

Erstellen gametischer Verwandtschaftsmatrizen mit SAS

Armin Tuchscherer
Forschungsinstitut für die
Biologie landwirtschaftlicher
Nutztiere (FBN) Dummerstorf
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
atuchsch@fbn-dummerstorf.de

Paul Eberhard Rudolph
Forschungsinstitut für die
Biologie landwirtschaftlicher
Nutztiere (FBN) Dummerstorf
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
pe.rudolph@fbn-
dummerstorf.de

Bernd Jäger
Institut für Biometrie und
Medizinische Informatik,
Ernst-Moritz-Arndt-Universität
Greifswald
Walter-Rathenau-Str. 48
17487 Greifswald
bjaeger@biometrie.uni-
greifswald.de

Margret Tuchscherer
Forschungsinstitut für die
Biologie landwirtschaftlicher
Nutztiere (FBN) Dummerstorf
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
mtuchsch@fbn-
dummerstorf.de

Zusammenfassung

Der sich ständig vertiefende Einblick in das Genom der Nutztiere kann zusätzliche Freiheitsgrade bei Zuchtentscheidungen ermöglichen. Die Nutzung der molekulargenetischen Information für die Tierzucht erfordert entsprechende Modelle und Verfahren zur Berücksichtigung dieser Information bei der Zuchtauswahl und Anpaarung von Tieren. Eine zentrale Rolle spielt dabei die Aufstellung der bedingten gametischen Verwandtschaftsmatrix G und ihrer Inversen unter der Bedingung gegebener Markerinformation. Die Gestalt von G ist abhängig von der Art der Gametenidentifikation. Wang et al. [6] und Abdel-Azim und Freeman [1] haben numerisch anspruchsvolle Algorithmen für die Aufstellung der gametischen Verwandtschaftsmatrix G und deren Invertierung vorgeschlagen. Im Falle von Abhängigkeiten in G , die z. B. auftreten, wenn der Mikrosatellit im Gen selbst sitzt (Rekombinationsrate = 0), bei monozygoten Zwillingen und Klonen oder bei X-chromosomaler Vererbung, ist eine Invertierung von G mit diesen Algorithmen nicht möglich. Tuchscherer et al. [5] haben die Verfahren von Wang et al. [6] und Abdel-Azim und Freeman [1] zur rekursiven Berechnung der gametischen Verwandtschaftsmatrix G und deren Inversen für den Fall von Abhängigkeiten in G verallgemeinert. Die Umsetzung dieses Algorithmus mit SAS/IML wird in diesem Beitrag vorgestellt und mit einem Demonstrationsbeispiel erläutert.

Schlüsselworte: Inverse gametische Verwandtschaftsmatrix, SAS/IML, MA-BLUP.

1 Einleitung und Problemstellung

Die Nutzung der molekulargenetischen Information für die Tierzucht erfordert entsprechende Modelle und Verfahren zur Berücksichtigung dieser Information bei der Zuchtauswahl und Anpaarung von Tieren. Die Anwendung von gemischten linearen Modellen, bei denen die genetische Varianz in einen polygenen Teil und einen molekulargenetisch erklärbaren Teil aufgeteilt wird, erweist sich als sinnvoll.

1.1 Modell und Voraussetzungen

Betrachtet wird das folgende gemischte lineare Modell $\underline{y} = X\underline{f} + Z\underline{u} + ZT\underline{v} + \underline{e}$. Dabei ist $\underline{y}_{(m \times 1)}$ der m-dimensionaler Vektor der phänotypischen Beobachtungswerte von n Individuen, $\underline{f}_{(n_r \times 1)}$ ist der Vektor der festen Effekte, $\underline{u}_{(n \times 1)}$ ist der Vektor der zufälligen polygenen Effekte and $\underline{v}_{(2n \times 1)}$ ist der Vektor der zufälligen gametischen Effekte $(\underline{v}_1^1, \underline{v}_1^2, \dots, \underline{v}_i^1, \underline{v}_i^2, \dots, \underline{v}_n^1, \underline{v}_n^2)'$ eines MQTL (marked quantitative trait locus), der eng mit einem Marker Locus (ML) gekoppelt ist. Zwischen dem ML und dem MQTL wird Kopplungsgleichgewicht angenommen. Beobachtete Markergenotypen werden im folgenden mit M bezeichnet. $X_{(m \times n_r)}$, $Z_{(m \times n)}$ sind bekannte Versuchsplanmatrizen und $T_{(n \times 2n)} = I_n \otimes [1 \ 1]$, wobei \otimes das Kroneckerprodukt bezeichnet. Zufallsvariable sind unterstrichen und die Indices in Klammern an Vektoren und Matrizen bezeichnen ihre Dimension. Die Erwartungswerte von \underline{u} , \underline{v} and \underline{e} und die Kovarianzen zwischen ihnen seien Null. Außerdem sei $\text{Cov}(\underline{u}) = \sigma_u^2 A$, $\text{Cov}(\underline{v}) = \sigma_v^2 G$, $\text{Cov}(\underline{e}) = \sigma_e^2 I$, mit der $(n \times n)$ -dimensionalen Verwandtschaftsmatrix A, der $(m \times m)$ -dimensionalen Restkovarianzmatrix I und der $(2n \times 2n)$ -dimensionalen bedingten gametischen Verwandtschaftsmatrix G sowie den Varianzkomponenten σ_u^2 , σ_v^2 und σ_e^2 der polygenen Effekte, der MQTL-Effekte und der Resteffekte.

Bezeichne α_i^1 und α_i^2 , $i=1, \dots, n$ die MQTL-Allele des Individuums i, denen die additiven Effekte $\underline{v}_i = (\underline{v}_i^1, \underline{v}_i^2)'$ des Modells zugeordnet sind. $P(\alpha_i^k \leftarrow \alpha_j^t | M)$ sei die bedingte Wahrscheinlichkeit, dass das k-te Allel α_i^k , $k = 1, 2$, von Individuum i vom t-ten Allel α_j^t , $t = 1, 2$, des Eltern j abstammt unter der Bedingung der Markerinformation M und r sei die Rekombinationsrate zwischen ML und MQTL. Mit Wang *et al.* [6] kann man die möglichen Wahrscheinlichkeiten $P(\alpha_i^k \leftarrow \alpha_j^t | M)$ der Übertragung der elterlichen MQTL-Allele ($j=s$: Vater; $j=d$: Mutter) auf Tier i als Matrix

$$Q_i = \begin{bmatrix} P(\alpha_i^1 \leftarrow \alpha_s^1 | M) & P(\alpha_i^1 \leftarrow \alpha_s^2 | M) & P(\alpha_i^1 \leftarrow \alpha_d^1 | M) & P(\alpha_i^1 \leftarrow \alpha_d^2 | M) \\ P(\alpha_i^2 \leftarrow \alpha_s^1 | M) & P(\alpha_i^2 \leftarrow \alpha_s^2 | M) & P(\alpha_i^2 \leftarrow \alpha_d^1 | M) & P(\alpha_i^2 \leftarrow \alpha_d^2 | M) \end{bmatrix} \quad (1)$$

schreiben, die im folgenden als gegeben vorausgesetzt wird. Weiterhin sei vorausgesetzt, dass ein geordnetes Pedigree vorliegt, d. h. Eltern stehen immer vor ihren Nachkommen im Pedigree. Die Tiere, die keine Eltern im Pedigree haben, bezeichnen wir als Basistiere.

Unter diesen Voraussetzungen erhalten wir mit den Lösungen der folgenden Misch-Modell-Gleichungen

$$\begin{bmatrix} X'X & X'Z & X'ZT \\ Z'X & Z'Z + \sigma_e^2/\sigma_u^2 \cdot A^{-1} & Z'ZT \\ T'Z'X & T'Z'Z & T'Z'ZT + \sigma_e^2/\sigma_v^2 \cdot G^{-1} \end{bmatrix} \cdot \begin{bmatrix} \hat{\underline{f}} \\ \hat{\underline{u}} \\ \hat{\underline{v}} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ Z'\underline{y} \\ T'Z'\underline{y} \end{bmatrix} \quad (2)$$

beste lineare erwartungstreue Schätzungen $\hat{\underline{f}}$ für \underline{f} und beste lineare erwartungstreue Vorhersagen $\hat{\underline{u}}$ und $\hat{\underline{v}}$ für \underline{u} und \underline{v} . Eine zentrale Rolle bei der Aufstellung des Gleichungssystems (2) spielt die Aufstellung der gametischen Verwandtschaftsmatrix und ihrer Inversen.

1.2 Berechnung der gametischen Verwandtschaftsmatrix G und ihrer Inversen - Problem Abhängigkeiten in G

Wang *et al.* [6] und Abdel-Azim und Freeman [1] haben Algorithmen für die Aufstellung der gametischen Verwandtschaftsmatrix G und deren Invertierung aufgestellt. Bei Wang *et al.* [6] wird G rekursiv nach

$$G_1 = I_2 \text{ und } G_i = \left[\begin{array}{c|c} G_{i-1} & G_{i-1}A'_i \\ \hline A_iG_{i-1} & C_{ii} \end{array} \right], \text{ mit } C_{ii} = \begin{bmatrix} 1 & f_i \\ f_i & 1 \end{bmatrix}, i = 2, \dots, n \text{ berechnet,} \quad (3)$$

wobei I_2 die Einheitsmatrix und f_i die bedingte Wahrscheinlichkeit dafür ist, dass 2 homologe Allele am MQTL von i herkunftsgleich sind unter der Bedingung der beobachteten Markergenotypen M (Berechnung: Formel (11) in Wang *et al.* [6]). Die Inverse von G erhält man ebenfalls rekursiv aus

$$G_1^{-1} = I_2 \text{ und } G_i^{-1} = \left[\begin{array}{c|c} G_{i-1}^{-1} & 0 \\ \hline 0 & 0 \end{array} \right] + \left[\begin{array}{c|c} A'_iD_i^{-1}A_i & -A'_iD_i^{-1} \\ \hline -D_i^{-1}A_i & D_i^{-1} \end{array} \right], \quad (4)$$

mit $D_i = (C_{ii} - A_iG_{i-1}A'_i)$, $i = 2, \dots, n$.

A_i ist eine $(2 \times 2(i-1))$ -dimensionale Matrix, die man erhält, indem die $(2s-1)$ -te und die $(2s)$ -te Spalte gleich der ersten bzw. zweiten Spalte von Q_i und deren $(2d-2)$ -te und $(2d)$ -te Spalte gleich der dritten bzw. vierten Spalte von Q_i gesetzt werden. Alle anderen Elemente von A_i sind Null, s und d sind die Nummer des Vaters (s) und der Mutter (d) von i im geordneten Pedigree.

Abdel-Azim und Freeman [1] zerlegen G mit $G = BDB'$, wobei B eine untere Dreiecksmatrix ist und D eine Blockdiagonalmatrix mit den (2×2) -Matrizen D_i (4) im i -ten Block. B kann dann rekursiv bestimmt werden aus

$$B_1 = I_2 \text{ und } B_i = \begin{bmatrix} B_{i-1} & 0 \\ A_i B_{i-1} & I_2 \end{bmatrix}, \quad i = 2, \dots, n, \quad (5)$$

wobei I_2 die Einheitsmatrix ist und die A_i wie in (3) und (4) definiert sind. Die Inverse von G erhält man dann aus

$$G^{-1} = (B')^{-1} D^{-1} B^{-1}, \text{ mit } D^{-1} = \text{diag}(D_1^{-1}, \dots, D_n^{-1}) \quad (6)$$

und rekursiv berechnetem B^{-1} :

$$B_1^{-1} = I_2 \text{ und } B_i^{-1} = \begin{bmatrix} B_{i-1}^{-1} & 0 \\ -A_i & I_2 \end{bmatrix}, \quad i = 2, \dots, n.$$

Im Falle von Abhängigkeiten in G , die z. B. auftreten, wenn der Mikrosatellit im Gen selbst sitzt (Rekombinationsrate = 0), bei monozygoten Zwillingen und Klonen oder bei X-chromosomaler Vererbung, ist eine Invertierung von G mit diesen Algorithmen nicht möglich, da D_i^{-1} in (4) und (6) nicht für alle i existiert.

1.3 Berechnung der gametischen Verwandtschaftsmatrix G und ihrer Inversen - Lösung des Problems der Abhängigkeiten in G

Tuchscherer *et al.* [5] haben die Verfahren von Wang *et al.* [6] und Abdel-Azim und Freeman [1] zur rekursiven Berechnung der gametischen Verwandtschaftsmatrix G und deren Inversen für den Fall von Abhängigkeiten in G verallgemeinert. Die Abhängigkeiten in G kommen durch identische Komponenten in \underline{y} bzw. durch als Linearkombination von Komponenten aus \underline{y} darstellbaren Komponenten von \underline{y} zustande. Die Idee ist nun, die Komponenten von \underline{y} auf die unabhängigen zu reduzieren, indem man das Ausgangsmodell umformt zu $\underline{y} = Xf + Zu + ZT \cdot L\underline{v}^* + \underline{e}$ mit $\underline{v} = L\underline{v}^*$, wobei \underline{v}^* der n^* -dimensionale Vektor der n^* verbleibenden unabhängigen Komponenten von \underline{y} und L eine $(2n \times n^*)$ -dimensionale Linkmatrix mit Zeilensummen von eins sind mit $E(\underline{v}^*) = 0$ und $\text{Cov}(\underline{v}^*) = \sigma_v^2 G^*$. G^* wird als "eingedampfte" gametische Verwandtschaftsmatrix bezeichnet, die den vollen Rang n^* hat und invertierbar ist. Die gametische Verwandtschaftsmatrix G kann man dann darstellen als $G_{(2n \times 2n)} = L G_{(n^* \times n^*)}^* L'$. Der Algorithmus von Tuchscherer *et al.* [5] gliedert sich dann in die drei Schritte:

1. Bestimmen der Indices der n^* verbleibenden gametischen Effekte
2. Bestimmen von L und
3. rekursive Berechnung von G^* und der Inversen von G^*

Es seien $N = ((N_{i,j}))$, $i = 1, \dots, n$, $j = 1, 2$ die $(n \times 2)$ -dimensionale Matrix mit den Indices der verbleibenden Komponenten \underline{v}^* von \underline{v} der n Tiere und $N_i = (N_{i,1}; N_{i,2})$ die i -te Zeile von N sowie n_b die Anzahl der Basistiere an der Spitze des Pedigrees, die unverwandt und nicht ingezüchtet sind und $n_{i-1}^{\max} = \max_{\substack{j=1, \dots, i-1 \\ k=1, 2}} \{N_{j,k}\}$, $i = 1, \dots, n$.

1. Schritt: Berechnen der Indexmatrix N

$i \leq n_b$ (Basistiere):

$$N_i = (2i - 1; 2i).$$

$i > n_b$ (keine Basistiere) and $k, j = 1, 2$:

$$N_i = \begin{cases} (N_{s(i),k}; N_{d(i),j}) & , \text{ falls } \langle Q_i(1,k) = 1 \wedge Q_i(2, j+2) = 1 \rangle \vee \langle SQ_i^k = 1 \wedge SQ_i^{2+j} = 1 \rangle \\ (N_{d(i),j}; N_{s(i),k}) & , \text{ falls } \langle Q_i(2,k) = 1 \wedge Q_i(1, j+2) = 1 \rangle \\ (N_{s(i),k}; n_{i-1}^{\max} + 1) & , \text{ falls } \langle Q_i(1,k) = 1 \wedge SQ_i^{2+j} \neq 1, \forall j \rangle \\ (N_{d(i),j}; n_{i-1}^{\max} + 1) & , \text{ falls } \langle Q_i(1, 2+j) = 1 \wedge SQ_i^k \neq 1, \forall k \rangle \\ (n_{i-1}^{\max} + 1; N_{s(i),k}) & , \text{ falls } \langle Q_i(2,k) = 1 \wedge SQ_i^{2+j} \neq 1, \forall j \rangle \\ (n_{i-1}^{\max} + 1; N_{d(i),j}) & , \text{ falls } \langle Q_i(2, 2+j) = 1 \wedge SQ_i^k \neq 1, \forall k \rangle \\ (n_{i-1}^{\max} + 1; n_{i-1}^{\max} + 2) & , \text{ falls } \langle SQ_i^k \neq 1, \forall k \wedge SQ_i^{2+j} \neq 1, \forall j \rangle \end{cases}$$

Dabei bezeichnen $N_{s(i),k}$ die Nummer des k -ten MQTL-Alleles ($k = 1, 2$) des Vaters $s(i)$ und $N_{d(i),j}$ die Nummer des j -ten MQTL-Alleles ($j = 1, 2$) der Mutter $d(i)$ von Tier i . $Q_i(o, t)$ ($o = 1, 2, t = 1, \dots, 4$) sei das t -te Element der o -ten Zeile von Q_i in (1) und SQ_i^k sei die Summe der k -ten Spalte von Q_i in (1).

2. Schritt: Berechnen der Linkmatrix L

Für jedes Tier i ($i = 1, \dots, n$) gibt es zwei Zeilen in L. $L_{2i-1,t}$ bezeichne die Elemente der ersten Zeile von i und $L_{2i,t}$ ($t = 1, \dots, n^*$) die Elemente der zweiten Zeile von i . Die Nichtnullelemente von L erhält man dann aus

$$L_{2i-1,t} = \begin{cases} \begin{cases} Q_i(1,k) \\ Q_i(1,j) \end{cases} & , \text{ falls } \langle SQ_i^k = 1 \wedge SQ_i^j = 1 \rangle \wedge \langle 1 \notin Q_i \rangle \text{ f\"ur } \begin{cases} t = N_{i,1} \\ t = N_{i,2} \end{cases} \\ 1 & , \text{ andernfalls f\"ur } t = N_{i,1} \end{cases}$$

und

$$L_{2i,t} = \begin{cases} \begin{cases} Q_i(2,k) \\ Q_i(2,j) \end{cases} & , \text{ falls } \langle SQ_i^k = 1 \wedge SQ_i^j = 1 \rangle \wedge \langle 1 \notin Q_i \rangle \text{ f\"ur } \begin{cases} t = N_{i,1} \\ t = N_{i,2} \end{cases} \\ 1 & , \text{ andernfalls f\"ur } t = N_{i,2} \end{cases}$$

wobei $1 \notin Q_i$ bedeutet, dass Q_i kein Element enthält, welches gleich eins ist.

3. Schritt: a) Rekursive Berechnung von G^*

Verallgemeinerung des Algorithmus von Wang *et al.* [6]:

Für $i=1$ ist $G_1^* = I_2$ und für $i = 2, \dots, n$ erhält man

$$G_i^* = \begin{cases} G_{i-1}^* & , \text{ falls } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ \left[\begin{array}{c|c} G_{i-1}^* & G_{i-1}^* A_i^{*k'} \\ \hline A_i^{*k} G_{i-1}^* & 1 \end{array} \right] & , \text{ falls } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max} ; k \neq j ; k, j = 1, 2 \\ \left[\begin{array}{c|c} G_{i-1}^* & G_{i-1}^* A_i^{*j'} \\ \hline A_i^{*j} G_{i-1}^* & C_{ii} \end{array} \right] & , \text{ falls } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \end{cases}$$

A_i^* ist eine $(2 \times n_{i-1}^{\max})$ -dimensionale Matrix, deren Nichtnullelemente, die $N_{s(i),1}$ -te, $N_{s(i),2}$ -te, $N_{d(i),1}$ -te, $N_{d(i),2}$ -te Spalte identisch mit der ersten, zweiten, dritten bzw. vierten Spalte von Q_i sind. A_i^{*k} ist die k -te Zeile von A_i^* , wobei $k=1$ ist falls $j=2$ oder $k=2$ ist falls $j=1$.

3. Schritt: b) Rekursive Berechnung der Inversen G^{*-1} von G^*

Verallgemeinerung des Algorithmus von Wang *et al.* [6]:

Für $i=1$ ist $G_1^{*-1} = I_2$ und für $i = 2, \dots, n$ erhält man

$$G_i^{*-1} = \begin{cases} G_{i-1}^{*-1} & , \text{ falls } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ \left[\begin{array}{c|c} G_{i-1}^{*-1} & 0 \\ \hline 0 & 0 \end{array} \right] + \left[\begin{array}{c|c} \frac{1}{d_i^*} A_i^{*k'} A_i^{*k} & -\frac{1}{d_i^*} A_i^{*k'} \\ \hline -\frac{1}{d_i^*} A_i^{*k} & \frac{1}{d_i^*} \end{array} \right] & , \text{ falls } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ \left[\begin{array}{c|c} G_{i-1}^{*-1} & 0 \\ \hline 0 & 0 \end{array} \right] + \left[\begin{array}{c|c} A_i^{*'} D_i^{*-1} A_i^{*'} & -A_i^{*'} D_i^{*-1} \\ \hline -D_i^{*-1} A_i^{*'} & D_i^{*-1} \end{array} \right] & , \text{ falls } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \end{cases}$$

mit $d_i^* = (1 - A_i^{*k} G_{i-1}^* A_i^{*k'})$ und $D_i^* = (C_{ii} - A_i^* G_{i-1}^* A_i^{*'})$.

Mit $G^* = B^* D^* B^{*'}$, $D^* = \text{diag}(\widehat{D}_1^*, \dots, \widehat{D}_n^*)$ und

$$\widehat{D}_i^* = \begin{cases} - & , \text{ falls } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ d_i^* & , \text{ falls } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ D_i^* & , \text{ falls } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \end{cases}$$

kann auch der Algorithmus von Abdel-Azim und Freeman [1] verallgemeinert werden mit rekursiv berechnetem B^* :

Für $i=1$ ist $B_1^* = I_2$ und für $i=2, \dots, n$ gilt

$$B_i^* = \begin{cases} B_{i-1}^* & , \text{ falls } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ \left[\begin{array}{c|c} B_{i-1}^* & 0 \\ \hline A_i^{*k} B_{i-1}^* & 1 \end{array} \right] & , \text{ falls } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ \left[\begin{array}{c|c} B_{i-1}^* & 0 \\ \hline A_i^* B_{i-1}^* & I_2 \end{array} \right] & , \text{ falls } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \end{cases}$$

und $G^{*-1} = (B^{*'})^{-1} D^{*-1} B^{*-1}$ mit $D^{*-1} = \text{diag}(\widehat{D}_1^{*-1}, \dots, \widehat{D}_n^{*-1})$ und rekursiv bestimmtem B^{*-1} :

Für $i=1$ ist $B_1^{*-1} = I_2$ und für $i=2, \dots, n$ erhält man

$$\mathbf{B}_i^{*-1} = \begin{cases} \mathbf{B}_{i-1}^{*-1} & , \text{ falls } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ \left[\begin{array}{c|c} \mathbf{B}_{i-1}^{*-1} & 0 \\ \hline -\mathbf{A}_i^{*k} & 1 \end{array} \right] & , \text{ falls } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ \left[\begin{array}{c|c} \mathbf{B}_{i-1}^{*-1} & 0 \\ \hline -\mathbf{A}_i^* & \mathbf{I}_2 \end{array} \right] & , \text{ falls } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \end{cases} .$$

2 Umsetzung des Algorithmus in SAS/IML

Die Verallgemeinerung des Verfahrens von Wang *et al.* [6] soll hier mit der Umsetzung des Algorithmus von Tuchscherer *et al.* [5] in SAS/IML demonstriert werden. Wir verwenden zur Demonstration zwei Beispiele.

Tabelle 1. Beispielpedigree, Markergenotypen und Übertragungswahrscheinlichkeiten Q_i (Abdel-Azim and Freeman [1]).

Tier (i)	Vater (s)	Mutter (d)	Marker- genotyp	Q_i (Rekombinationsrate = 0.1)			
1	0	0	A_1A_1	0.00	0.00	0.00	0.00
2	0	0	A_2A_2	0.00	0.00	0.00	0.00
3	0	0	A_1A_2	0.00	0.00	0.00	0.00
4	1	2	A_1A_2	0.50	0.50	0.00	0.00
				0.00	0.00	0.50	0.50
5	3	4	A_1A_1	0.45	0.05	0.45	0.05
				0.45	0.05	0.45	0.05
6	1	4	A_1A_2	0.50	0.50	0.00	0.00
				0.00	0.00	0.10	0.90
7	5	6	A_1A_2	0.50	0.50	0.00	0.00
				0.00	0.00	0.10	0.90

Mit den Daten von Tabelle 1 ist der Algorithmus von Abdel-Azim and Freeman [1] ohne Probleme anwendbar, da die resultierende gametische Verwandtschaftsmatrix den vollen Rang hat. Dagegen scheitert der Algorithmus von Abdel-Azim and Freeman [1] bei Verwendung der Daten aus Tabelle 2 bei der Erstellung der Inversen der gametischen Verwandtschaftsmatrix. Ursache des Scheiterns sind Abhängigkeiten in G . Wie wir sehen werden, führt dagegen die Verallgemeinerung des Algorithmus durch Tuchscherer *et al.* [5] in in beiden Fällen zum Ziel.

Tabelle 2. Beispielpedigree, Markergenotypen und Übertragungswahrscheinlichkeiten Q_i (Tuchscherer *et al.* [5]).

Tier (i)	Vater (s)	Mutter (d)	Marker- genotyp	Q_i (Rekombinationsrate = 0.0)			
1	0	0	A_1A_1	0.00	0.00	0.00	0.00
2	0	0	A_2A_2	0.00	0.00	0.00	0.00
3	0	0	A_1A_2	0.00	0.00	0.00	0.00
4	1	2	A_1A_2	0.50	0.50	0.00	0.00
				0.00	0.00	0.50	0.50
5	3	4	A_1A_1	0.50	0.00	0.50	0.00
				0.50	0.00	0.50	0.00
6	1	4	A_1A_2	0.50	0.50	0.00	0.00
				0.00	0.00	0.00	1.00
7	5	6	A_1A_2	0.50	0.50	0.00	0.00
				0.00	0.00	0.00	1.00

1.1 Ausgangsdatei

Voraussetzung für die Anwendung des Programms

Gametische Verwandtschaftsmatrix.sas

ist das Vorliegen der Pedigreedaten und der Übertragungswahrscheinlichkeiten (1) als SAS-Datei, wie hier am Beispiel der Daten von Tabelle 1 demonstriert.

```
/* SAS-Datei Ped: Pedigree Übertragungswahrscheinlichkeiten r=0.1*/
data Ped;
input Tier Vater Mutter Q11 Q12 Q13 Q14 Q21 Q22 Q23 Q24;
cards;
    1      0      0      0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
    2      0      0      0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
    3      0      0      0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
    4      1      2      0.50 0.50 0.00 0.00 0.00 0.00 0.50 0.50
    5      3      4      0.45 0.05 0.45 0.05 0.45 0.05 0.45 0.05
    6      1      4      0.50 0.50 0.00 0.00 0.00 0.00 0.10 0.90
    7      5      6      0.50 0.50 0.00 0.00 0.00 0.00 0.10 0.90
run;
```

In der SAS-Datei `Ped` steht die Tiernummer in der ersten Spalte, gefolgt von der Vaternummer des Tieres und der Mutternummer. Ist ein Elter unbekannt, so steht an der entsprechenden Stelle eine "0". Die Spalten vier bis sieben enthalten die vier Elemente der ersten Zeile von Q_i , gefolgt von den vier Elementen der zweiten Zeile von Q_i in den Spalten neun bis elf.

```

Gametische Verwandtschaftsmatrix.sas
34
35 /*****
36 *
37 * Programm zur Berechnung der gametischen Verwandtschaftsmatrix und ihrer Inversen mit SAS/IML *
38 *
39 * Ausgangsdatei: Pedigree und Übertragungswahrscheinlichkeiten *
40 * ----- *
41 * data Ped; *
42 * input Tier Vater Mutter Q11 Q12 Q13 Q14 Q21 Q22 Q23 Q24; *
43 * cards; *
44 * 1 0 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 *
45 * 2 0 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 *
46 * 3 0 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 *
47 * 4 1 2 0.50 0.50 0.00 0.00 0.00 0.00 0.50 0.50 *
48 * 5 3 4 0.50 0.00 0.50 0.00 0.50 0.00 0.50 0.00 *
49 * . . . *
50 * run; *
51 *
52 * Vorgabe Genauigkeitsschranke für Übertragungswahrscheinlichkeiten: EPS *
53 * (Zeile 66) *
54 *
55 *****/
56
57 proc iml;
58 use Ped; read all into Ped;
59 nt=nrow(Ped);

```

Abbildung 1: Vorgabegrößen

Neben der Pedigreeinformation und den Übertragungswahrscheinlichkeiten ist auch die Genauigkeitsschranke EPS für die Übertragungswahrscheinlichkeiten (1) vorzugeben. Standardwert ist EPS=0.

1.2 Beispielsrechnung

Wir beginnen mit Tabelle 1 und erzeugen zunächst die temporäre SAS-Datei Ped.

	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
7	Tier	Vater	Mutter	Q11	Q12	Q13	Q14	Q21	Q22	Q23	Q24	
1	1	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
2	2	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
3	3	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
4	4	1	2	0.50	0.50	0.00	0.00	0.00	0.00	0.50	0.50	
5	5	3	4	0.45	0.05	0.45	0.05	0.45	0.05	0.45	0.05	
6	6	1	4	0.50	0.50	0.00	0.00	0.00	0.00	0.10	0.90	
7	7	5	6	0.50	0.50	0.00	0.00	0.00	0.00	0.10	0.90	

Abbildung 2a: Temporäre SAS-Datei PED mit Pedigree und Übertragungswahrscheinlichkeiten von Tabelle 1.

Für die Genauigkeitsschranke EPS verwenden wir den Standardwert. Als Ergebnis des ersten und zweiten Schrittes des Algorithmus von Tuchscherer *et al.* [5] erhält man die Indices der verbleibenden 14 Effekte. Es erfolgte also keine Reduzierung. Folglich muß

auch die Linkmatrix L eine Einheitsmatrix sein, wie man in Abbildung 2b leicht erkennen kann.

WORK.N

7	2	Int	Int
	N1	N2	
1	1	2	
2	3	4	
3	5	6	
4	7	8	
5	9	10	
6	11	12	
7	13	14	

WORK.L

14	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
14	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10	COL11	COL12	COL13	COL14	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	

Abbildung 2b: Temporäre SAS-Dateien N mit den Indices der verbleibenden Effekte und L mit der Linkmatrix L (Daten: Tabelle 1).

Im dritten Schritt erhält man dann die gametische Verwandtschaftsmatrix G und ihre Inverse G^{-1} , die in den entsprechenden temporären SAS-Dateien (Abbildung 2c) abgelegt werden.

WORK.G

14	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
14	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10	COL11	COL12	COL13	COL14
1	1.000	0.000	0.000	0.000	0.00	0.00	0.50	0.00	0.2250	0.2250	0.500	0.050	0.2250	0.0950
2	0.000	1.000	0.000	0.000	0.00	0.00	0.50	0.00	0.2250	0.2250	0.500	0.050	0.2250	0.0950
3	0.000	0.000	1.000	0.000	0.00	0.00	0.00	0.50	0.0250	0.0250	0.000	0.450	0.0250	0.4050
4	0.000	0.000	0.000	1.000	0.00	0.00	0.00	0.50	0.0250	0.0250	0.000	0.450	0.0250	0.4050
5	0.000	0.000	0.000	0.000	1.00	0.00	0.00	0.00	0.4500	0.4500	0.000	0.000	0.4500	0.0000
6	0.000	0.000	0.000	0.000	0.00	1.00	0.00	0.00	0.0500	0.0500	0.000	0.000	0.0500	0.0000
7	0.500	0.500	0.000	0.000	0.00	0.00	1.00	0.00	0.4500	0.4500	0.500	0.100	0.4500	0.1400
8	0.000	0.000	0.500	0.500	0.00	0.00	0.00	1.00	0.0500	0.0500	0.000	0.900	0.0500	0.8100
9	0.225	0.225	0.025	0.025	0.45	0.05	0.45	0.05	1.0000	0.0000	0.225	0.090	0.5000	0.1035
10	0.225	0.225	0.025	0.025	0.45	0.05	0.45	0.05	0.0000	1.0000	0.225	0.090	0.5000	0.1035
11	0.500	0.500	0.000	0.000	0.00	0.00	0.50	0.00	0.2250	0.2250	1.000	0.050	0.2250	0.1450
12	0.050	0.050	0.450	0.450	0.00	0.00	0.10	0.90	0.0900	0.0900	0.050	1.000	0.0900	0.9050
13	0.225	0.225	0.025	0.025	0.45	0.05	0.45	0.05	0.5000	0.5000	0.225	0.090	1.0000	0.1035
14	0.095	0.095	0.405	0.405	0.00	0.00	0.14	0.81	0.1035	0.1035	0.145	0.905	0.1035	1.0000

WORK.GINV

14	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
14	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10	COL11	COL12	COL13	COL14
1	2	1	0.0	0.0	0.00	0.0000	-1.0000	0.0000	0.0000	0.0000	-1.0000	0.0000	0	0.0000
2	1	2	0.0	0.0	0.00	0.0000	-1.0000	0.0000	0.0000	0.0000	-1.0000	0.0000	0	0.0000
3	0	0	1.5	0.5	0.00	0.0000	0.0000	-1.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
4	0	0	0.5	1.5	0.00	0.0000	0.0000	-1.0000	0.0000	0.0000	0.0000	0.0000	0	0.0000
5	0	0	0.0	0.0	3.25	0.2500	2.2500	0.2500	-2.5000	-2.5000	0.0000	0.0000	0	0.0000
6	0	0	0.0	0.0	0.25	1.0278	0.2500	0.0278	-0.2778	-0.2778	0.0000	0.0000	0	0.0000
7	-1	-1	0.0	0.0	2.25	0.2500	4.3056	0.7500	-2.5000	-2.5000	0.0000	-0.5556	0	0.0000
8	0	0	-1.0	-1.0	0.25	0.0278	0.7500	6.5278	-0.2778	-0.2778	0.0000	-5.0000	0	0.0000
9	0	0	0.0	0.0	-2.50	-0.2778	-2.5000	-0.2778	3.7778	2.7778	0.0000	0.0000	-1	0.0000
10	0	0	0.0	0.0	-2.50	-0.2778	-2.5000	-0.2778	2.7778	3.7778	0.0000	0.0000	-1	0.0000
11	-1	-1	0.0	0.0	0.00	0.0000	0.0000	0.0000	0.0000	0.0000	2.0585	0.5263	0	-0.5848
12	0	0	0.0	0.0	0.00	0.0000	-0.5556	-5.0000	0.0000	0.0000	0.5263	10.2924	0	-5.2632
13	0	0	0.0	0.0	0.00	0.0000	0.0000	0.0000	-1.0000	-1.0000	0.0000	0.0000	2	0.0000
14	0	0	0.0	0.0	0.00	0.0000	0.0000	0.0000	0.0000	0.0000	-0.5848	-5.2632	0	5.8480

Abbildung 2c: Temporäre SAS-Dateien G mit der gametischen Verwandtschaftsmatrix G sowie GINV mit ihrer Inversen G^{-1} (Daten: Tabelle 1).

Die Ergebnisse in Abbildung 2c stimmen mit den Ergebnissen von Abdel-Azim and Freeman [1] überein.

Bei den Daten von Tabelle 2 verbleiben mit dem Algorithmus von Tuchscherer *et al.* [5] 10 Effekte (Abbildung 3a). Mit der Linkmatrix L in Abbildung 3a wird ersichtlich, dass

The image shows two SAS temporary files. The first, WORK.N, is a small table with 7 rows and 3 columns. The second, WORK.L, is a larger table with 14 rows and 12 columns, representing a link matrix L.

	Int	Int
7	N1	N2
1	1	2
2	3	4
3	5	6
4	7	8
5	5	7
6	9	8
7	10	8

10	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
14	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10	
1	1	0	0	0	0.0	0	0.0	0	0	0	
2	0	1	0	0	0.0	0	0.0	0	0	0	
3	0	0	1	0	0.0	0	0.0	0	0	0	
4	0	0	0	1	0.0	0	0.0	0	0	0	
5	0	0	0	0	1.0	0	0.0	0	0	0	
6	0	0	0	0	0.0	1	0.0	0	0	0	
7	0	0	0	0	0.0	0	1.0	0	0	0	
8	0	0	0	0	0.0	0	0.0	1	0	0	
9	0	0	0	0	0.5	0	0.5	0	0	0	
10	0	0	0	0	0.5	0	0.5	0	0	0	
11	0	0	0	0	0.0	0	0.0	0	1	0	
12	0	0	0	0	0.0	0	0.0	1	0	0	
13	0	0	0	0	0.0	0	0.0	0	0	1	
14	0	0	0	0	0.0	0	0.0	1	0	0	

Abbildung 3a: Temporäre SAS-Dateien N mit den Indices der verbleibenden Effekte und L mit der Linkmatrix L (Daten: Tabelle 2).

The image shows two SAS temporary files. The first, WORK.G, is a table with 10 rows and 11 columns showing the damped gametic relationship matrix G*. The second, WORK.GINV, is a table with 10 rows and 11 columns showing the inverse of G*.

10	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
10	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	1.00	0.00	0.0	0.0	0.0	0	0.5	0.0	0.50	0.25
2	0.00	1.00	0.0	0.0	0.0	0	0.5	0.0	0.50	0.25
3	0.00	0.00	1.0	0.0	0.0	0	0.0	0.5	0.00	0.00
4	0.00	0.00	0.0	1.0	0.0	0	0.0	0.5	0.00	0.00
5	0.00	0.00	0.0	0.0	1.0	0	0.0	0.0	0.00	0.50
6	0.00	0.00	0.0	0.0	0.0	1	0.0	0.0	0.00	0.00
7	0.50	0.50	0.0	0.0	0.0	0	1.0	0.0	0.50	0.50
8	0.00	0.00	0.5	0.5	0.0	0	0.0	1.0	0.00	0.00
9	0.50	0.50	0.0	0.0	0.0	0	0.5	0.0	1.00	0.25
10	0.25	0.25	0.0	0.0	0.5	0	0.5	0.0	0.25	1.00

10	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
10	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	2	1	0.0	0.0	0.0	0	-1.0	0	-1	0
2	1	2	0.0	0.0	0.0	0	-1.0	0	-1	0
3	0	0	1.5	0.5	0.0	0	0.0	-1	0	0
4	0	0	0.5	1.5	0.0	0	0.0	-1	0	0
5	0	0	0.0	0.0	1.5	0	0.5	0	0	-1
6	0	0	0.0	0.0	0.0	1	0.0	0	0	0
7	-1	-1	0.0	0.0	0.5	0	2.5	0	0	-1
8	0	0	-1.0	-1.0	0.0	0	0.0	2	0	0
9	-1	-1	0.0	0.0	0.0	0	0.0	0	2	0
10	0	0	0.0	0.0	-1.0	0	-1.0	0	0	2

Abbildung 3b: Temporäre SAS-Dateien G mit der eingedampften gametischen Verwandtschaftsmatrix G^* sowie GINV mit ihrer Inversen G^{*-1} (Daten: Tabelle 2).

die Effekte des zweiten QTL-Allels der Tiere 4, 6 und 7 identisch sind und die Effekte der QTL-Allele des Tieres 5 sich jeweils als Linearkombination der Effekte der ersten Allele der Elterntiere ergeben. Somit wären die Algorithmen von Wang *et al.* [6] und Abdel-Azim and Freeman [1] für $i = 5, 6$ und 7 gescheitert, da die D_i in (4) und (6) wegen $\det(D_i) = 0$ nicht invertierbar sind.

Literatur

- [1] Abdel-Azim G., Freeman A.E. (2001): A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked Quantitative Trait Locus, *Genet. Sel. Evol.* 33, 153-173.
- [2] SAS Institute Inc. (1999): SAS Macro Language: Reference, Version 8, Cary, NC: SAS Institute Inc.
- [3] SAS Institute Inc. (1999): SAS/IML User's Guide, Version 8, Cary, NC: SAS Institute Inc.
- [4] SAS Institute Inc. (1999): SAS/STAT® User's Guide, Version 8, Cary, NC: SAS Institute Inc.
- [5] Tuchscherer A., Mayer, M., Reinsch, N. (2004): Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse, *Genet. Sel. Evol.* 36, 621-642.
- [6] Wang T., Fernando R.L., van der Beek S., Grossman M., van Arendonk J.A.M. (1995): Covariance between relatives for a marked quantitative trait locus, *Genet. Sel. Evol.* 27, 251-274.