

Güte und Genauigkeit eines reduzierten zufälligen Modells zur multiplen QTL-Kartierung

M. Mayer und D. Zimmer
FB Genetik und Biometrie
Leibniz-Institut für Nutztierbiologie (FBN)



LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

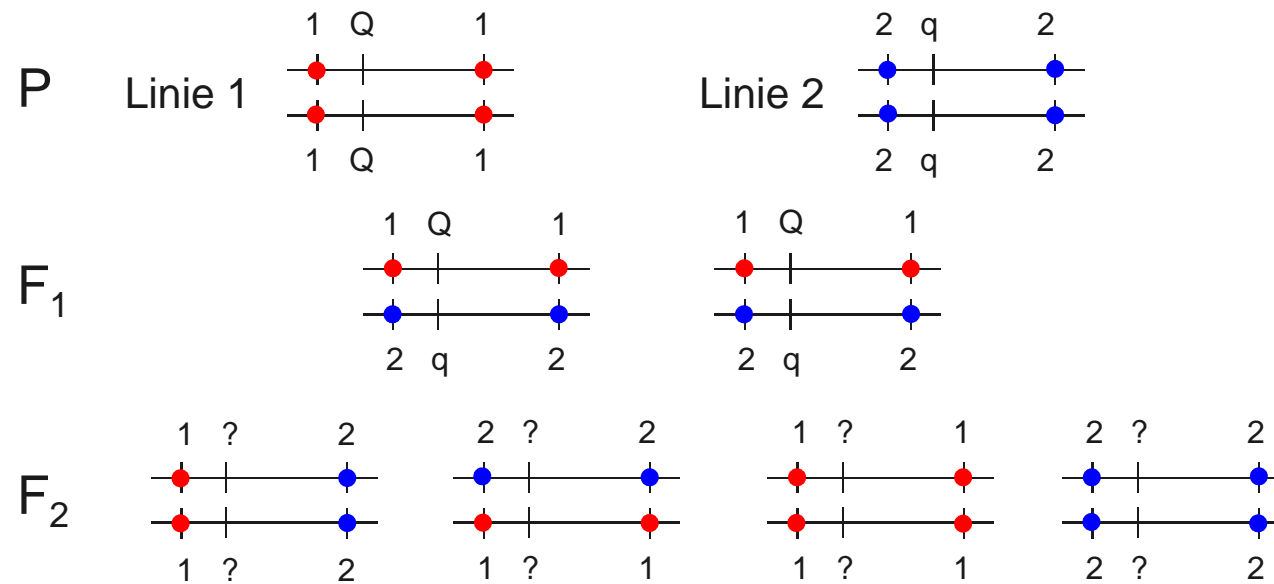
Gliederung

- Methodik – Ableitung des reduzierten Modells
- Methodenvergleich mittels Simulationsstudien
- Methodenvergleich für multiple Familien



Versuchsdesign

F₂-Populationen abgeleitet aus Inzuchtlinien



Individuelles Modell

Lineares gemischtes Modell mit v QTL

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{\ell=1}^v \mathbf{z}_{\ell} (\mathbf{u}_{a_{\ell}} + \mathbf{u}_{d_{\ell}}) + \sum_{\ell=1}^{v-1} \sum_{k=\ell+1}^v \mathbf{z}_{\ell k} (\mathbf{u}_{aa_{\ell k}} + \mathbf{u}_{ad_{\ell k}} + \mathbf{u}_{da_{\ell k}} + \mathbf{u}_{dd_{\ell k}}) + \mathbf{e}$$

\mathbf{y} ... Vektor der Phänotypen (Länge n)

$\boldsymbol{\beta}$... Vektor der fixen Effekte mit Designmatrix \mathbf{X}

\mathbf{u}_{τ} ... Vektoren der zufälligen **individuellen** genetischen Effekte (Länge n), $E(\mathbf{u}_{\tau}) = \mathbf{0}$, $Var(\mathbf{u}_{\tau}) = \mathbf{V}_{\tau} \sigma_{\tau}^2$

$$\tau \in \{a_{\ell}, d_{\ell}, aa_{\ell k}, ad_{\ell k}, da_{\ell k}, dd_{\ell k}\}$$

$\mathbf{z}_{\ell}, \mathbf{z}_{\ell k}$... zugehörige Designmatrizen

\mathbf{e} ... Restfehler mit $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I} \sigma_e^2)$



Verwandtschaftsmatrizen Individuelles Modell

Aufstellen der Verwandtschaftsmatrizen mit

1. **Bedingten QTL-Genotypwahrscheinlichkeiten**
→ Nutzung von flankierenden Markergenotypen
2. **Elementare Kovarianzmatrizen** → Kovarianzen von Effekten für bekannte QTL-Genotypen

$$\mathbf{A} = \begin{matrix} & G_{QQ} & & \\ & \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} & & \\ & & G_{Qq} & \\ & & & G_{qq} \end{matrix}$$

$$\mathbf{D} = \begin{matrix} & G_{QQ} & & \\ & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & & \\ & & G_{Qq} & \\ & & & G_{qq} \end{matrix}$$

Elementare paarweise epistatische Kovarianzmatrizen nach Zimmer *et al.* (2011) $\mathbf{A} \otimes \mathbf{A}$ $\mathbf{A} \otimes \mathbf{D}$ $\mathbf{D} \otimes \mathbf{A}$ $\mathbf{D} \otimes \mathbf{D}$

Verwandtschaftsmatrizen Individuelles Modell

Jedes F_2 -Individuum erhält einen genetischen Effekt!

➤ Dimension der Verwandtschaftsmatrizen $n \times n$

Xie *et al.* (1998): $\mathbf{V}_{a_\ell} = \{a_{st}^{(\ell)}\}_{s,t=1}^n$ am ℓ -ten QTL ($i, j \in \{1, \dots, 9\}$)

$$a_{st}^{(\ell)} = \mathbf{I}_i^\ell \mathbf{A} (\mathbf{I}_j^\ell)' = (2p_i^{QQ} + p_i^{Qq})p_j^{QQ} + p_j^{Qq} + (p_i^{Qq} + 2p_i^{qq})p_j^{qq}$$

$$a_{ss}^{(\ell)} = \mathbf{I}_i^\ell \text{diag}(\mathbf{A}) = 2p_i^{QQ} + p_i^{Qq} + 2p_i^{qq}$$

➤ Annahme: QTL und Marker nicht identisch, höchstens ein QTL innerhalb des Markerintervalls

Verwandtschaftsmatrizen Individuelles Modell

Beispiel:

$$a_{12}^{\ell} = \mathbf{I}_{22,22}^{\ell} \mathbf{A} \left(\mathbf{I}_{22,12}^{\ell} \right)'$$

Individuum	Marker-genotyp
1	22,22
2	22,12
3	22,22
4	22,22
5	22,12

$$= (0,000 \quad 0,042 \quad 0,958) \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0,009 \\ 0,437 \\ 0,554 \end{pmatrix}$$

$$= 2 \cdot 0,000 \cdot 0,009 + 0,000 \cdot 0,437 + \dots + 2 \cdot 0,958 \cdot 0,554$$

$$= 1,52$$

$$a_{11}^{\ell} = \mathbf{I}_{22,22}^{\ell} \mathit{diag}(\mathbf{A})$$

$$= 2 \cdot 0,000 + 0,042 + 2 \cdot 0,958$$

$$= 1,96$$

Verwandtschaftsmatrizen Individuelles Modell

Beispiel:

$$a_{12}^{\ell} = \mathbf{I}_{22,22}^{\ell} \mathbf{A} \left(\mathbf{I}_{22,12}^{\ell} \right)' = 1,52$$

$$a_{11}^{\ell} = \mathbf{I}_{22,22}^{\ell} \text{diag}(\mathbf{A}) = 1,96$$

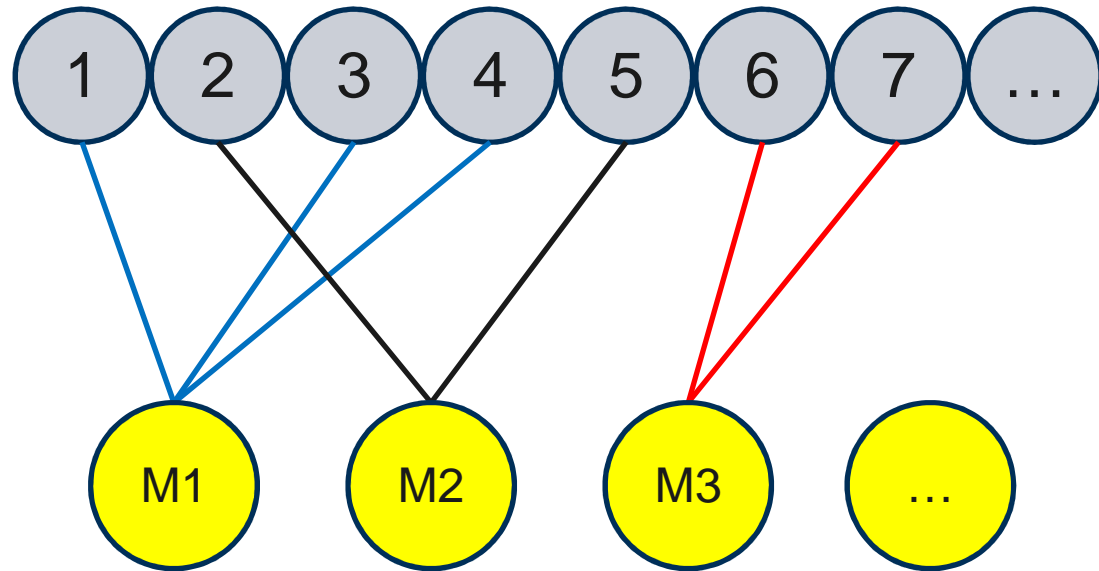
Individuum	Marker-genotyp
1	22,22
2	22,12
3	22,22
4	22,22
5	22,12

$$\mathbf{V}_{a_{\ell}} = \begin{pmatrix} 1,96 & 1,52 & 1,92 & 1,92 & 1,52 \\ 1,52 & 1,56 & 1,52 & 1,52 & 1,30 \\ 1,92 & 1,52 & 1,96 & 1,92 & 1,52 \\ 1,92 & 1,52 & 1,92 & 1,96 & 1,52 \\ 1,52 & 1,30 & 1,52 & 1,52 & 1,56 \end{pmatrix}$$

Berechnung von $\mathbf{V}_{d_{\ell}}$, $\mathbf{V}_{aa_{\ell k}}$, $\mathbf{V}_{ad_{\ell k}}$, $\mathbf{V}_{da_{\ell k}}$, $\mathbf{V}_{dd_{\ell k}}$ analog, aber \mathbf{D} bzw. Kronecker-Produkt $\left(\mathbf{I}_{ij}^{\ell k} \right)$ nutzen

Gruppierung des F₂-Individuen

Individuum	Marker-genotyp
1	12,11
2	22,22
3	12,11
4	12,11
5	22,22
6	11,11
7	11,11
...	



Gruppierung der F₂-Individuen nach dem Markergenotyp

Äquivalentes Modell mit \emptyset -Effekten

Äquivalente Schreibweise

$$\mathbf{z}_\tau \mathbf{u}_\tau = \tilde{\mathbf{z}}_\tau \tilde{\mathbf{u}}_\tau + \mathbf{m}_\tau \quad \text{mit} \quad \tau \in \{a_\ell, d_\ell, aa_{\ell k}, ad_{\ell k}, da_{\ell k}, dd_{\ell k}\}$$

$\tilde{\mathbf{z}}_\tau$... zugehörige Designmatrizen

$\tilde{\mathbf{u}}_{a_\ell}, \tilde{\mathbf{u}}_{d_\ell}$... Vektor der additiven bzw. dominanten \emptyset -Effekte (Längen n_ℓ)

$\tilde{\mathbf{u}}_s$... Vektor der epistatischen \emptyset -Effekte (Länge $n_{\ell k}$)

$$E(\tilde{\mathbf{u}}_\tau) = \mathbf{0}, \quad \text{Var}(\tilde{\mathbf{u}}_\tau) = \tilde{\mathbf{V}}_\tau \sigma_\tau^2 \quad s \in \{aa_{\ell k}, ad_{\ell k}, da_{\ell k}, dd_{\ell k}\}$$

\mathbf{m}_τ ... Vektor der genetischen „sampling“ Effekte (Länge n)

$$E(\mathbf{m}_\tau) = \mathbf{0}, \quad \text{Var}(\mathbf{m}_\tau) = \mathbf{V}_{m_\tau} \sigma_\tau^2$$

➤ $n_\ell, n_{\ell k}$ Anzahl Markerklassen

➤ $n_\ell = 9, n_{\ell k} \in \{27, 81\}$ voll informative Marker



Äquivalentes Modell mit \emptyset -Effekten

Reduzierte additive Verwandtschaftsmatrix $\tilde{\mathbf{V}}_{a_\ell} = \{\tilde{a}_{ij}^\ell\}_{i,j=1}^{n_\ell}$

➤ Diagonalelement

$$\tilde{a}_{ii}^\ell = \begin{cases} \mathbf{I}_i^\ell \text{diag}(\mathbf{A}) & \text{wenn } n_i^\ell = 0 \\ \frac{1}{n_i^\ell} \mathbf{I}_i^\ell \text{diag}(\mathbf{A}) + \left(1 - \frac{1}{n_i^\ell}\right) \mathbf{I}_i^\ell \mathbf{A} (\mathbf{I}_i^\ell)' & \text{sonst} \end{cases}$$

➤ Nebendiagonalelement

$$\tilde{a}_{ij}^\ell = \mathbf{I}_i^\ell \mathbf{A} (\mathbf{I}_j^\ell)' \quad (i \neq j)$$

➤ Berechnung von $\tilde{\mathbf{V}}_{d_\ell}, \tilde{\mathbf{V}}_{aa_{\ell k}}, \tilde{\mathbf{V}}_{ad_{\ell k}}, \tilde{\mathbf{V}}_{da_{\ell k}}, \tilde{\mathbf{V}}_{dd_{\ell k}}$ analog

Äquivalentes Modell mit \emptyset -Effekten

$$\mathbf{V}_{a_\ell} = \begin{pmatrix} 1,96 & 1,52 & 1,92 & 1,92 & 1,52 \\ 1,52 & 1,56 & 1,52 & 1,52 & 1,30 \\ 1,92 & 1,52 & 1,96 & 1,92 & 1,52 \\ 1,92 & 1,52 & 1,92 & 1,96 & 1,52 \\ 1,52 & 1,30 & 1,52 & 1,52 & 1,56 \end{pmatrix}$$

Beispiel:

Individuum	Marker-genotyp
1	22,22
2	22,12
3	22,22
4	22,22
5	22,12

$$\tilde{a}_{12}^l = \mathbf{I}_{22,22}^l \mathbf{A} \left(\mathbf{I}_{22,12}^l \right)' = 1,52$$

$$\begin{aligned} \tilde{a}_{11}^l &= \frac{1}{n_{22,22}^l} \mathbf{I}_{22,22}^l \text{diag}(\mathbf{A}) + \left(1 - \frac{1}{n_{22,22}^l} \right) \mathbf{I}_{22,22}^l \mathbf{A} \left(\mathbf{I}_{22,22}^l \right)' \\ &= \frac{1}{3} \cdot 1,96 + \left(1 - \frac{1}{3} \right) \cdot 1,92 = 1,93 \end{aligned}$$

$$\tilde{\mathbf{V}}_{a_\ell} = \begin{pmatrix} 1,93 & 1,52 \\ 1,52 & 1,43 \end{pmatrix}$$

➤ Differenz besteht in Matrix $\mathbf{V}_{m_{a_\ell}}$



Äquivalentes Modell mit \emptyset -Effekten

Kovarianz des äquivalenten Ausdrucks

$$(\tilde{\mathbf{Z}}_{\tau} \tilde{\mathbf{V}}_{\tau} \tilde{\mathbf{Z}}_{\tau}' + \mathbf{V}_{m_{\tau}}) \sigma_{\tau}^2$$

➤ Nebendiagonale in $\mathbf{V}_{m_{\tau}}$ abhängig vom Stichprobenumfang:

$$\lim_{n \rightarrow \infty} \mathbf{V}_{m_{\tau}} = \Delta_{\tau} \quad \text{mit Diagonalmatrix } \Delta_{\tau} \text{ (Ordnung } n)$$

Reduziertes Modell

Annahme: Asymptotische Varianzen der genetischen „sampling“ Effekte **gleich** für alle Markerklassen

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{\ell=1}^v \tilde{\mathbf{z}}_{\ell} (\tilde{\mathbf{u}}_{a_{\ell}} + \tilde{\mathbf{u}}_{d_{\ell}}) + \sum_{\ell=1}^{v-1} \sum_{k=\ell+1}^v \tilde{\mathbf{z}}_{\ell k} (\tilde{\mathbf{u}}_{aa_{\ell k}} + \tilde{\mathbf{u}}_{ad_{\ell k}} + \tilde{\mathbf{u}}_{da_{\ell k}} + \tilde{\mathbf{u}}_{dd_{\ell k}}) + \boldsymbol{\varepsilon}$$

Annahme:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_{\varepsilon}^2) \quad \text{mit} \quad \mathbf{I}\sigma_{\varepsilon}^2 \approx \sum_{\tau} \mathbf{V}_{m_{\tau}} \sigma_{\tau}^2 + \mathbf{I}\sigma_e^2$$

!!! Approximation der Kovarianzmatrix der Phänotypen des individuellen Modells durch das reduzierte Modell

Gewichtetes reduziertes Modell

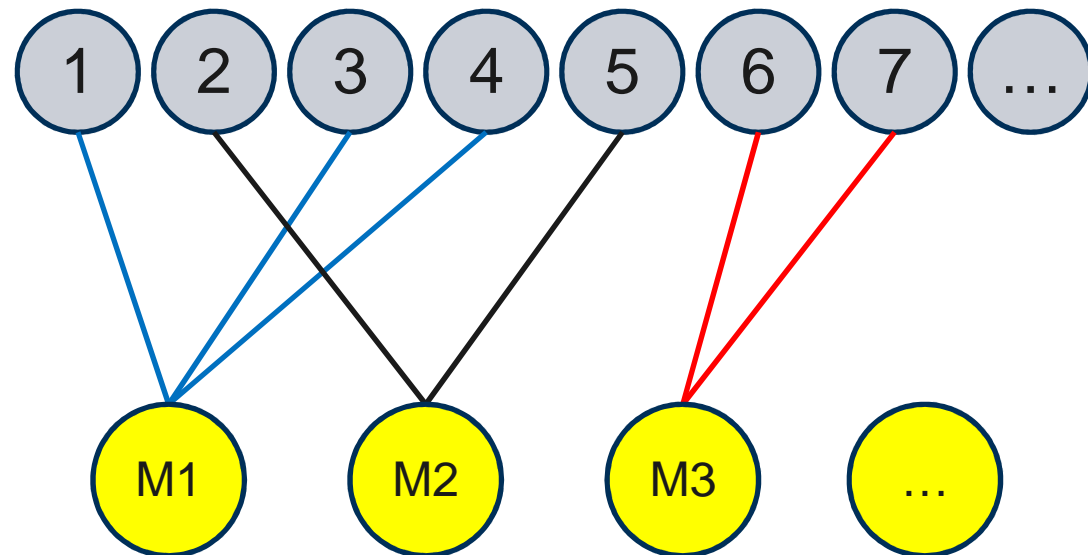
- Verwerfung der Annahme von gleichen sampling Varianzen für alle Markerklassen

Hierauf soll aber an dieser Stelle nicht weiter eingegangen werden

Individuelles vs. reduziertes Modell

- Individuelles Modell: Jedes F_2 -Individuum erhält einen genetischen Effekt

Individuum	Marker-genotyp
1	12,11
2	22,22
3	12,11
4	12,11
5	22,22
6	11,11
7	11,11



- Reduziertes Modell: Jede Markerklasse erhält einen Durchschnittseffekt

Dimension der Verwandtschaftsmatrizen \mathbf{V}_τ

Individuelles Modell	Reduziertes Modell
$\dim(\mathbf{V}_{a_\ell}) = n \times n$	$\dim(\mathbf{V}_{a_\ell}) = n_\ell \times n_\ell$
$\dim(\mathbf{V}_{d_\ell}) = n \times n$	$\dim(\mathbf{V}_{d_\ell}) = n_\ell \times n_\ell$
$\dim(\mathbf{V}_s) = n \times n$	$\dim(\mathbf{V}_s) = n_{\ell k} \times n_{\ell k}$

$$s \in \{aa_{\ell k}, ad_{\ell k}, da_{\ell k}, dd_{\ell k}\}$$

$$n_\ell = 9, \quad n_{\ell k} \in \{27, 81\}$$

- Beispiel: $n = 500 \Rightarrow \dim(\mathbf{V}_{a_\ell}) = 500 \times 500$ vs. $\dim(\tilde{\mathbf{V}}_{a_\ell}) = 9 \times 9$
- Dimension von $\tilde{\mathbf{V}}_\tau \downarrow$ mit \uparrow # Ind., # Marker, \uparrow # Varianzkomponenten.

Multiple Familien

- Entstehenden Kovarianzmatrizen für multiple Familien sind blockdiagonal
- Dimension von $\tilde{\mathbf{V}}_{\tau}$ ↓ mit ↑ Familiengröße, ↓ # Familien
- # Familien ↑↑ → Rechenzeit ↑↑↑

Simulationen

- Evaluierung des reduzierten Modells durch umfangreiche Simulationen
- Vergleich mit individuellem Modell, Regressionsintervallkartierung, multiple Intervallkartierung
- Güte der Methoden
- Genauigkeit geschätzter QTL-Positionen und QTL-Effekte (Mittelwert, mittlerer quadrierter Fehler)
- Ein- und Zwei-QTL-Modelle mit additiven und additiv-mal-additiven Effekte, # Individuen, relative QTL-Varianz, multiple Familien usw.



Ergebnisse

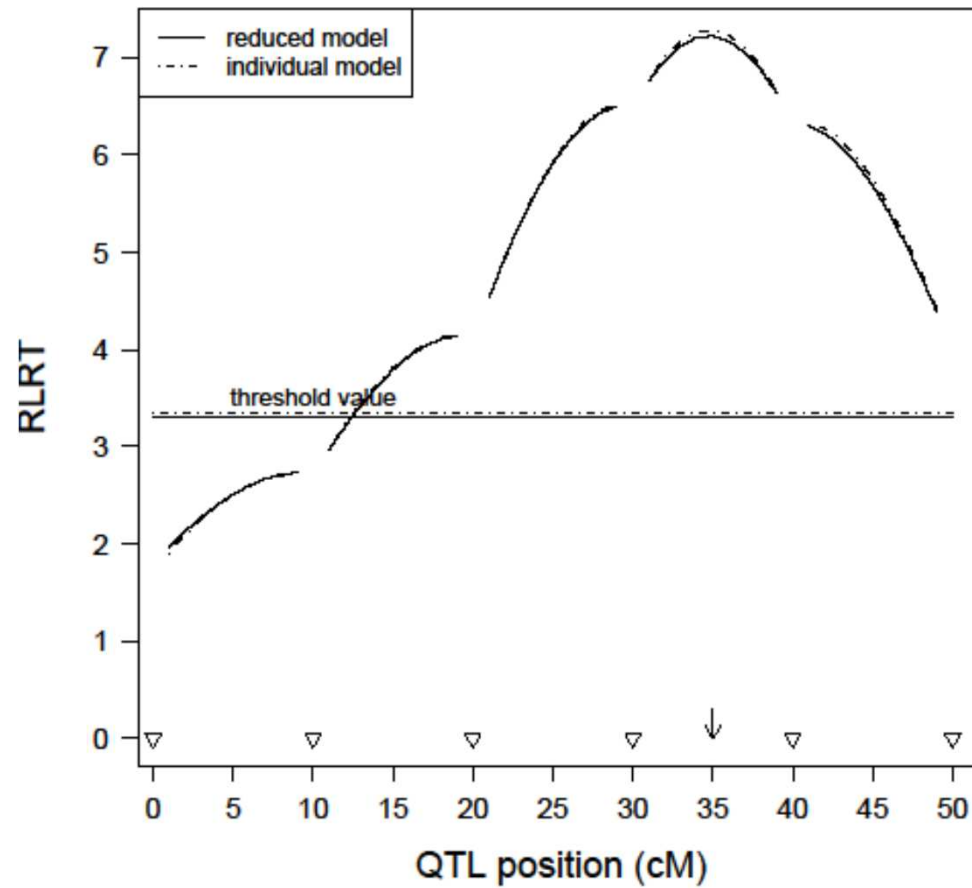
- Kaum Methodenunterschiede
 - Güte ähnlich
 - Genauigkeit der geschätzten QTL-Positionen
 - **Reduziertes und individuelles Modell sehr ähnliche Ergebnisse**

Simulationen

Ein QTL-Szenario

- 1000 Wiederholungen
- 200 Individuen
- Marker bei 0,10,20,30,40,50 cM
- QTL bei 35 cM
- Nur additiv-genetischer Effekt
- QTL-Varianz 5%





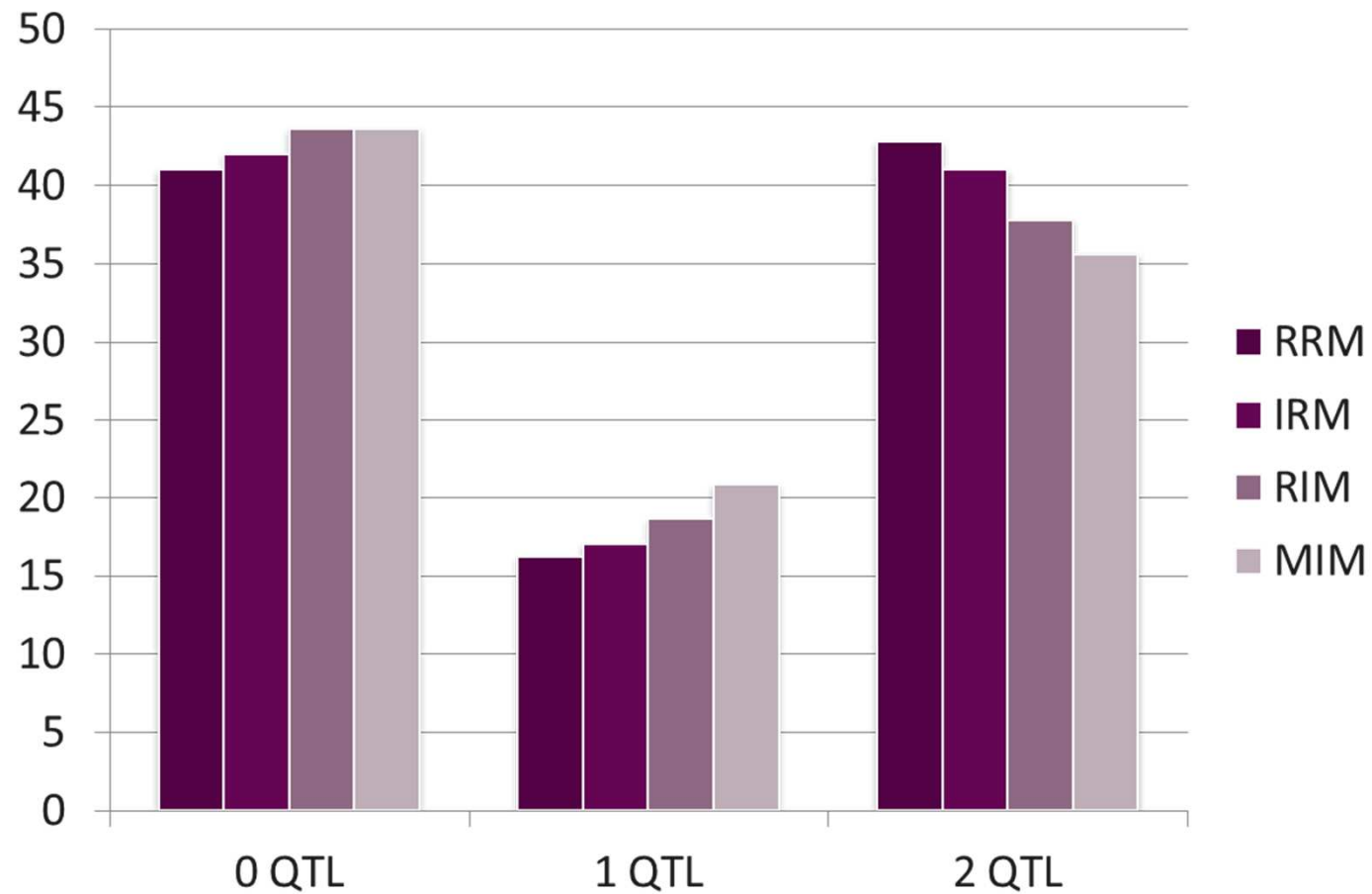
Simulationen

Zwei QTL-Szenarium

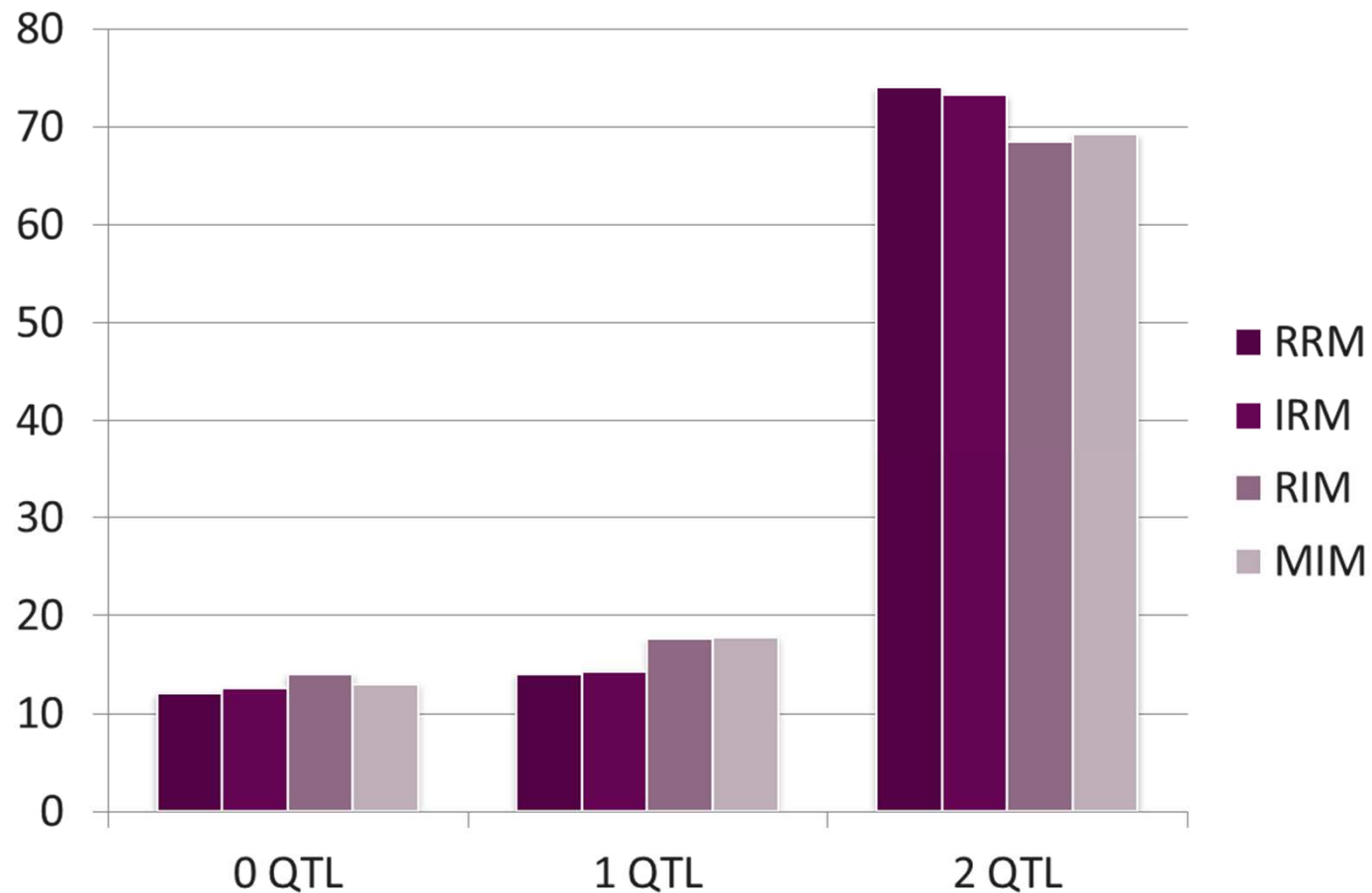
- 500 Wiederholungen
- 200 bzw. 500 Individuen
- Marker bei 0,10,20,30,40,50 cM
- QTL bei 35 cM und 45 cM
- $a_1 = 0.5$, $a_2 = -1.0$, $aa = 1.0$
- Relative QTL-Varianz 5%, 10% und 25%
- Vergleich des reduzierten Modells (RRM) mit individuellem Modell (IRM), Regressionsintervallkartierung (RIM) und multipler Intervallkartierung (MIM)



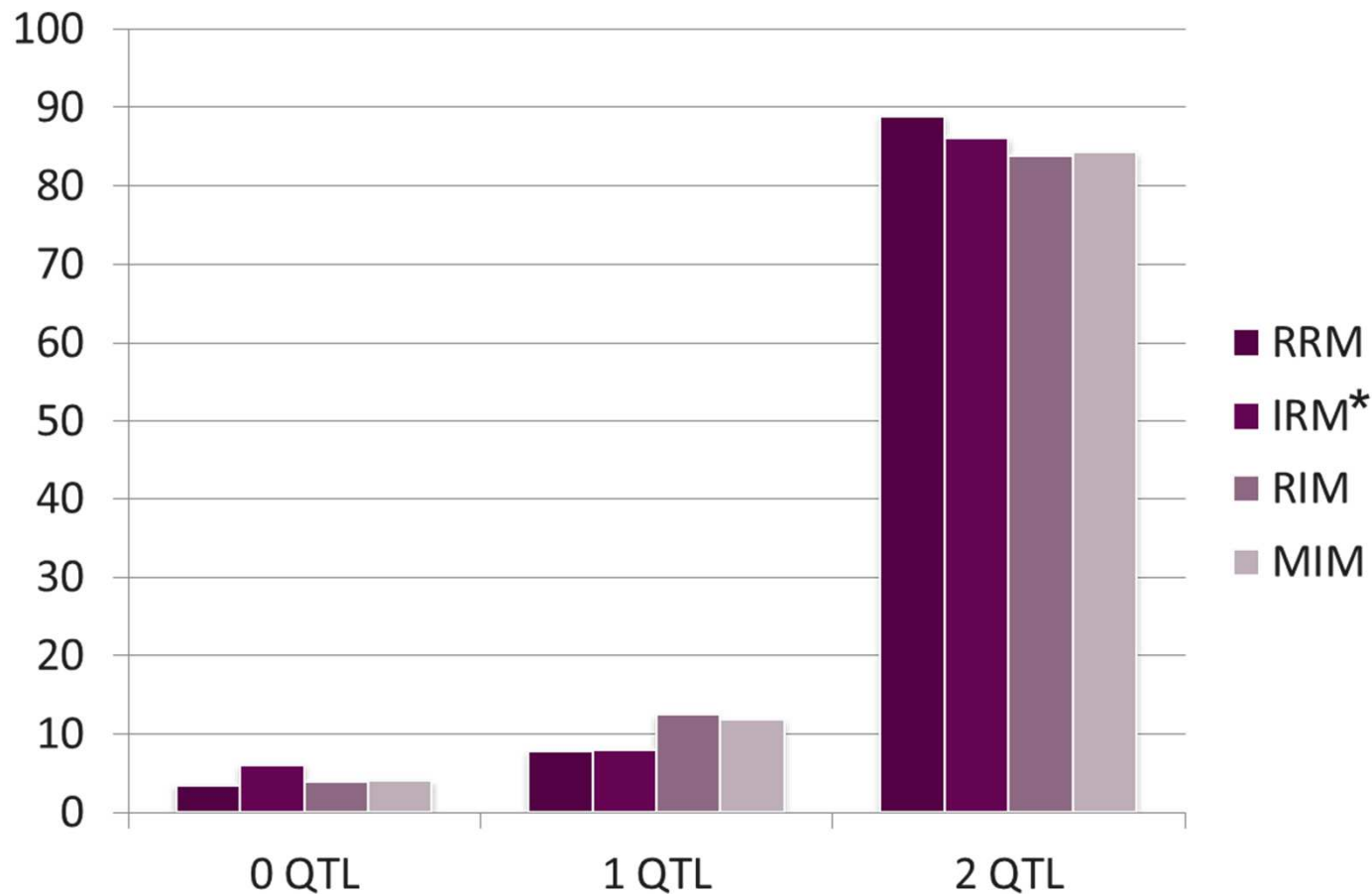
Trennschärfe – $R^2 = 5\%$, $n = 200$



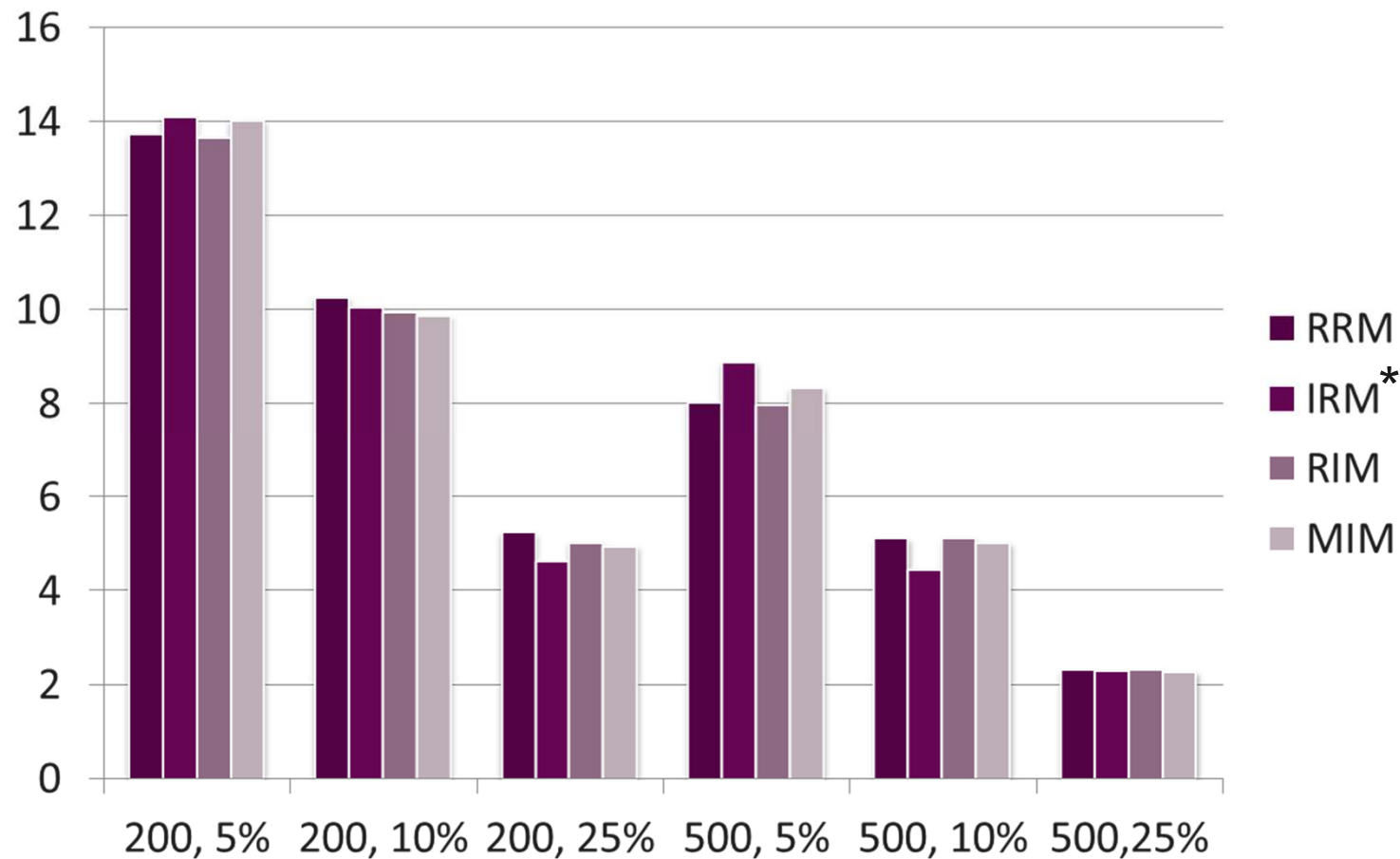
Trennschärfe – $R^2 = 10\%$, $n = 200$



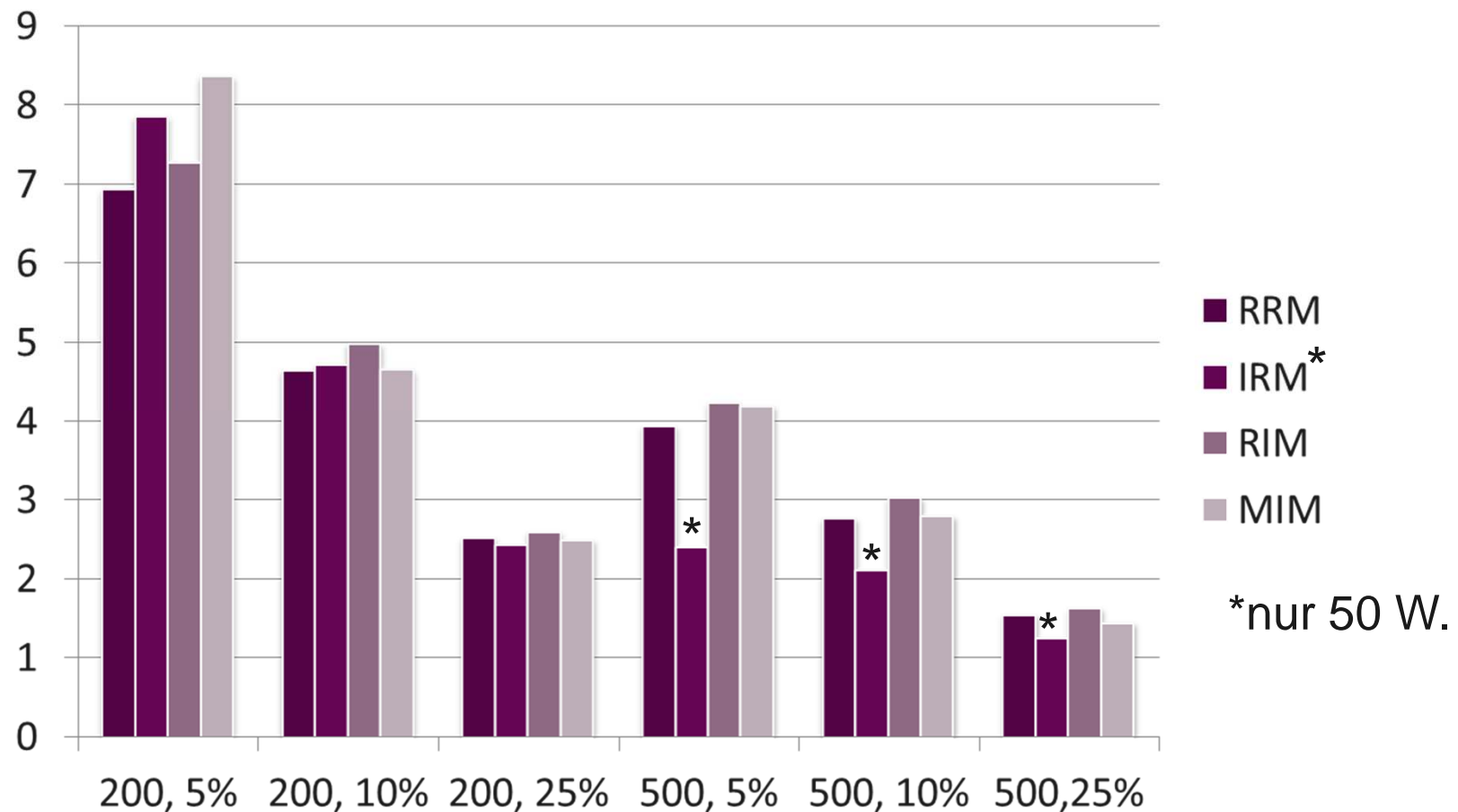
Trennschärfe – $R^2 = 5\%$, $n = 500$



RMSE für 1. QTL-Position (Familiengröße, Relative QTL-Varianz)



RMSE für 2. QTL-Position (Familiengröße, Relative QTL-Varianz)



Simulationen multiple Familien

Zwei QTL-Szenarium

- 500 Wiederholungen
- 500 Individuen
- Marker bei 0,10,20,30,40,50 cM
- QTL bei 35 cM und 45 cM
- $a_1 = 0.5$, $a_2 = -1.0$, $aa = 1.0$
- Relative QTL-Varianz 5%
- Wahrscheinlichkeit in P-Generation für Q_iQ_i bzw. q_iq_i jeweils 50 Prozent



Simulationen multiple Familien

Anzahl Familien	Familiengröße
50	10
20	25
10	50
5	100
2	250



Trennschärfe – Vergleich reduziertes / individuelles Modell

Trennschärfe reduz. / indiv.	0	1	2
50_10	34.2% / 34.8%	38.2% / 37.0%	27.6% / 28.2%
20_25	18.4% / 18.2%	41.0% / 39.8%	40.6% / 42.0%
10_50	11.4% / 11.4%	49.6% / 49.8%	39.0% / 38.8%
5_100	12.0% / 12.2%	43.6% / 43.0%	44.4% / 44.8%
2_250	22.4% / 22.2%	47.2% / 46.4%	30.4% / 31.4%

➤ Keine bedeutsamen Unterschiede



Trennschärfe

Trennschärfe reduz. / indiv.	0	1	2
50_10	34.2% / 34.8%	38.2% / 37.0%	27.6% / 28.2%
20_25	18.4% / 18.2%	41.0% / 39.8%	40.6% / 42.0%
10_50	11.4% / 11.4%	49.6% / 49.8%	39.0% / 38.8%
5_100	12.0% / 12.2%	43.6% / 43.0%	44.4% / 44.8%
2_250	22.4% / 22.2%	47.2% / 46.4%	30.4% / 31.4%



Mittelwerte, Standardfehler und RMSE (nur Ergebnisse mit 2 signifikanten QTL)

	Position 1 reduziert	Position 1 individuell	Position 2 Reduziert	Position 2 individuell
50_10	16.5 ± 13.9 23.2	16.8 ± 14.1 23.1	42.6 ± 7.5 7.8	42.9 ± 7.2 7.5
20_25	21.4 ± 13.1 18.8	21.5 ± 13.0 18.8	43.4 ± 5.5 5.7	43.4 ± 5.3 5.6
10_50	22.6 ± 12.8 17.8	22.7 ± 12.6 17.6	44.2 ± 4.9 4.9	44.2 ± 4.9 4.9
5_100	26.2 ± 12.0 14.8	26.2 ± 12.1 14.9	43.6 ± 5.5 5.7	43.5 ± 5.4 5.6
2_250	27.1 ± 11.9 14.3	27.3 ± 12.0 14.3	42.9 ± 6.2 6.5	43.1 ± 6.1 6.4

Anzahl QTL kartiert im korrekten Markerintervall

	reduziertes Modell	individuelles Modell
50_10	27	30
20_25	63	64
10_50	67	68
5_100	109	114
2_250	81	88



Diskussion multiple Familien

Xu, Genetics 148 (1989): 517 für Ein QTL-Designs:

... there seems to be an optimal sampling strategy (10_50) that leads to the highest statistical power and smallest estimation errors of QTL parameters ...



Zusammenfassung

- Asymptotische Äquivalenz vom reduzierten und individuellen Modell
- **Effektanzahl** ↓↓ im reduzierten Modell → (deutlich) schneller im Vergleich zum individuellen Modell
(Ausnahme: viele multiple Familien mit kleiner Familiengröße)
- Reduziertes Modell unabhängig von der Anzahl an F_2 -Ind.
- Güte und Genauigkeit von geschätzten QTL-Positionen nahezu gleich

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT



Literatur

- Zimmer, D., M. Mayer* and N. Reinsch (2011): Complex genetic effects in QTL identification: a computationally tractable random model for use in F2 populations. *Genetics* 187: 261-270.
- Zimmer, D. (2011): Modified random models for QTL discovery in F2 populations. Diss. agr. Kiel.

