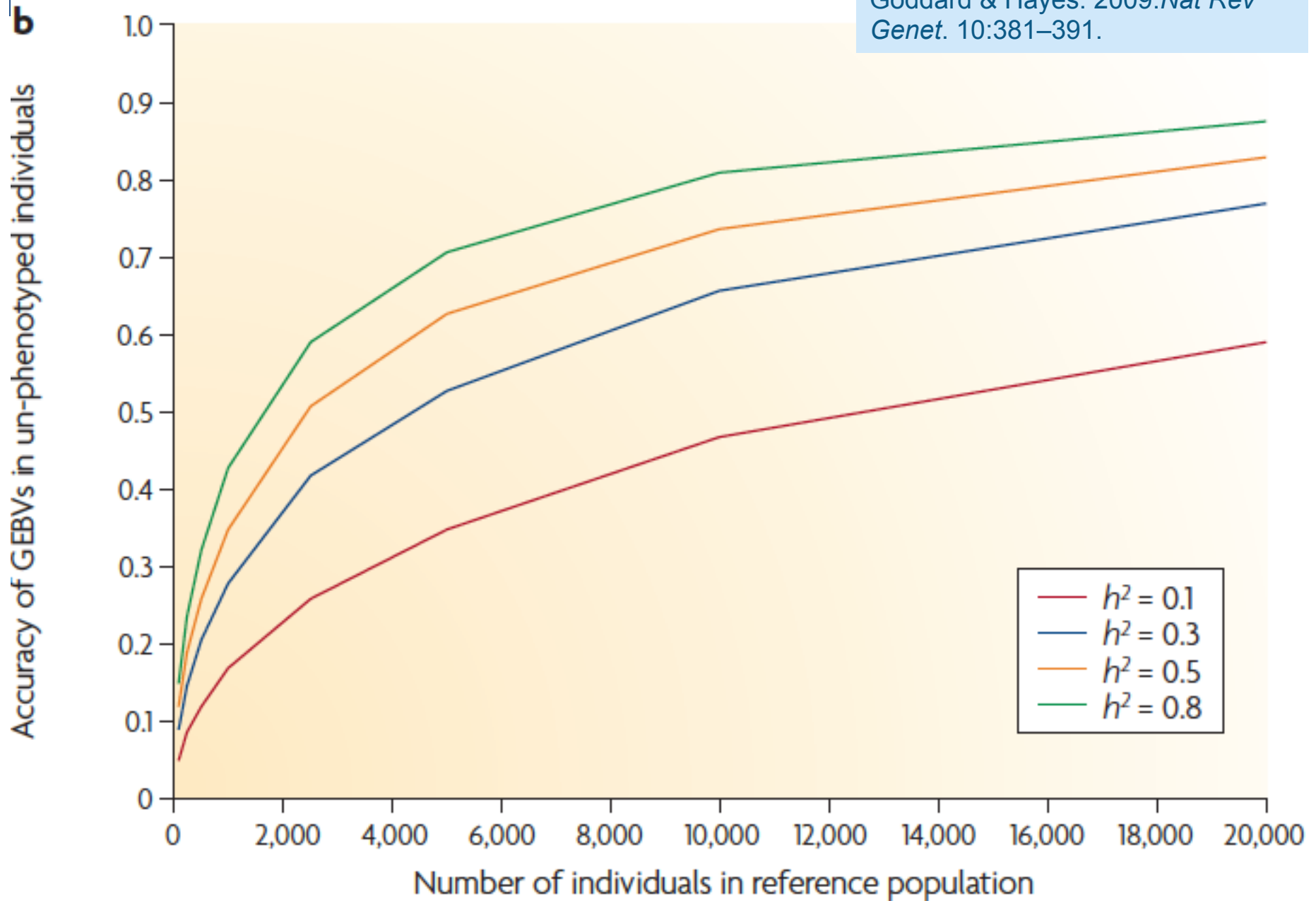
$$r = \sqrt{Nh^2 / (Nh^2 + q)}$$

- N = Anzahl Tiere in Referenzpopulation
- h^2 = Heritabilität des Merkmals
- $q = 2N_eL$ (Anzahl QTLs mit Effekt auf ein Merkmal)

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung







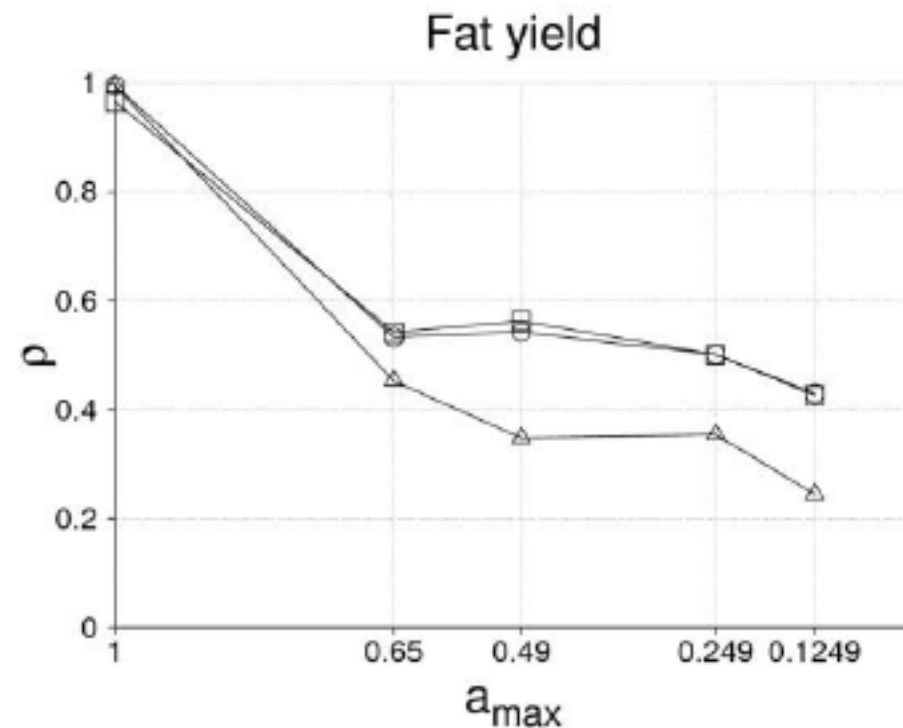
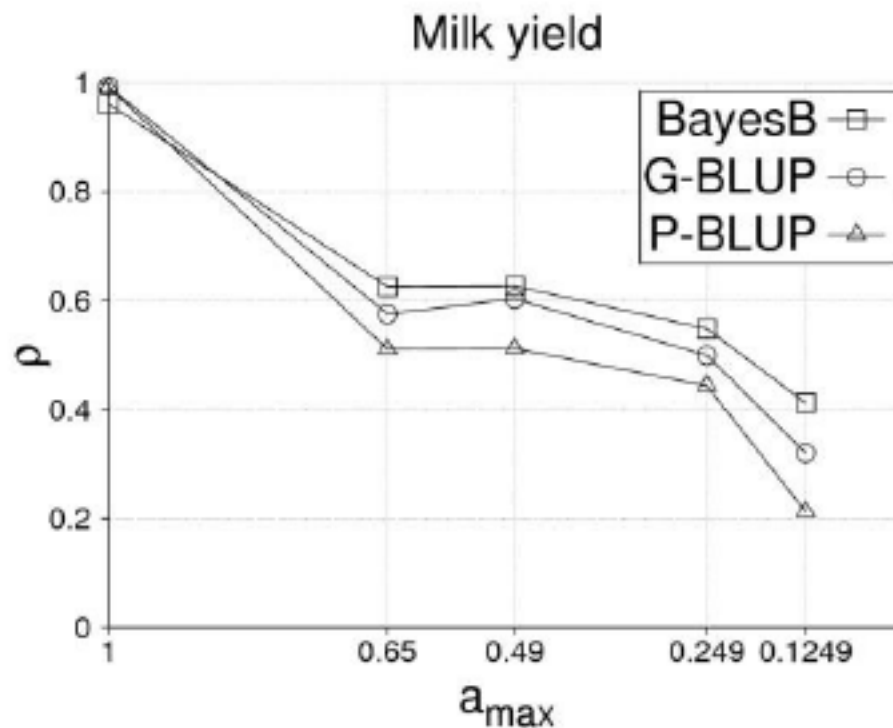
Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- **Verwandtschaft** zwischen Referenzpopulation und Selektionskandidaten beeinflusst Sicherheit
- Hohe Sicherheit, wenn Tiere in der Referenzpopulation hoch verwandt sind mit Selektionskandidaten
- Gleichzeitig soll die Referenzpopulation aber so verschieden wie möglich sein!!

Einfluss auf die Sicherheit der genomischen Zuchtwertschätzung

- **Verwandtschaft** zwischen Referenzpopulation und Selektionskandidaten beeinflusst Sicherheit

Habier et al. 2010 Gen. Sel. Evol. 42:5



Genomische Selektion im Zuchtprogramm

ORIGINAL ARTICLE

Strategy for applying genome-wide selection in dairy cattle

L.R. Schaeffer

Summary

Department of Anima

Canada

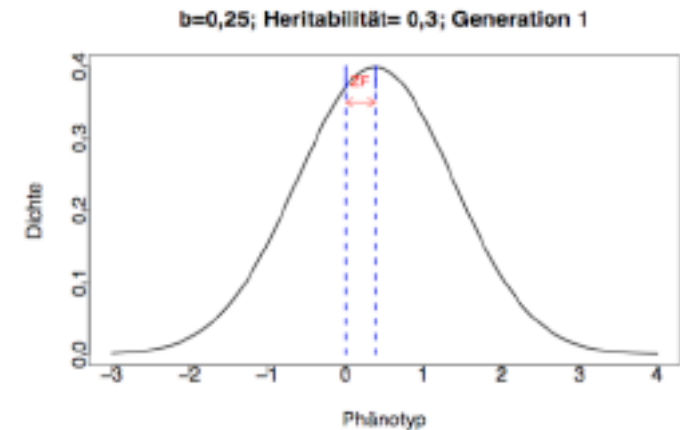
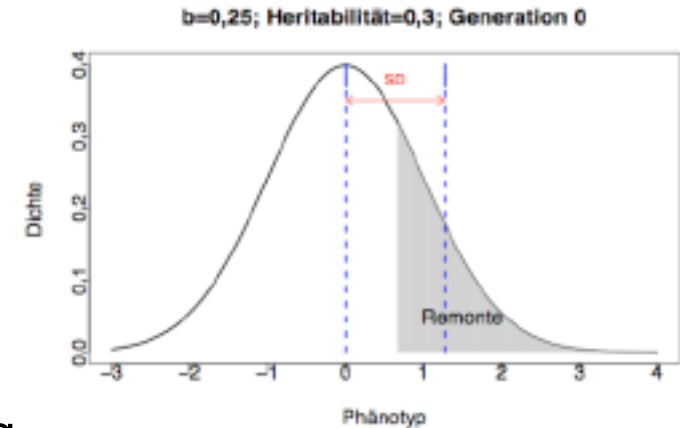
Animals can be genotyped for thousands of single nucleotide polymorphisms (SNPs) at one time, where the SNPs are located at roughly 1-cM intervals throughout the genome. For each contiguous pair of SNPs there are four possible haplotypes that could be inherited from the sire. The effects of each interval on a trait can be estimated for all intervals simultaneously in a model where interval effects are random factors. Given the estimated effects of each haplotype for every interval in the genome, and given an animal's genotype, a 'genomic' estimated breeding value is obtained by summing the estimated effects for that genotype. The accuracy of that estimator of breeding values is around 80%. Because the genomic estimated breeding values can be calculated at birth, and because it has a high accuracy, a strategy that utilizes these advantages was compared with a traditional progeny testing strategy under a typical Canadian-like dairy cattle situation. Costs of proving bulls were reduced by 92% and genetic change was increased by a factor of 2. Genome-wide selection may become a popular tool for genetic improvement in livestock.

Genomische Selektion im Zuchtprogramm

- **Zuchtfortschritt (ZF):**

$$ZF / \text{Jahr} = \frac{\sigma_a * i * r_{\hat{A}A}}{G}$$

- σ_a = additive genetische Standardabweichung
- i = Selektionsintensität
- $r_{\hat{A}A}$ = Sicherheit Zuchtwert (Korrelation zw. geschätztem und wahren Zuchtwert)
- G = Generationsintervall



Schwarzenbacher, 2013

Konventionelles Zuchtprogramm

Zeit (Monate)	Massnahme
0	Stiermutter wird besamt
9	Stierkalb geboren
27	Prüfbesamungen der jungen Stiere werden durchgeführt
36	Töchter der Prüfstiere werden geboren
54	Töchter der Prüfstiere werden besamt
63	Töchter kalben ab und beginnen 1. Laktation
68	Erste Milchzuchtwerte aus Testtagsmodell vorhanden
73	Töchter schliessen 1. Laktation ab Stierenklassierung

Konventionelles Zuchtprogramm

↓
6 Jahre

Zeit (Monate)	Massnahme
0	Stiermutter wird besamt
9	Stierkalb geboren
27	Prüfbesamungen der jungen Stiere werden durchgeführt
36	Töchter der Prüfstiere werden geboren
54	Töchter der Prüfstiere werden besamt
63	Töchter kalben ab und beginnen 1. Laktation
68	Erste Milchzuchtwerte aus Testtagsmodell vorhanden
73	Töchter schliessen 1. Laktation ab Stierenklassierung


Leben



Tod



Konventionelles Zuchtprogramm

	Zeit (Monate)	Massnahme
Stier R.B. Peter JANITO *TM * 19.11. 2008	0	Stiermutter wird besamt
	9	Stierkalb geboren
Teststier 16.06.2010 – 15.06.2011	27	Prüfbesamungen der jungen Stiere werden durchgeführt
Gislers Janito ALEGRA * 30.05.2011	36	Töchter der Prüfstiere werden geboren
	54	Töchter der Prüfstiere werden besamt
	63	Töchter kalben ab und beginnen 1. Laktation
	68	Erste Milchzuchtwerte aus Testtagsmodell vorhanden
Stierenklassierung 03.12.2014 JANITO ist 6 Jahre alt	73	Töchter schliessen 1. Laktation ab Stierenklassierung

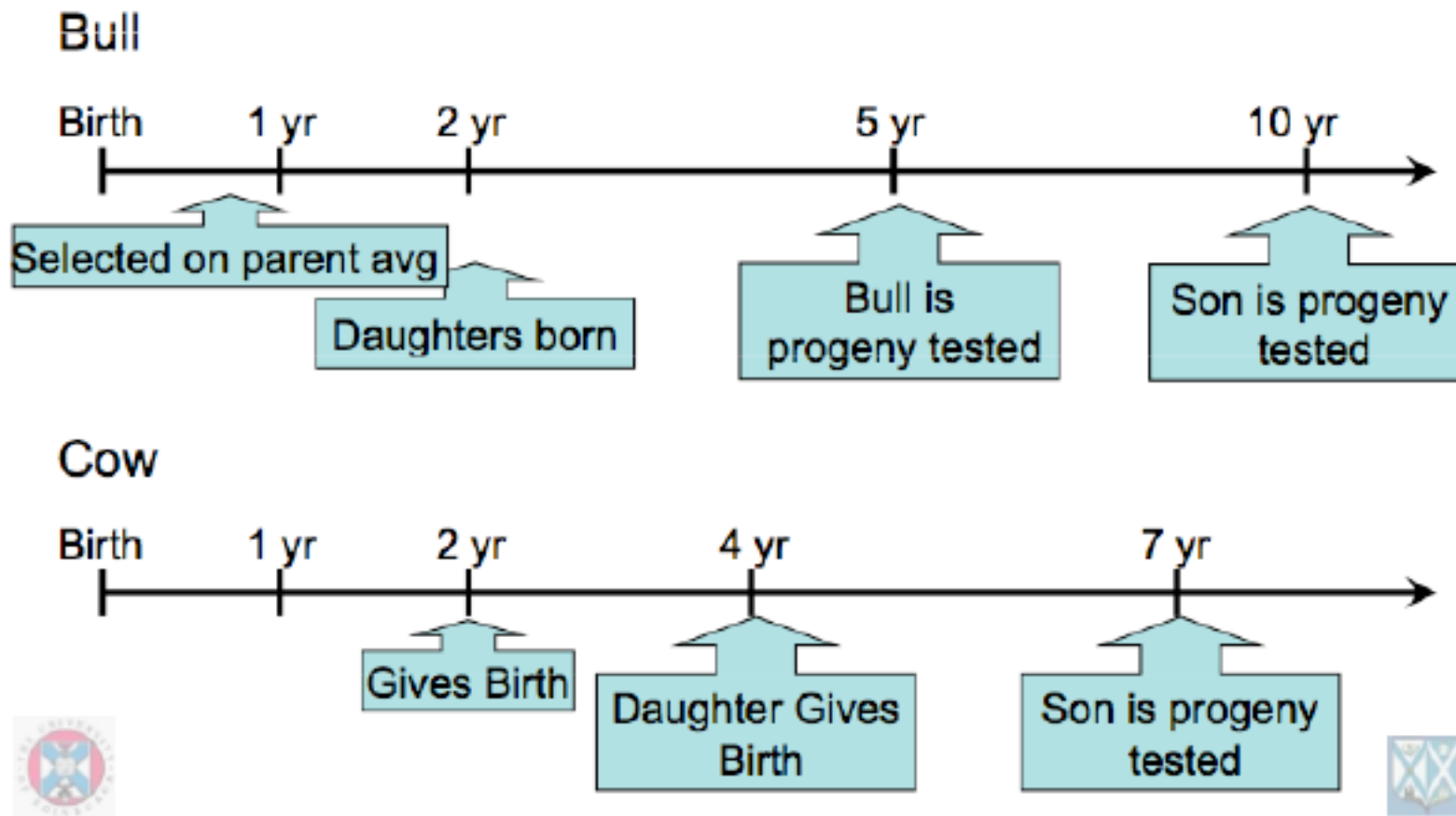
Leben



Tod



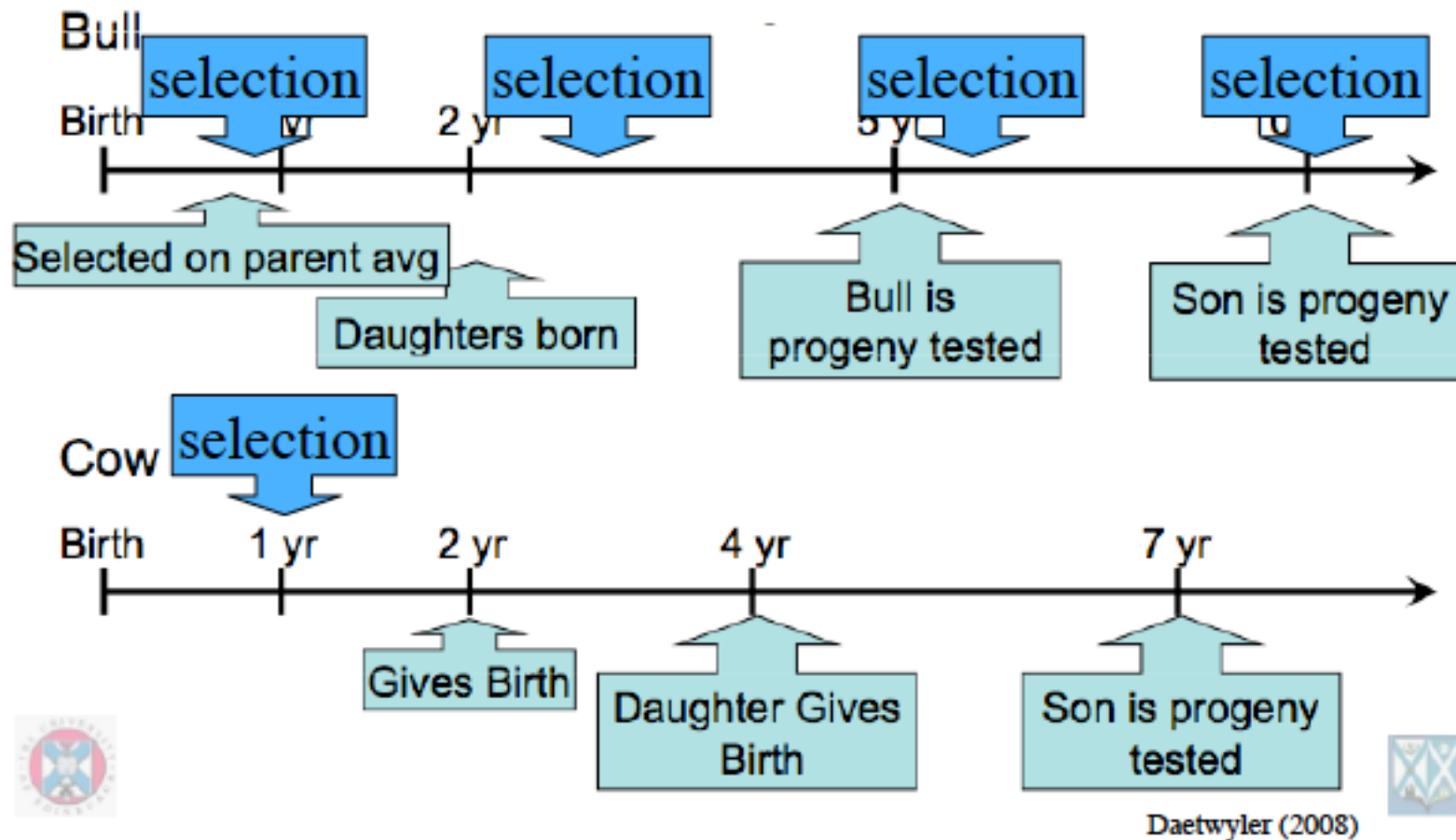
Konventionelles Zuchtprogramm



Daetwyler (2008)

Genomisches Zuchtprogramm

$$ZF / \text{Jahr} = \frac{\sigma_a * i * r_{AA}}{G}$$

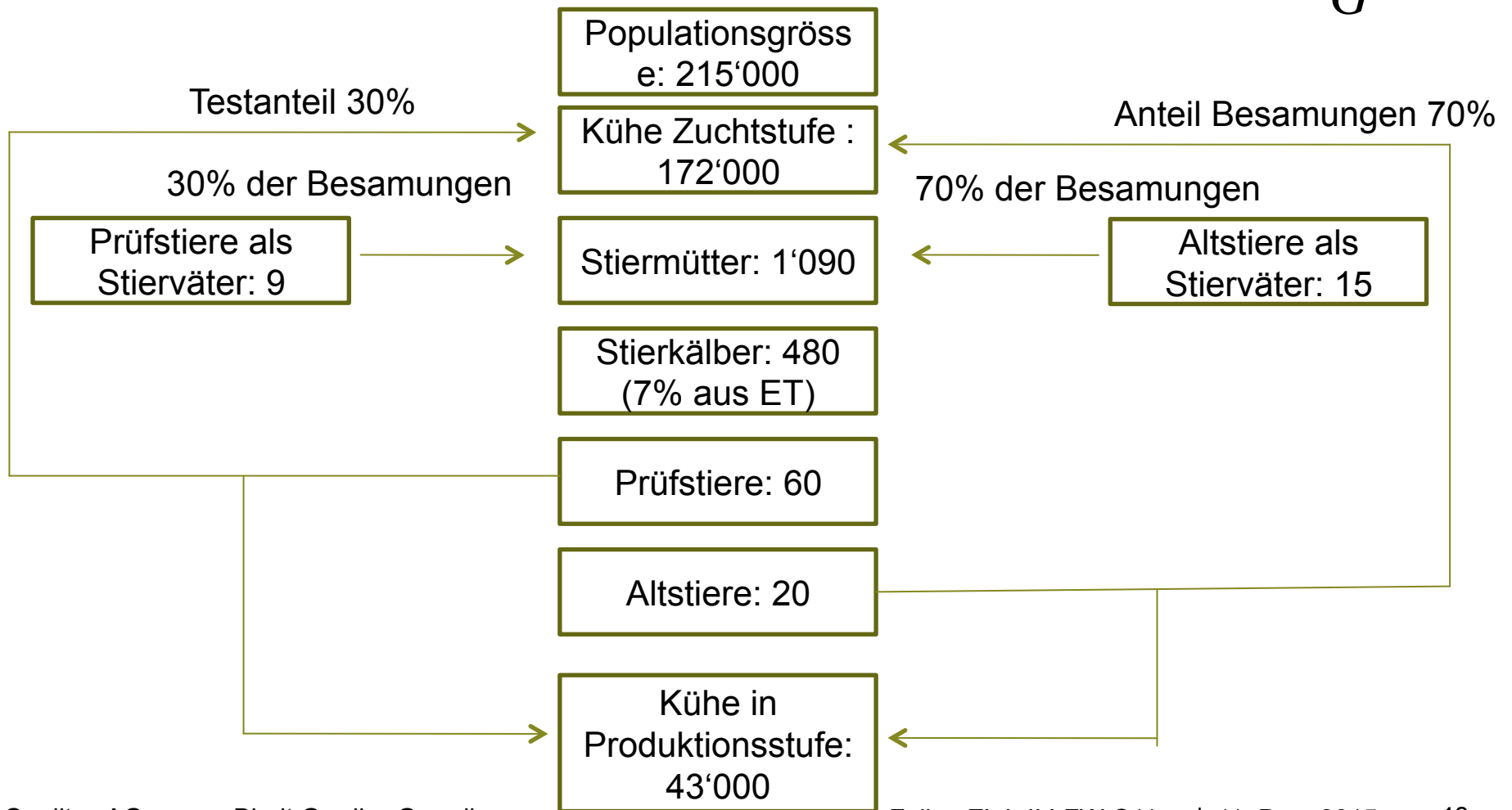


Daetwyler (2008)

Zuchtprogramm Braunvieh Schweiz

Masterarbeit Sabrina Bütler, 2014

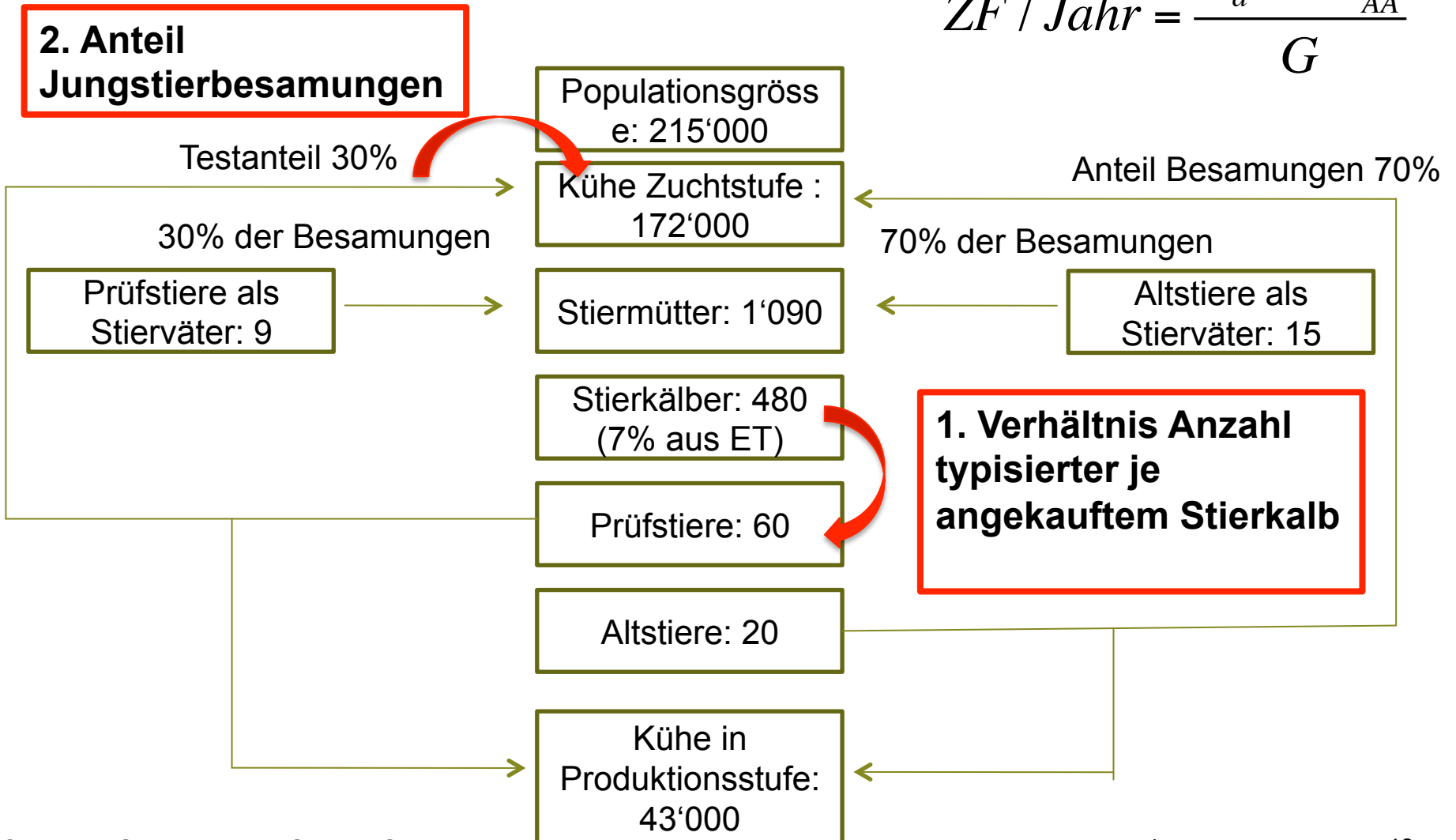
$$ZF / Jahr = \frac{\sigma_a * i * r_{AA}}{G}$$



Zuchtprogramm Braunvieh Schweiz

Masterarbeit Sabrina Bütler, 2014

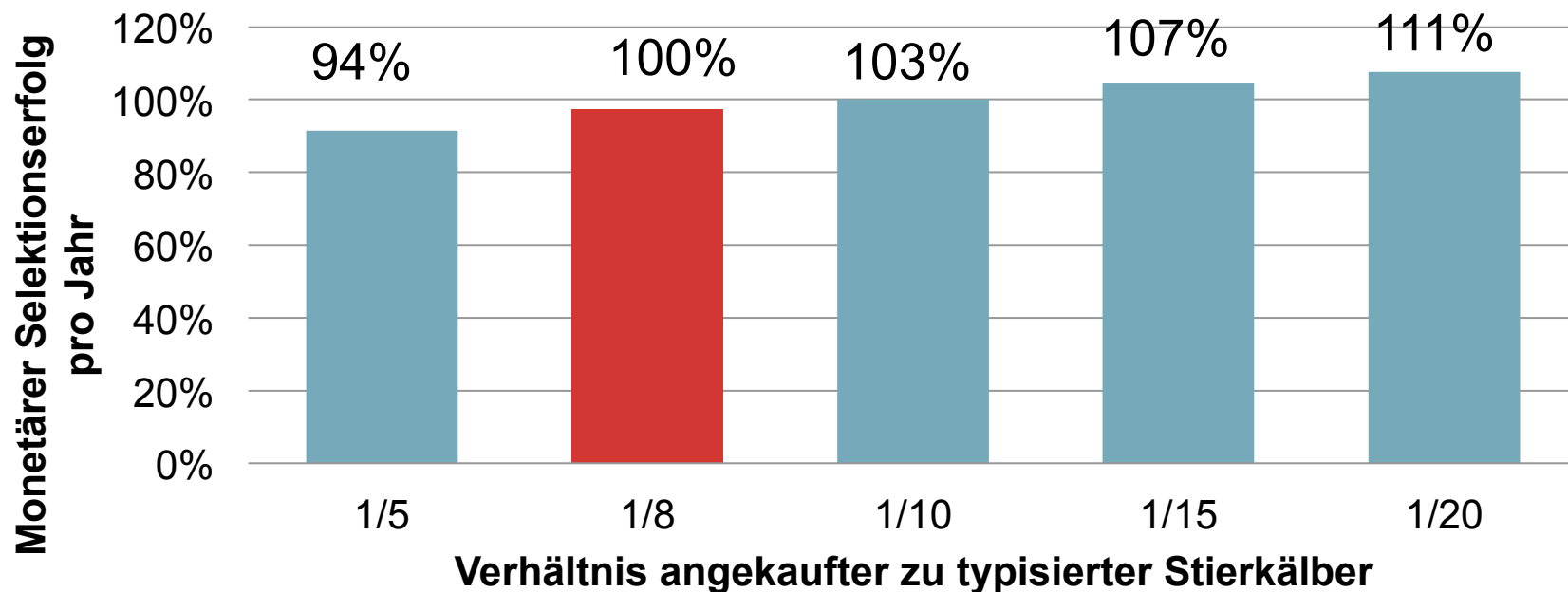
$$ZF / Jahr = \frac{\sigma_a * i * r_{AA}}{G}$$



Zuchtprogramm Braunvieh Schweiz

Masterarbeit Sabrina Bütler, 2014

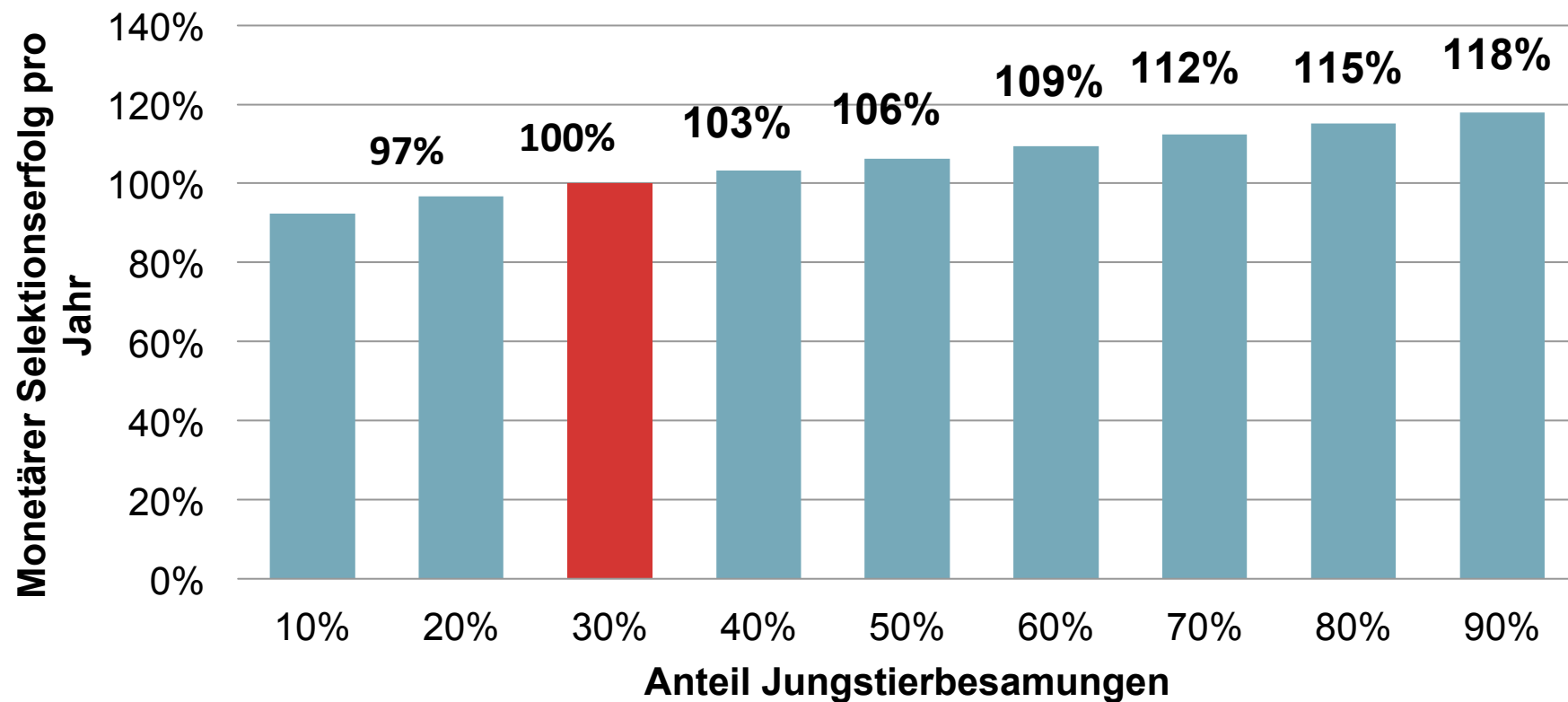
$$ZF / Jahr = \frac{\sigma_a * i * r_{AA}}{G}$$



Zuchtprogramm Braunvieh Schweiz

Masterarbeit Sabrina Bütler, 2014

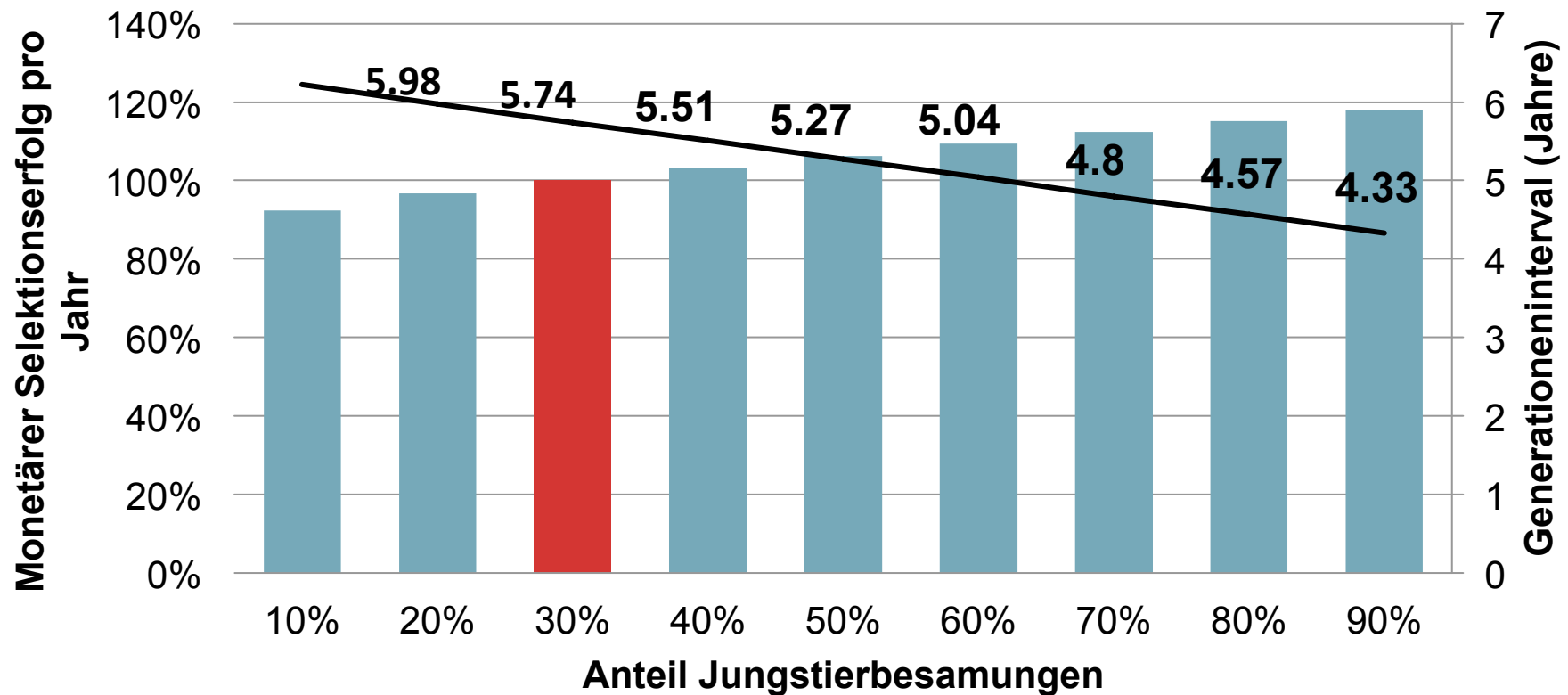
$$ZF / \text{Jahr} = \frac{\sigma_a * i * r_{AA}}{G}$$



Zuchtprogramm Braunvieh Schweiz

Masterarbeit Sabrina Bütler, 2014

$$ZF / \text{Jahr} = \frac{\sigma_a * i * r_{AA}}{G}$$



Schlussfolgerungen

Masterarbeit Sabrina Bütler, 2014

$$ZF / \text{Jahr} = \frac{\sigma_a * i * r_{AA}}{G}$$

- Verhältnis angekaufter zu typisierten Stierkälber:
 - Zuchtfortschritt steigt (schärfer selektieren)
- Anteil Jungstierbesamungen:
 - Zuchtfortschritt steigt mit höherem Anteil
 - (Generationsintervall sinkt)
- Erstkalbealter Stiermütter:
 - Zuchtfortschritt steigt mit jüngeren Stiermüttern (Generationsintervall sinkt)
- Typisierung Stiermütter:
 - Zuchtfortschritt steigt leicht an

Chancen und Risiken

- Neue Technologie – sehr junge Wissenschaft!!
 - Wurde sehr schnell in die Praxis eingeführt
- Grösse der Referenzpopulation entscheidend!
 - Kleine Populationen?
 - Multi-Breed Referenzpopulation
- Verwandtschaft entscheidend!
 - Wenn Selektionskandidaten wenig verwandt zur Referenzpopulation → Sicherheit ↓
- Inzuchtkontrolle!
 - Schnellerer Zuchtfortschritt

Chancen und Risiken

- Phänotypen werden immer wichtiger
 - Traditionelle Zuchtwertschätzung immer noch nötig → Schätzung von SNP-Effekten
- Vorteil für funktionale Merkmale und schwierig zu erfassende Merkmale
 - Fruchtbarkeit, Krankheiten, ...

**Genomische Selektion
revolutioniert die Tierzucht**

Prüfung

- Freitag, 18. Dezember 2015
- 10:00 bis 11:30
- Keine Hilfsmittel
- Nicht programmierbarer Taschenrechner
- Matrixrechenregeln – Addition, Multiplikation, Inversion einfacher Matrizen!!

Ausblick Frühjahrssemester 2016

- 751-6212-00L Angewandte Zuchtwertschätzung für Nutztiere (Birgit Gredler-Grandl)
- Angewandte Zuchtwertschätzung bei Rind (Merkmale, Interpretation traditioneller und genomischer Zuchtwerte)
- Angewandte Zuchtwertschätzung bei Schwein, Schaf, Ziege (Merkmale und Interpretation)
- Übungen
- Exkursion/Besuch bei Qualitas und Braunvieh Schweiz in Zug

QUALITAS⁺



Ausblick Frühjahrssemester 2016

- **751-7602-00L Angewandte statistische Methoden in den Nutztierwissenschaften (Peter von Rohr)**
- Einführung in multiple lineare Regression
- Problematik $n \ll p$ von Least Squares in der genomischen Selektion
- BLUP basierte Lösungsansätze
- LASSO (Least Absolute Shrinkage and Selection Operator) als Alternative zu den in der Tierzucht verwendeten Methoden
- Einführung in Bayes'sche Statistik und Parameterschätzung
- Anwendung der Bayes'schen Verfahren in der genomischen Selektion (BayesA, BayesB, BayesC, BayesN)
- Übungen



SCHWEIZERISCHE VEREINIGUNG FÜR TIERPRODUKTION
Association Suisse pour la Production Animale
Swiss Association for Animal Production

- Verein mit Sitz in Zug
- Förderung Tierwissenschaften – Züchtung, Haltung, Erhalt der genetischen Vielfalt landwirt. Nutztiere
- Zusammenarbeit/Austausch versch. Organisationen in der Tierproduktion
- Verbindung zwischen Forschung und Praxis
- Förderung wissenschaftlicher Nachwuchs
- Mitglied werden: siehe Flyer

(<http://www.svt-asp.ch/PDF/2014/Flyer-SVT-Kluge-Koepfe-sind-gesucht.pdf>)