The founders of a pedigreed population are the very first individuals with no further recorded ancestors. They are usually treated as unrelated and non-inbred for setting up relationship matrices. However, treating founders of a genealogy as related has been shown to be a useful concept (Legarra et al. 2015) when genomic relationships (VanRaden 2008) and pedigree information are to be combined into a joint relationship matrix. The main achievement of taking founder relatedness into account is a suitable scaling of pedigree relationships (Legarra et al. 2015), which makes them compatible with genomic relationships. Other benefits are reasonably interpretable estimates of genetic variance components and the prediction of genetic trends (Legarra et al. 2015). Founder relationships can be estimated from marker data of genotyped individuals (Christensen 2012; Legarra et al. 2015; Colleau et al. 2017), which are usually only available for younger generations in ongoing breeding programs. In the context of mapping of quantitative trait loci (QTL) in line-cross experiments, inferences on within-line relationships between QTL-genotypes of founders can also be made from variance components and related likelihoods (Rönneberg et al. 2008).

In the context of crossbreeding, founders comprise individuals from two or more genetically distinct populations. This requires relationship coefficients for each single population, in addition to a combination of populations (Legarra et al. 2015). Here, the aim is to model relationships between purebreds and also between purebreds and crossbreds, most frequently from the F1 generation. Applications are in genetic evaluations, where purebred and crossbred performances are treated as genetically correlated traits (Aguilar et al. 2010; Pszczola et al. 2012; de los Campos et al. 2013; Garcia-Baccino et al. 2017).
the interest is in the genetic (co-)variance components for these traits in the purebred populations, where selection takes place.

A somewhat different focus exists when composite populations are generated, e.g., for selection experiments in laboratory animals (Holt et al. 2005) or when building advanced intercross lines (AIL, Darvasi and Soller 1995) for fine-mapping purposes. In this case, two or more genetically distinct lines, in some cases inbred lines, are intercrossed. The population is further developed from generation to generation by inter-mating crossbreds. Only performance traits of the intercross are entered into the genetic makeup of the composite population.

THEORY

Underlying assumptions

We assume two distinct founder populations; A and B, that contribute to a composite crossbred population. All founders are assumed to be genotyped and line specific founder allele frequencies are known. For each marker $i$, founder frequencies are denoted as $p_i^A$ and $p_i^B$. Under the condition that all founders contribute equally to the composite population, the expected allele frequency in the F2 generation $p_i$ is fully determined as the average, $p_i = \frac{1}{2}(p_i^A + p_i^B)$, of the two line specific frequencies.

Observed genomic founder relationships

The genotypes of the founders can be summarized into a centered genotype matrix, $Z$, with one row per individual and one column per marker.

**Autosomes**: For autosomal markers, entries into the $Z$ matrix are $z_i^A \in (2 - 2p_i, 1 - 2p_i, 0 - 2p_i)$ for genotypes AA, Aa and aa, respectively. The observed genomic relationship matrix, $G_0$, between founders is

$$G_0 = Z_0 Z_0' / S$$

where $S$ is the scaling factor; $S = 2\sum p_i(1 - p_i)$. $G_0$ is a standard genomic relationship matrix, except when using $p_i$ for centering and scaling, as previously described by Van Raden (2008).

**X-Chromosome**: The observed X-chromosomal genomic relationship matrix is set up in accordance with the rules for autosomal markers. Extra details apply to the definitions of average gene frequencies and the treatment of male (hemizygous) individuals. For X-chromosome markers we define the mean allele frequency $\bar{p}_i$, again, as $\bar{p}_i = \frac{1}{2}(p_i^A + p_i^B)$. However, on the X-chromosome this is only equal to the gene frequency to be ultimately reached in later generations if the two founder lines contribute equally through males and females to the

Expected founder genomic relationships

**Autosomes**: An expectation of $G_0$ (denoted $E(G_0)$) can be derived under the assumption that alleles at independent loci are randomly sampled from each founder line’s particular gene pool (Binomial sampling), as defined by their known founder allele frequencies. We consider two gametes randomly chosen from base population A. The $2 \times 2$ matrix $R_{AA}$ has the expected sum of squared centered coefficients $\bar{z}_i^A (\bar{z}_i^A \in (1 - p_i, 0 - p_i))$, for alleles A and a, respectively for each of the gametes on its diagonal and the corresponding expected sum of cross products on its off diagonal.

$$R_{AA} = \begin{bmatrix} d_{AA} & r_{AA} \\ r_{AA} & d_{AA} \end{bmatrix}$$

$$= \frac{1}{5} \left[ \sum_i [p_i^A (1 - p_i) + (p_i^A - \bar{p}_i)^2] \sum_i (p_i^A - \bar{p}_i)^2 \right]$$

$$- \sum_i (p_i^A - \bar{p}_i)^2 \sum_i [p_i^A (1 - p_i) + (p_i^A - \bar{p}_i)^2].$$

For base population B, the equivalent matrix $R_{BB}$ can be derived from its line specific allele frequencies $p_i^B$. In the following matrices, $R_{AA}$ and $R_{BB}$ are referred to as the expected covariance matrices between gametes from the same founder line.

Furthermore, we can set up $R_{AB}$ as the equivalent relationship matrix of two randomly chosen gametes from populations A and B:

$$R_{AB} = \begin{bmatrix} d_{AB} & r_{AB} \\ r_{AB} & d_{AB} \end{bmatrix}$$

$$= \frac{1}{5} \left[ \sum_i [p_i^A (1 - p_i) + (p_i^A - \bar{p}_i)^2] \sum_i (p_i^A - \bar{p}_i)(p_i^B - \bar{p}_i) \right]$$

$$- \sum_i (p_i^A - \bar{p}_i)(p_i^B - \bar{p}_i) \sum_i [p_i^B (1 - p_i) + (p_i^B - \bar{p}_i)^2].$$

From the distinct elements in $R_{AA}$ and $R_{AB}$ (also $R_{BB}$) we can compute all necessary expected relationships between individuals, which may occur in $E(G_0)$. First, we have the expected self-relationship of an individual from a particular founder line (e.g., line A)

$$\omega^A = 2d_{AA} + 2r_{AA}.$$
The expected relationship between two individuals from two different lines, respectively, and the parents of another, as known from the tabular method. The resulting expected genomic relationship matrix is denoted by \( v \). The expected self-relationships and relationships are:

\[
\begin{align*}
\text{Self-relationship} & \\
\omega^A_A & = 4r^A_A. \\
\omega^B_B & = 4r^B_B. \\
\end{align*}
\]

**X-chromosome:** With regard to expected X-chromosomal relationships, different combinations between either males or females may occur, in addition to the same or different founder lines. In total, there are eight different kinds of possible relationships: the self-relationship of a male originating from a certain line, \( \omega_{xx}^x \); the expected relationship between two males from the same line, \( \omega_{xx}^{AA, AB} \); the self-relationship, \( \omega_{xx}^A \), of a female from a certain line; the relationships \( \omega_{xx,xx}^A \) and \( \omega_{xx,xx}^{AB} \) between two females from the same line and different lines, respectively, and finally the relationships \( \omega_{xx,xx}^{AA, AB} \) between a male and a female from different lines. Formulas for all eight cases are summarized in Table 1.

### Extending expected genomic relationships to later generations

Expected founder genomic relationships can be extended to all descendants by following the paths of Mendelian transmission, as specified in the pedigree. The resulting expected genomic relationship matrix is denoted by \( \tilde{A} \). The diagonal elements \( \tilde{g}_{kk}^x \) and off-diagonal elements \( \tilde{g}_{kl}^x \) of matrix \( \tilde{A} \) are computed by a modified version of the tabular method (Emik and Terrill 1949; Cruden 1949).

**Autosomes:** The expected autosomal self-relationships (diagonal elements of \( \tilde{A} \)) consist of three parts: the expected self-relationships of gametes inherited from the sire, from the dam and the gametic relationship between these parental gametes. Relationships between individuals (off-diagonal elements) are an average of the relationships of one candidate and the parents of another, as known from the tabular method. The expected self-relationships and relationships are:

\[
\begin{align*}
\tilde{g}_{kk}^x & = s_{sire} + s_{dam} + \frac{1}{2} \tilde{g}_{sire,dam}^x \\
\tilde{g}_{kl}^x & = \tilde{g}_{sire,l}^x + \frac{1}{2} \tilde{g}_{sire,dam}^x + \frac{1}{2} \tilde{g}_{dam,l}^x - \frac{1}{4} \tilde{g}_{sire,dam}^x
\end{align*}
\]

**X-chromosome:** For the X-chromosome, equations (7) and (8) are modified for females to give

\[
\tilde{g}_{kl}^x = \tilde{g}_{sire,l}^x + \tilde{g}_{dam,l}^x + \frac{1}{2} \tilde{g}_{sire,dam}^x
\]

and for males to give

\[
\tilde{g}_{kl}^x = \tilde{g}_{dam,l}^x + \frac{1}{2} \tilde{g}_{sire,dam}^x
\]

---

### Table 1 Formulas for calculating the different types of expected X-chromosomal relationships of males and females in a crossbred population derived from two founder lines A and B

| Relationship | Formula |
|--------------|---------|
| Self-relationship | \( \omega^A_A = 4r^A_A \) |
| Relationship between individuals | \( \omega_{xx}^{AA, AB} = \frac{1}{2} \left( \omega_{xx}^A + \omega_{xx}^B \right) \) |
| Expected founder genomic relationships | \( \tilde{g}_{kk}^x = s_{sire} + s_{dam} + \frac{1}{2} \tilde{g}_{sire,dam}^x \) |
| Autosomal | \( \tilde{g}_{kl}^x = \tilde{g}_{sire,l}^x + \tilde{g}_{dam,l}^x + \frac{1}{2} \tilde{g}_{sire,dam}^x \) |
| X-chromosomal | \( \tilde{g}_{kl}^x = \tilde{g}_{sire,l}^x + \tilde{g}_{sire,dam}^x + \tilde{g}_{dam,l}^x + \frac{1}{2} \tilde{g}_{sire,dam}^x \) |
Note that Fernando and Grossman (1990) introduced similar equations for calculating the pedigree-derived relationships for the X-chromosome, were the self-relationship for males, \(s_{xe}\), is 0.5 in all generations, meaning \(s_{xem}\) must also be 0.5. The underlying assumption is that allele frequencies of both sexes are equal, as it is the case if the population is in an equilibrium state (Sef and Liang 1987). Equations (9) to (11), in contrast, allow \(s_{xe}\) and \(s_{xem}\) to fluctuate along with male and female marker frequencies in early generations, after mating male and female founders with differing allele frequencies.

**Expected genomic co-variances in an infinitely large F2 population**

In a cross between two populations, A and B, the gametes that an individual receives in the F2 generation are of types A and B, with equal probability. Such an individual will have a probability of 0.25 of inheriting two A gametes, a probability of 0.25 of inheriting two B gametes and a probability of 0.5 of inheriting one A gamete and one B game.

The average self-relationship of an individual in an infinitely large F2 population, is therefore

\[
\frac{1}{4}(2d^{AA} + 2r^{AA}) + \frac{1}{4}(2d^{BB} + 2r^{BB}) + \frac{1}{2}(d^{AB} + d^{BA} + 2r^{AB})
\]

\[
= \frac{1}{4}(2d^{AA} + 2r^{AA}) + \frac{1}{2}(d^{AB} + d^{BA} + 2r^{AB})
\]

\[
= d^{AA} + d^{BB}
\]

\[
= 1
\]

as \(r^{AA} = r^{BB} = -r^{AB}\) in a cross of two lines.

The average covariance between F2 individuals can be derived as the weighted average of covariances in nine possible combinations of two individuals, each of them may carry two (AA), one (AB) or no (BB) gametes from the A line. These nine single pair-wise covariances can be expressed in terms of relationships between gametes, i.e. \(r^{AA}, r^{BB}\) and \(r^{AB}\). The weights are the probabilities of the occurrence of all these combinations:

\[
\frac{1}{16}4r^{AA} + \frac{2}{16}(2d^{AA} + 2r^{AB}) + \frac{1}{16}2r^{AB}
\]

\[
+ \frac{2}{16}(2d^{AA} + 2r^{AB}) + \frac{4}{16}(d^{AA} + d^{BB} + 2r^{AB}) + \frac{2}{16}(2d^{BB} + 2r^{AB})
\]

\[
+ \frac{1}{16}4r^{AB} + \frac{2}{16}(2d^{AB} + 2r^{BB}) + \frac{1}{16}2r^{BB} = 0.
\]

The expected covariance matrix \(\tilde{A}\) for unrelated and non-inbred F2 individuals is, therefore, an identity matrix. This means that such a hypothetical population can be viewed as a reference for the actual composite population derived from the genotyped founders. Variance components estimated with an \(\tilde{A}\) matrix can, after proper adjustment for founder relationships (Legarra et al. 2015), be interpreted as the combined relationship matrix, \(\tilde{A}\), and genomic relationships into a joint matrix, \(H\). In our case we used \(H\) as a modification of \(\tilde{A}\) that is corrected for the observed founder relationships in \(G_0\). We denote \(\tilde{A}\) as the expected genomic relationships of founders. In terms of the inverse of \(H\) (Christensen and Lund 2010; Aguilar et al. 2010), we then get

\[
H^{-1} = \tilde{A}^{-1} + \begin{bmatrix} G_0^{-1} - \tilde{A}_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}
\]

\[
= \tilde{A}^{-1} + \begin{bmatrix} G_0^{-1} - E(G_0)^{-1} \\ 0 & 0 \end{bmatrix}
\]

The inverse of matrix \(\tilde{A}\) is described in the supplement. The observed founder genomic relationship matrix, \(G_0\), may be singular. In this case, one may capitalize on the idea of blending (Garcia-Baccino et al. 2017). We used \(G_0^{BLEND} = 0.98 \times G_0 + 0.02 \times E(G_0)\), instead of \(G_0\), when computing \(H^{-1}\).

**Joint relationships from simulated genotypes**

For the sake of comparison with \(H\), a joint relationship matrix, \(G\), was generated from observed genotypes of founders plus simulated genotypes of non-founders. The alleles were randomly sampled from observed founder genotypes and simulated marker genotypes of offspring in later generations were derived by gene-drop. The expected self-relationships for autosomes were calculated as

\[
g^E_{k1,k2} = \frac{1}{S} \sum_i E(\{\frac{1}{2} c^1_i - 2 \tilde{p}_i\} \{\frac{1}{2} c^2_i - 2 \tilde{p}_i\})
\]

where \(g^E_{k1,k2}\) is the expected relationship between individuals \(k1\) and \(k2\), \(c^1_i\) and \(c^2_i\) are the gene counts of individuals \(k1\) and \(k2\) at locus \(i\). For self-relationship we used \(g^E_{k,k} = \frac{1}{S} \sum_i E[(c^1_i - 2 \tilde{p}_i)^2]\). Expectations were obtained by averaging over 10000 replicates of the gene-drop simulation.

For X-chromosomes, the sex of the descendants must also be considered. For female descendants equation (13) remains. For male descendants equation (13) will be modified by applying the centered X-chromosomal genotypes \(x_i^k \in (1 - \tilde{p}_i, 0 - \tilde{p}_i)\), which gives

\[
g^E_k = \frac{1}{S} \sum_i E[(c_i^k - \tilde{p}_i)^2]
\]

**APPLICATION EXAMPLE**

**Animals, pedigree, phenotypes and genotypes**

The advanced intercross mouse line (AIL) bred in the Leibniz Institute for Farm Animal Biology (FBN) was established by randomly choosing and intercrossing four females from the long-term selected, high-fecundity line, FL1 (Langhammer et al. 2014), and four males from a highly inbred (theoretical inbreeding coefficient > 0.999) control line, DUKsi (Alm et al. 2010). Both lines were derived from the same initial gene pool (Dietl et al. 2004).

The high fecundity FL1 line was selected for an index trait that combines litter size (LSO) and litter weight (LWO) at birth in primiparous females (Index I = 1.6 × LSO + LWO) up to generation 131. As a result of selection over 131 generations an average of 17.14 ± 3.25 pups per litter had been reached. This is a 1.8 fold higher fecundity than observed in the control line (see Table 2). An outbred control line, DUKs, was maintained at approximately the same population size for 79 generations by
random mating and without any selection pressure. The inbred derivative DUKs was split from DUKs in generation 79.

Four male founders were chosen for the experiment after 38 generations of full sib mating in the DUKs line. Each of four females from generation 131 of the FL1 selection line was mated with one male from the control line. The F1 litters were standardized to four male and eight female pups immediately after birth in order to maintain a surplus of females for further reproduction. Full sibs from the four initial F1 families were then repeatedly (at least four times) inter-mated by rotating males and females within the family. Thus, each of the four pairs of founder parents constitutes a family of its own, with descendants up to generation F3. Offspring of only one of these families were then maintained and became the ancestors of all further generations of the AIL.

A total of 19266 mice (9453 males and 9813 females) were used for this study. They were distributed unevenly across all generations; 44 in F1, 1483 in F2, 5235 in F3, 1025, 1058 and 1070 in F23, F24 and F25, respectively, and between 312 and 431 for other generations.

Reproductive ability was measured as litter size at birth (LS0) and litter weight at birth (LW0). The litter traits were recorded for 4430 females for further reproduction. Full sibs from the four initial F1 families were then repeatedly (at least four times) inter-mated by rotating males and females within the family. Thus, each of the four pairs of founder parents constitutes a family of its own, with descendants up to generation F3. Offspring of only one of these families were then maintained and became the ancestors of all further generations of the AIL.

A total of 19266 mice (9453 males and 9813 females) were used for this study. They were distributed unevenly across all generations; 44 in F1, 1483 in F2, 5235 in F3, 1025, 1058 and 1070 in F23, F24 and F25, respectively, and between 312 and 431 for other generations.

Reproductive ability was measured as litter size at birth (LS0) and litter weight at birth (LW0). The litter traits were recorded for 4430 females (from 9813 females) for their first litter. Among these females, 1481 also had a record for their second litter (there were no second litter records for generations from F2 to F21). Growth traits that were recorded for all females for further reproduction. Full sibs from the four initial F1 families were then repeatedly (at least four times) inter-mated by rotating males and females within the family. Thus, each of the four pairs of founder parents constitutes a family of its own, with descendants up to generation F3. Offspring of only one of these families were then maintained and became the ancestors of all further generations of the AIL.

A total of 19266 mice (9453 males and 9813 females) were used for this study. They were distributed unevenly across all generations; 44 in F1, 1483 in F2, 5235 in F3, 1025, 1058 and 1070 in F23, F24 and F25, respectively, and between 312 and 431 for other generations.

Reproductive ability was measured as litter size at birth (LS0) and litter weight at birth (LW0). The litter traits were recorded for 4430 females (from 9813 females) for their first litter. Among these females, 1481 also had a record for their second litter (there were no second litter records for generations from F2 to F21). Growth traits that were recorded for all generations were body weight at day 21, 42 and 63 (BM21, BM42, and BM63), in addition to body weight at first mating (BMM).

The six fertility and growth traits (see Table 3 for summary statistics) were analyzed using different kinds of relationship matrices, as described below.

All eight founders of this intercross line were genotyped with the JAX Mouse Diversity Genotyping Array (Yang et al. 2009) at the genotyping facility of the Jackson Laboratory, The Jackson Laboratory, Bar Harbor, Maine, USA.

**Comparative estimation of variance components**

Two fertility and four growth traits were comparatively analyzed with mixed models that comprised of different kinds of relationship matrices.

Fertility traits (LS0, LW0) were analyzed as traits of females (Langhammer et al. 2017). For both traits, the model for the iith observation (i = 1, 2) of animal a was

\[ y_{ia} = \mu + \beta_g a_i + \gamma w_{ia} + b_i + u_i + e_{ia} \]

(14)

where \( \beta_g \) is the fixed generation effect for females born in generations from F2 to F29 \( (g = 1, ..., 29) \) and \( \gamma \) is the linear regression of the body weight \( w_a \) at mating of each female’s mother. The random part of the model comprised the additive autosomal and X-chromosomal genetic effects \( a_i \) and \( x_i \) of animal \( a \) \((a=1, ..., 19266)\), the common litter environmental effect \( u_i \) \((i=1, ..., 2420)\), the permanent environmental effect \( u_p \) \((p=1, ..., 4430)\) and the residual \( e_{ia} \).

The covariance matrices of the common litter environmental and permanent environmental effects were assumed as equal to an identity matrix of proper size times the respective variance component.

Observations of growth traits (BM21, BM42, BM63, and BMM) were made from males and females from generations F2 to F30. Therefore, the fixed part of the model also included an additional sex effect, \( \beta_s \), and the number of levels was 29 for the generation effect. For BM21 and BMM no permanent environmental effect could be fitted because these traits were only measured once for all animals.

Table 4 gives more details of the model we applied.

All random effects were assumed to be mutually independent. Three different kinds of relationship matrices were compared for autosomal and X-chromosomal genetic effects: pedigree-derived relationship matrices, based on gene-drop simulations (denoted as “A”), and “G”, and “H”, respectively. Model variants “Aa”, “Ga” and “Ha” included only an autosomal relationship matrix of one of the specified types. Model variants “Aax”, “Gax” and “Hax”, additionally include the X-chromosomal relationship matrix of the same type. The Restricted maximum likelihood (REML) estimates of all variance components were obtained from the ASReml package (Gilmour et al. 2015). All estimated genetic variance components from model variants, which included founder relationships, were corrected for non-independence.
Table 4 Numbers of observations (N) and number of levels for fixed and random effects in models fitted to litter size (LS0), litter weight (LW0) and body mass traits at ages of 21, 42, 63 days and at mating (BM21, BM42, BM63, BM). Fixed effects are generation number, linear regression on mother’s weight at mating, litter number and sex; random effects are additive genetic, litter and permanent environmental effects

| Trait  | N     | Generation | Weight of Mother | Litter Number | Sex | Additive Genetic Effect | Litter env. | Permanent env. |
|--------|-------|------------|------------------|---------------|-----|-------------------------|-------------|---------------|
| LS0    | 5,911 | 28         | 1                | 2             | 1   | 4,430                  | 2,420       | 1481          |
| LW0    | 5,884 | 28         | 1                | 2             | 1   | 4,410                  | 2,397       | 1474          |
| BM21   | 19,080| 29         | 1                | —             | 2   | 19,080                 | 4,362       |                |
| BM42   | 14,586| 29         | 1                | —             | 2   | 13,959                 | 3,132       | 627           |
| BM63   | 13,910| 29         | 1                | —             | 2   | 13,283                 | 3,044       | 627           |
| BM    | 7,416 | 28         | 1                | —             | 2   | 7,416                  | 2,626       |                |

of founders by multiplication with a correction factor, $D_k$ (Searle 1982. p. 355; Legarra 2016):

$$D_k = \frac{\sum f g_{i,j} (i, i)}{f^2} = \frac{\sum f g_{j,i} (i, j)}{f^2}$$

where $g_{i,j}$ are elements of either the observed founder relationship matrix, $G_k$ (after blending), or a respective diagonal matrix for pedigree-derived relationships, and $f = 8$ is the number of founders.

The significance of X-chromosomal genetic effects was evaluated by comparing the full model for each trait with a reduced model without X-chromosomal genetic effects. Error probabilities were derived via restricted likelihood ratio tests (RLRT), with a single degree of freedom (Wiencierz et al. 2011).

Animal welfare declaration

The animal experiments were performed following national and international guidelines and were approved by the local authorities (Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei, Mecklenburg-Vorpommern, Germany).

Data availability

Marker genotypes of founders, a pedigree file for all AIL animals and a data file with observed phenotypes for the six analyzed traits can be found in the RADAR (research data repository) repository under https://doi.org/10.22000/88. An R-program that sets up the $\bar{A}$ matrix for a two-line cross according to the rules explained above (autosomal and X-chromosomal) is available from the first author. Supplemental material available at Figshare: https://doi.org/10.25387/g3.7110440.

RESULTS AND DISCUSSION

Founder relationship matrices

Marker data: The number of polymorphic autosomal SNPs (single nucleotide polymorphisms) was 140,532 in all founders (see Table 5). The number that segregated in the FL1 line alone was 44,827, while 67,450 segregated within the DUKsi control line. The numbers on the X-chromosome (non-pseudoautosomal, Perry et al. 2001) were 2,009 for all founders, 191 in the FL1 line and 1,055 in the control line. Opposite alleles with frequencies at hundred and zero per cent in the two lines (line-specific alleles) occurred with 38.9% on autosomes as well as the X-chromosome. Polymorphic markers were evenly distributed across the genome (see Figure S1 in supplement file) and the density (number per 1 Mbp) was between 34.0 and 70.4 (see Table S1 in supplement file).

Observed genomic relationships: The observed 8×8 genomic founder relationship matrices are shown in Figure 1 as triangular matrices for autosomes (above) and X-chromosomes (below). Observed autosomal self-relationships were fairly uniform in both the control (between 1.382 and 1.488) and the FL1 line (between 1.405 and 1.452). The expected self-relationships (lower triangle of 4×4 matrix, same panel) were 1.525 and 1.669 for the control and FL1 founders, respectively. The lower observed relationships (approximately 7% in controls and 14% in the FL1 line) indicate an excess of heterozygosity in both lines relative to within line Hardy-Weinberg proportions at observed allele frequencies. These deviations can be explained by the sampling of rare alleles, which are more likely to occur in a heterozygous condition when compared with more frequent alleles. Observed self-relationships are, however, considerably larger than one due to elevated homozygosity relative to the non-inbred F2 individuals that define the base population. Despite some fluctuations, relationships between founders of the same line (expected: 1.193) and different lines (expected: -1.193) barely deviate from the expectations. On the X-chromosome, the observed self-relationships of female control founders (range: 1.347 - 1.520) and the FL1 line (between 1.405 and 1.452). The expected self-relationships (lower triangle of 4×4 matrix, same panel) were 1.525 and 1.669 for the control and FL1 founders, respectively. The lower observed relationships (approximately 7% in controls and 14% in the FL1 line) indicate an excess of heterozygosity in both lines relative to within line Hardy-Weinberg proportions at observed allele frequencies. These deviations can be explained by the sampling of rare alleles, which are more likely to occur in a heterozygous condition when compared with more frequent alleles. Observed self-relationships are, however, considerably larger than one due to elevated homozygosity relative to the non-inbred F2 individuals that define the base population. Despite some fluctuations, relationships between founders of the same line (expected: 1.193) and different lines (expected: -1.193) barely deviate from the expectations. On the X-chromosome, the observed self-relationships of female control founders (range: 1.347 - 1.520) deviate more (about 26%) from the expectation of 1.906, while in contrast, observed self-relationships of male control founders and all observed X-chromosomal relationships between male and female founders agree well with the expectations. On both kinds of chromosomes, large negative relationships between individuals from different lines are predominantly a result of SNPs with line-specific alleles. Their high proportion and the resulting negative between-line relationships reflect the long lasting separation of the two lines and their selection for different goals. Consequently, the expectation of $\omega_{AB} = 4r_{AB} = 1.193$ translates into a proportion of $-2r_{AB} = 2 \cdot 0.29825 \approx 0.6$ of the total genetic variance that can be attributed to segregation variance.

Both the autosomal and X-chromosomal observed genomic relationship matrices are singular, with rank seven. In the case of the autosomal

Table 5 Numbers of SNP-markers which were segregating within all eight founders, and within both founder lines (FL1, control). Line-specific SNPs do not segregate within founders of the same line but have alternative kinds of alleles in each line

| Chromosomal Location | All Founders | FL1 | Control | Line-specific SNP |
|----------------------|-------------|-----|--------|-----------------|
| autosomal            | 140,532     | 44,827 | 67,450 | 54,654          |
| X-chromosomal        | 2,009       | 191  | 1,055  | 869             |
markers, this is caused by the relationship of 1.36 between the third and fourth founders (g3,4), which is close to the self-relationship of the same animals (1.38). This translates into a very close correlation of almost one. The background is that the inbred control line actually consisted of several sublines that can be traced back to the same pair of ancestors. Sublines were generated by branching the main line in different generations and maintained by repeated full-sib mating. Unintentionally, two male founders were sampled from the same subline and the other male founders from two different sublines. In the case were all control founders had been drawn from the same subline, the rank for the observed relationship matrix is expected to be five, as almost no genetic variation is expected within the subline.

As a consequence of this rank deficiency, the observed relationship matrices are not invertible. This was solved by blending them with their expected counterparts (see Theory). Alternatively, one could have averaged the two columns and rows of the highly correlated animals and used this average as a replacement in a 7×7 relationship matrix, thereby assigning a single genetic effect to both founders (e.g., Tuchscherer et al. 2004).

**Evolution of self-relationships over generations:** The mean self-relationships, as derived from different kinds of relationship matrices, develop differently over generations (Figure 2). The classical pedigree-derived matrix A has diagonal elements of one in the base generations and the F1, followed by a jump to 1.25, which indicates inbreeding of F2 animals due to full-sib mating in the F1. From then on there is only a very slight increase of the mean inbreeding coefficient. This pattern is present for autosomal relationships (Figure 2, upper left), as well as for X-chromosomal self-relationships in females (lower left panel). Self-relationships from A, in contrast, show strong fluctuations from high position values in founders to high negative values in the F2 (same panels). Autosomal self-relationships from the A matrix reach an average of larger than one in the F2, which increases only slightly in further generations (upper left panel). The A matrix is scaled in such a way that non-inbred F2 animals would have a self-relationship of one. Therefore, larger values are a sign of a higher expected homozygosity when compared with this reference population. The fluctuations of average generational autosomal self-relationships come to an end from generation F3 onwards (upper left), while they continue, albeit with decreasing amplitudes, for X-chromosomal self-relationships in males (middle left) and females (lower left). The underlying reason is that the genetic equilibrium is reached after two generations for autosomal markers, when initial allele frequencies differ in males and females (Crow and Kimura 1970). This process takes longer for X-chromosomal loci (Li 1976, p. 137). In line with this, the amplitudes for male X-chromosomal self-relationships have the opposite sign to those for females.

The mean X-chromosomal self-relationships of males stabilize at around 0.54, which is somewhat larger than 0.5. The reason is that the actual equilibrium allele frequencies approach \( p_1 = \frac{1}{2} p_{FL1} + \frac{1}{2} p_{control} \) instead of \( p_1 = \frac{1}{2} (p_{FL1} + p_{control}) \), since all founders from the FL1 line were females with two alleles and all founders from the control line had only a single allele at each X-chromosome locus. The X-chromosomal A matrix was, however, computed under the assumption of equal contributions of both founder lines to the F2, which would require equal numbers of male and female founders from both lines.

Values of 0.54 therefore indicate somewhat more X-chromosomal variability, as in a reference population were \( p_1 = \frac{1}{2} (p_{FL1} + p_{control}) \).

Average self-relationships from H and G types of relationship matrices can be seen on the panels to the right of Figure 2. The initial fluctuation patterns already described for A are also present in the average self-relationships of these two matrices. The three curves for the G matrices are similar but not identical to A. In comparison, averages for the H matrices are lower in all cases, which is a result of correcting A to lower observed homozygosity than expected, under the assumptions made for the construction of A.

**Genetic parameters:** The genetic variance components and heritability for six selected traits can be found in the Table 6 and Table S2. The X-chromosomal genetic variance proved to be significant at the 5% level for the three growth traits; BM21, BM42 and BM63, regardless of what type of relationship matrix was used as part of the model.
were taken into account (matrix $\mathbf{G}$). Consequently, heritability for LS0 was 20% when founder relationships were not taken into account (matrix $\mathbf{H}$), vs. 12% when they were not taken into account (Table 6). For LW0 and growth traits, the respective comparison were 34% and 32% vs. 23% (LW0), 23% and 22% vs. 17% (BM21), 41% and 40% vs. 35% (BM42), 47% and 45% vs. 41% (BM63), and 52% and 50% vs. 45% (BMM). For LS0, in particular, the increase in the estimates of the genetic variance using $\mathbf{H}$ matrices was considerably larger estimates for autosomal additive genetic variances. $h^2 = 12\%$. Similarly, Gutiérrez et al. (2006) published heritability values for litter size from 0.099 to 0.101. Gutiérrez et al. (2006) also reported considerably lower heritability, from 0.112 to 0.148 (derived with different models), for litter weight.

Estimated heritability values for body weight traits in mice vary widely in the literature. Falconer (1953) reported the heritability for body weight measured at day 60 as approximately 20% for upward selection and 50% for downward selection. Interestingly, Wilson et al. (1971) found that the realized heritability for body weight at day 60 declined from 0.32 in the first 10 generations to 0.08 between generations 61 and 70, in a selection experiment. Eisen (1978) reported a heritability of 0.44 (or 0.55, depending on the method of estimation) for six week body weight. Heath et al. (1995) reported a heritability of 0.25 before selection and a mean heritability of $0.216 \pm 0.0077$ (from $0.191 \pm 0.016$ to $0.242 \pm 0.014$ for different pairs of lines) for six week body weight.
body weight. In an intercross-population, Kramer et al. (1998) estimated high heritability values of 0.54 ± 0.24 for three week body weight, 0.76 ± 0.04 for six week body weight and 0.81 ± 0.01 for nine week body weight in a cross-fostering experiment. Although they are within the upper ranks, our estimates of approximately 40% at ages from 42 to 63 days may be seen as well within the range of literature values.

Estimated genetic trends: Figure 3 shows estimated genetic trends for LS0 and BM42 from models with different relationship matrices. The upper panels (Figure 3A) show genetic trends from models with autosomal relationships only, while the other panels (Figure 3B) depict those from applying the autosomal and X-chromosomal relationship matrices, simultaneously. In general, trends are very similar by shape and also by level, in accordance with the absence of any serious trend. An exception is the jump at generation F3 for LS0. This reflects the fact that from generation F1 onwards only the descendants of a single family were maintained in order to breed the later generations, which makes the differences between the four founder families manifest. See also Figure S2 for genetic trends in other traits.

Underlying assumptions: In deriving , the expected overall heterozygosity, as described by the scale parameter , is taken as being fully determined by the observed allele frequencies of the genotyped founders and the assumption that both lines contribute equally to the new composite population. The latter can easily be adapted to more than two founder lines, even with unequal contributions, by an alteration of the definition of the average allele frequency . With more than two founder lines, the equilibrium state with foreseen autosomal heterozygosity, , may also be reached later than in the second crossbred generation. This depends on the mating scheme that generates the new composite population and, as with two founder lines, is an asymptotic process for X-chromosome markers. The initial numbers of male and female founders may be different in each line to be crossed later. Autosomal and X-chromosome reference populations should be comparable, especially in the interpretation of genetic parameters. Therefore, it seems reasonable to define as for X-chromosome markers, and hence , as if males and females from all founder lines contribute equally. However, this does not need to be the case, as demonstrated by our mouse example. A further assumption entered into is that founder genotype frequencies meet line-specific Hardy-Weinberg equilibrium. Observed founder genomic self-relationships that deviate from their expected counterparts indicate an excess or lack of heterozygosity relative to this assumption. Actual marker heterozygosity values are then accounted for using the combined relationship matrix .

Matrix reflects average genomic relationships that result from repeatedly sampling alleles at observed line-specific frequencies from founders and their forward transmission to later generations, in accordance with the pedigree. No attempts were made to estimate IBD-based founder relationships (e.g., Powell et al. 2010) within lines. There are typically only a few founders of experimental populations, as in our mouse example, and they may provide information on within-line IBD-relationships only with high sampling errors, unless a larger sample is genotyped. However, we did not treat founders as a sample from their respective lines but their genotypes were treated as a complete inventory of all possible alleles that can be further transmitted to the F2 generation and beyond. Therefore, they fully determine the genetic

---

### Table 6 Estimates of genetic parameters for all traits obtained by three different kinds of relationship matrices

| Trait | Relationship | $\sigma^2_s$ | $\sigma^2_e$ | $\sigma^2_h$ | $\sigma^2_p$ | $\sigma^2_r$ | $h^2$ |
|-------|--------------|-------------|-------------|-------------|-------------|-------------|------|
| LS0   | A            | 1.37 (0.30) | 0.31 (0.16) | 2.06 (0.33) | 7.79 (0.27) | 0.12 (0.03) |
|       | G            | 2.25 (0.42) | 0.30 (0.16) | 1.76 (0.34) | 7.81 (0.27) | 0.19 (0.03) |
|       | H            | 2.44 (0.45) | 0.31 (0.16) | 1.70 (0.34) | 7.81 (0.27) | 0.20 (0.03) |
| LW0   | A            | 5.76 (0.83) | 0.53 (0.33) | 3.86 (0.70) | 14.5 (0.5)  | 0.23 (0.03) |
|       | G            | 8.46 (1.10) | 0.49 (0.33) | 3.15 (0.72) | 14.5 (0.5)  | 0.32 (0.03) |
|       | H            | 9.09 (1.17) | 0.50 (0.33) | 3.02 (0.73) | 14.6 (0.5)  | 0.34 (0.04) |
| BM21  | A            | 0.54 (0.07) | 0.02 (0.01) | 2.15 (0.06) | 0.61 (0.03) | 0.17 (0.02) |
|       | G            | 0.73 (0.09) | 0.02 (0.01) | 2.13 (0.06) | 0.59 (0.03) | 0.22 (0.02) |
|       | H            | 0.77 (0.10) | 0.03 (0.02) | 2.12 (0.06) | 0.59 (0.03) | 0.23 (0.03) |
| BM42  | A            | 2.47 (0.23) | 0.35 (0.11) | 1.71 (0.09) | 2.82 (0.11) | 0.68 (0.04) |
|       | G            | 3.09 (0.29) | 0.30 (0.10) | 1.70 (0.09) | 2.82 (0.11) | 0.68 (0.04) |
|       | H            | 3.28 (0.31) | 0.34 (0.11) | 1.70 (0.09) | 2.82 (0.11) | 0.68 (0.04) |
| BM63  | A            | 3.89 (0.33) | 0.77 (0.20) | 1.19 (0.10) | 4.73 (0.17) | 0.80 (0.05) |
|       | G            | 4.89 (0.42) | 0.64 (0.16) | 1.19 (0.10) | 4.73 (0.17) | 0.80 (0.05) |
|       | H            | 5.17 (0.44) | 0.72 (0.19) | 1.19 (0.10) | 4.73 (0.17) | 0.80 (0.05) |
| BMM   | A            | 5.59 (0.46) | 1.24 (0.16) | 5.74 (0.24) | 0.45 (0.03) |
|       | G            | 7.01 (0.57) | 1.23 (0.16) | 5.75 (0.24) | 0.50 (0.03) |
|       | H            | 7.42 (0.61) | 1.23 (0.16) | 5.75 (0.24) | 0.52 (0.03) |

Traits: litter size (LS0), litter weight (LW0), body weights at ages of 21, 42, 63 days and at mating (BM21, BM42, BM63, BMM). Kinds of relationship matrices: pedigree-derived numerator relationship matrix (A), gene-drop derived (G), combined expected and observed genomic relationships (H); subscripts indicate that autosomal relationships only (a) or both autosomal and X-chromosomal relationships (a+x) were part of the model. Genetic parameters: $\sigma^2_s$: autosomal additive genetic variance, $\sigma^2_e$: common litter environmental variance, $\sigma^2_h$: permanent environmental variance, $\sigma^2_p$: residual variance, $h^2$: heritability, standard errors in brackets.
makeup of the later crossbred population, as mirrored by the construction of $\mathbf{A}$. The usefulness of $\mathbf{A}$, and hence $\mathbf{H}$, for estimating the genetic variance will depend on how well the frequency spectrum of markers reflect the frequency spectrum of QTL (quantitative trait loci) for a trait (Walsh and Lynch 2018, p. 631-668). The latter requirement is likely to be largely fulfilled in a cross of two divergently selected lines, where a large contribution by segregation variance can be expected to be picked up by a large proportion of markers with line-specific alleles.

Fields of application: In the analysis of selection experiments, mixed models have largely replaced other methods due to their flexibility (Walsh and Lynch 2018, p. 631-668). Data from AIL lines can be seen as a special case, with no artificial selection applied. Pedigree-based relationship matrices traditionally treat founders as unrelated and non-inbred. As only a limited number of founders are often genotyped, due to cost, this unrealistic assumption can be overcome by taking genomic founder relationships into account. Using mixed models, with the described $\mathbf{A}$ based version of $\mathbf{H}$, accounts for the initial disequilibrium in allele and genotype frequencies, which exists for more generations on the X chromosome, compared with only two generations for autosomal loci. Meaningful estimates of genetic variances and heritability may be calculated in crosses of two lines in which line-specific alleles can be expected to prevail both at marker loci and QTL due to divergent selection histories. Moreover, in contrast to a gene-drop derived matrix, additional genotypes from later generations can easily be integrated by setting up a joint genomic relationship matrix for all genotyped individuals and joining it with $\mathbf{A}$ into $\mathbf{H}$. Thus $\mathbf{A}$ leads to more realistic assumptions and more flexibility in the analysis of selection experiments if all founders are genotyped.

Conclusion
An approach for constructing expected autosomal and X-chromosomal genomic relationship matrices for founders from an arbitrary number of founder lines was developed. Extension to non-genotyped individuals in later generations can be performed by an adapted version of the tabular method using pedigree information. The resulting matrix $\mathbf{A}$ expresses relationships as average genomic relationships, as one would expect from repeated random sampling of alleles from founders at the observed frequencies. Implicitly, $\mathbf{A}$ accounts for any proportion of segregation variance between zero and one, which is not possible using only pedigree data. Observed marker data of founders and non-founders can then be combined into a joint relationship matrix, $\mathbf{H}$, and its inverse can be used in mixed models for estimating the genetic variance in the crossbred population.

Acknowledgments
Many thanks are to our colleagues from the mouse facility (LIN) in Dummerstorf for their technical assistance. J Meng gratefully acknowledges the financial support by the China Scholarship Council.

Literature Cited
Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta et al., 2010 Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. J. Dairy Sci. 93: 743–752. https://doi.org/10.3168/jds.2009-2730
Alm, H., S. Kuhlmann, M. Langhammer, A. Tuchscherer, H. Torner et al., 2010 Occurrence of Polyovular Follicles in Mouse Lines Selected for High Fecundity. J. Reprod. Dev. 56: 449–453. https://doi.org/10.1262/jrd.09-224H
