G3.6 cflury@gwdg.de

Mean Epistatic Kinship - A New Tool for the Analysis of Short-Term Phylogenetic Structures

C. Flury and H. Simianer
University of Göttingen, Institute of Animal Breeding and Genetics, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany

Abstract

The kinship coefficient (Malécot, 1948) focuses a single locus only. The extension to chromosomal segments leads to a new similarity index called ‚Epistatic Kinship', which basically is the probability that chromosome segments of a given length are identical by descent. This parameter reflects the number of meioses separating individuals or populations. Thus, it might be used as measure to quantify the genetic distance of sub-populations that have been separated only few generations ago, which may prove especially useful in studies with farm or experimental animals. In this study algorithms to calculate Epistatic Kinship and Epistatic Inbreeding are presented. The properties of the approach are investigated in a Monte Carlo simulation. Further a test based on the mean Epistatic Kinship is demonstrated for the assignment of animals to their population of origin. Based on the results the optimum design for microsatellite haplotypes will be derived and the respective genotyping will be done for the three subpopulations of the Goettingen Minipig. The typing results will be used to calculate the Epistatic Kinship, Epistatic Inbreeding and the diversity between populations based on the average Epistatic Kinship. We expect the project to provide a novel approach to the estimation of genetic similarity of closely related individuals or populations.

Introduction

The coefficient of kinship C_{ij} is defined as the probability that two randomly sampled alleles from the same locus in two individuals i and j are identical by descent (*ibd*) (Malécot, 1948). Another concept for the estimation of genetic similarity between individuals is the coefficient of relationship $R_{xy,}$ specified by Wright (Wright, 1922). The link between the two parameters is $R_{ij} = 2 * C_{ij}$.

Coefficients of kinship refer to the *ibd* probability for a randomly chosen single locus or as an average over all loci (Simianer, 1994). This presumes independently segregating loci. For the differentiation of breeds and the genetic control of important traits the formation of genecomplexes over multiple loci and epistatic interactions are important (Brockmann et al., 2000). Therefore an extension of the perspective from single loci to chromosomal segments seems reasonable.

Eding and Meuwissen (2001) proposed mean coefficients of kinship between and within populations as tool to assess genetic diversity in livestock populations. The estimates of kinship were compared with genetic distance measures. Marker based estimates of kinship yielded higher correlations with pedigree-based kinships than genetic distance measures. An extension from this approach to chromosomal segments, called the average Epistatic Kinship as measure of genetic diversity is suggested in this study.

Various studies investigate the properties of conserved haplotypes around a polymorphism. Visscher et al. (1996) describe haplotype sharing in the context of marker based introgression while several other authors use it in context of ibd-mapping of genes and QTL's (for example: Meuwissen and Goddard, 2000, Nezer et al., 2003). The length of conserved haplotypes depends on the timespan since separation.

In this study the coefficient of kinship will be extended from single loci to chromosomal segments of length x Morgan. This leads to a new similarity index called Epistatic Kinship C_{ij}^{x}. The Epistatic Kinship describes the probability of chromosomal segments being identical by descent. Taking into account *ibd*-segments, more accurate statements of kinships or inbreeding and the inference of relationships without pedigree information are possible. Considering populations the average Epistatic Kinship can be used as measurement for genetic diversity.

Methods

An existing FORTRAN-Code was extended for the simulations done in this study. Further we used SAS (Release 8.01) and Maple V (Release 5). Crossing over are assumed to follow a Poisson distribution without interference, thus the Haldane mapping function was applied. If the distances are measured in Morgans, then the rate of the Poisson process is one.

Epistatic Kinship

The extension from single locus to chromosomal segments requires a correction for the probability that recombination occurs. The rules to set up the numerator relationship matrix with a recursive tabular method (Lynch and Walsh, 1998) were extended. Given a list of animals sorted by age in such a way, that parents always precede their offspring, the diagonal element for animal i with parents $s, d < i$ is

$$A_{ii}^{x} = 1 + 0.5 A_{sd}^{x} * (1 - x)^2$$

The relationship of this animal i with all older animals $k < i$ is

$$A_{ki}^{x} = \frac{A_{ks}^{x} + A_{kd}^{x}}{2} * (1 - x)$$

where: $x =$ length of considered chromosomal segment in Morgan

Epistatic Kinship relates to this extended numerator relationship matrix such that

$$C_{ij}^{x} = \frac{1}{2} A_{ij}^{x}$$

C_{ij}^{x} is the probability, that two randomly chosen homologue chromosomes of two individuals i, j are *ibd* for a chromosome segment of length x Morgan.

C_{ii}^{x} is derived from the diagonal element of the numerator relationship matrix.

$$C_{ii}^{x} = \frac{1}{2} A_{ii}^{x}$$

C_{ii}^{x} describes the probability that for animal i pairs of segments of the length x Morgan are *ibd*. The Epistatic Inbreeding F_{i}^{x} equals

$$F_{i}^{x} = A_{ii}^{x} - 1 .$$

Simulation and Test

The efficiency of the suggested apporach in the analysis of short-term phylogenetic structures was analysed in a simulation study.

A base population was simulated consisting of 50 male and 50 female animals. After 7 generations of closed reproduction it was separated in a second population. The two populations consisted each of 50 males and females. 9 Generations after fission the total pedigree contained 2500 animals. The populations were assumed to be ideal, i.e. equal number of male and female animals, random mating and random contributions to the next generation. 10 replications had been implemented resulting with a minimal standard error. The history of the two populations is depicted in figure 1.

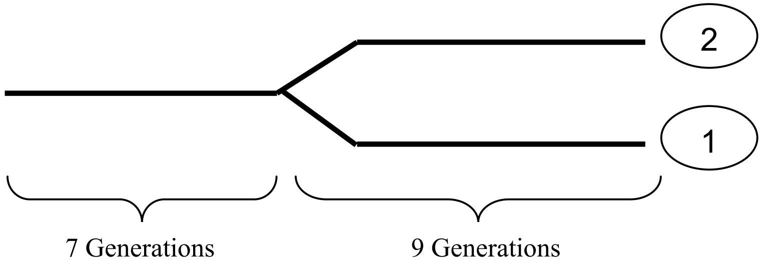
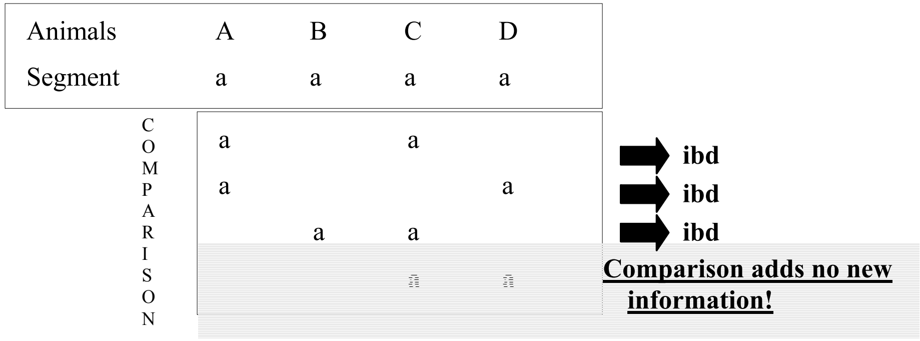


Figure 1: History of the two populations

The epistatic A-matrix was set up to calulate the average Epistatic Kinship within and between the two populations of the last nine generations. This was done for segment length x=0, 0.05, 0.1, 0.15, and 0.2. This leads to 5 different scenarios for the „true state of nature“ for the probability p_w having *ibd* alleles within population and p_b having *ibd* alleles between population for each x.

In generation t after fission, it was tested whether the two populations were genetically identical or different. The test was based on a random sample of M individuals in each of the two populations. For these indiviudals, L chromosome segments of length x were genotyped. For each pair of the 2M individuals the expected *ibd*-status was developed. At this stage, the number of informative comparisons needs to be taken into account. This will be illustrated with the following example. Consider individuals A and B in population 1 and C and D in population 2. We find that for one chromosome segment A is *ibd* with both C and D. B is also found to be *ibd* with C, then C has to be *ibd* with D as well. In this case, only three of the four comparisons are in fact informative.



Animals	A	B	C	D	
Segment	a	a	a	a	
COMPARISON	a		a		ibd
	a			a	ibd
		a	a		ibd
			a	a	Comparison adds no new information!

Figure 2: Effective comparisons between population 1 and population 2 for segment a

For the number of informative comparisons where N_w stands for the effective pairwise comparisons within populations and N_b for the effective pairwise comparisons between populations the following approximations were used.

$$N_w = 2\left[(M-1) + (M^2/2 - 3M/2 + 1)(1-p_w^2)^{(M-2)}\right]$$

$$N_b = M + M(M-1)(1-p_b^3)$$

Note that the proportion of informative comparisons within and between populations is inversely proportional to the *ibd* probabilities p_w and p_b, respectively. Note further that for $p_w = p_b = 0$, $N_w = M(M-1)$ and $N_b = M^2$, respectively. Because in each comparison four different pairs of chromosome segments can be compared, the number of pairwise comparisons within (V_w) and between (V_b) populations are:

$$V_w = N_w * 4L$$

$$V_b = N_b * 4L$$

The proportion of total *ibd* animals and total not-*ibd* animals was taken as expected probability p_0, calculated with the fomula below.

$$p_0 = \frac{p_w * V_w + p_b * V_b}{V_w + V_b}$$

The Nullhypothesis was $H_0 : p_b = p_w = p_0$ accordingly the alternative hypothesis was $H_A : p_b < p_w$. Thus the aim was to test if two samples were taken from different populations or not. Finally the teststatistic $E(X^2)$ was reached following the formula below.

$$E(X^2) = \frac{(p_w - p_0)^2 * V_w + (p_b - p_0)^2 * V_b}{p_0} + \frac{(p_w - p_0)^2 * V_w + (p_b - p_0)^2 * V_b}{(1 - p_0)}$$

Asymptotically, this test statistic follows a χ^2-distribution with 1 df.

Results

In Figure 3 the behaviour of the C_{ij}^x within population 1 and 2 and the average C_{ij}^x between population 1 and 2 is depicted for all generations and $x = 0$. As x was set to 0, i.e only one locus considered the graph below reflects the case where the Epistatic Kinship C_{ij}^x equals the Kinship C_{ij}.

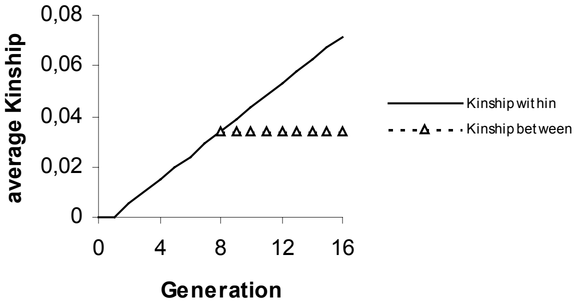


Figure 3: Average Epistatic Kinship within and between populations 1,2 ; $x = 0.0$

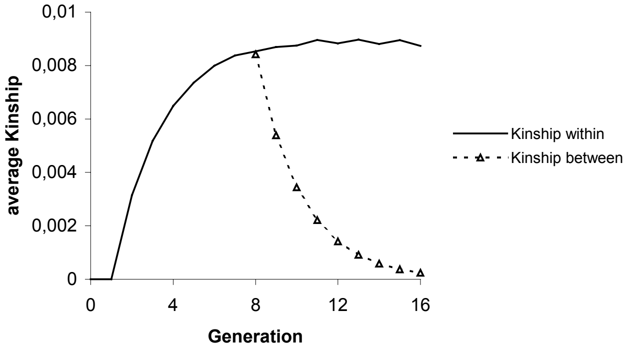


Figure 4: Average Epistatic Kinship within and between populations 1,2 ; $x = 0.2$

Generally, the level of average Epistatic Kinship is decreasing with increasing segment length. Now the Epistatic Kinship within population loses the linear behaviour over generations. After generation 9 the increase of Kinship within population resulting from conancestry is balanced by the loss of *ibd*-status due to recombination. So, the average Kinship within a population remains constant. Since *ibd*-status between populations is based on „old“ coancestries, recombination quickly destroys *ibd*-segments so that the average Kinship between populations is eroding within few generations, thus stays not constant as it did under the one locus approach. Figure 5 shows the log of the proportion of average Epistatic Kinship within and average between populations of the five different x (0.0, 0.05, 0.1, 0.015 and 0.2). The curves are nearly linear. With Epistatic Kinship, the number of generations since fission is directly proportional to the log, where the slope is a function of the length of the chromosome segment x.

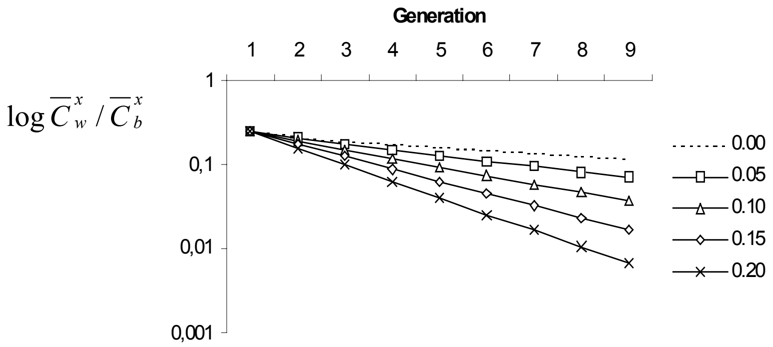


Figure 5: Log of the proportions of Epistatic Kinship within and Epistatic Kinship between for 5 different values of x

The characteristics of the $E(X^2)$ are depicted in figure 6 for $M = 10$, $L = 5$ for the five different x. The curve for $x = 0.0$, i.e. considering one locus only, results in each generation with lower $E(X^2)$ than $x > 0$. Further it can be seen that we have different most informative segment lengths for different generations since fission.

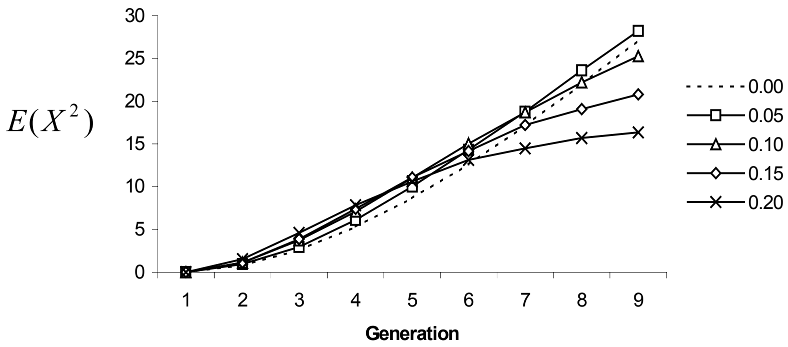


Figure 6: $E(X^2)$ for x = 0.0, 0.05, 0.1, 0.15 and 0.2

Discussion

Methods for the estimation of genetic distances have been developed to study differentiation of species, thus for processes in an evolutionary timespan. For livestock populations the timspan between separation is much shorter. Meiosisequivalents may better account for this fact. Regarding the results, the Epistatic Kinship seems a promising approach for the analysis of short term phylogenetic structures. Figure 6 shows, that the choice of the segment length influences the resolution of the approach. Thus long segments fit better for shorter timespan since separation and short segments fit better for longer timespan since separation.

The variable L used in the test stands for the number of loci or chromosome segments considered. A differentiation for the higher typing effort when working with chromosomal segments was not introduced yet. This has to be taken in account while seeking the economically optimum approach.

Further it can be seen, that the number of Loci or chromosome segments used L has a linear influence on the derivation of expected *ibd*-animals while the influence of the sample size M is squared. Thus the sample size of tested animals has a stronger effect on $E(X^2)$ than the number of considered loci or segments. This means that with a given budget, it is more informative to genotype many animals for few markers, rather than few animals with many markers.

For the simulation fully informative markers were assumed. In reality this assumption does not hold. Therefore the approach has to be tested with real marker data to estimate the information loss.

Assuming not unique alleles in the founder population the probability for having alleles identical by descent is interferred by the proability of having alleles alike in state. Eding and Meuwissen (2001) show, that the influence of the distribution of the founder alleles is considerable when kinships or numbers of alleles are small. This loss of information one can overcome by genotyping short chromosomal segments with highly polymorphic microsatellites. The suggested approach will be used to study the genetic structures of the Goettingen Minipig population. The three subpopulations of the Goettingen minipig will be genotyped. According to the genotyping result the theoretical results will be evaluated and the optimum design for the mean Epistatic Kinship as tool for the analysis of short term phylogenetic structures will be further improved. This approach may be a useful extension to improve the power to study short-term phylogenies we do have in farm animals. This is essential to an efficient management of farm animal genetic diversity.

Acknowledgements

The project is kindly supported by the German Research Foundation (DFG) .

References

Brockmann, G.A., Kratzsch, J., Haley, C.S., Renne, U., Schwerin, M., Karle, S. 2000. Single QTL effects, epistasis, and pleiotropy account for two-thirds of the phenotypic F_2 variance of growth and obesity in DU6i x DBA/2 mice. Genome Res. **10**: 1941-57.

Eding, H., Meuwissen, T.H.E. 2001. Marker based estimates of between and within population kinships for the conservation of genetic diversity, Journal of Animal Breeding and Genetics **118**: 141-159

Lynch, M. and Walsh, B. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

Malécot, G. 1948. Les mathématiques de l'hérédité. Paris, Masson et Cie.

Meuwissen, T.H.E., Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics 155: 421-30.

Nezer, C., Collette, C., Moreau, L., Brouwers, B., Kim, J.J., Giuffra, E., Buys, N., Andersson, L., Georges, M. 2003. Haplotype Sharing Refines Location of an Imprinted Quantitative Trait Locus With Major Effect on Muscle Mass to a 250-kb Chromosome Segment Containing the Porcine *IGF2* Gene. Genetics: **165**: 277-285.

Simianer, H. 1994. Derivation of single locus relationship coefficients conditional on marker information. Theor. Appl. Genet. **88**: 548-556.

Visscher, P.M., Haley, C.S., Thompson, R. 1996. Marker assisted introgression in backcross breeding programs. Genetics **144**: 1923–1932.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. **56**: 330-339.