J. Appl. Genet. 39(4), 1998, pp. 309-319

Properties of the transmission-disequilibrium test for the detection of quantitative trait loci in livestock

Joanna SZYDA[1,2], Zengting LIU[3], Volker WILD[4]

[1] Department of Animal Genetics, Agricultural University of Wrocław, Wrocław, Poland
[2] Department of Epidemiology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada
[3] Biometrical Unit, United Datasystems for Animal Production (VIT), Verden, Germany
[4] Animal Genetics Group, Department of Animal Husbandry and Animal Breeding, University of Hohenheim, Stuttgart, Germany

Abstract. The transmission-disequilibrium test (TDT) is a model-free method to detect linkage between a marker and a trait locus. Originally developed to map disease genes in human genetics, this statistic has been recently extended to deal with quantitative characters. The emphasis of current research is on investigating statistical properties of the test applied to data from livestock populations. For various constellations of sample parameters, it is shown via simulation that the empirically derived null hypothesis distribution of TDT remains in good agreement with its asymptotic distribution while its power is satisfactory only for very close linkage. TDT is then applied to a real data set from milk production data of a dairy cattle population.

Key words: dairy cattle, model-free methods, QTL detection, transmission-disequilibrium test.

Introduction

The transmission-disequilibrium test (TDT; SPIELMAN et al. 1993) was originally designed for binary traits and has been widely used in human genetics to map disease loci. Recently the application of TDT has also been extended to quantitative traits (ALLISON 1997, MARTIN et al. 1997). This statistic belongs to the family of model-free methods, as it does not require assumptions on the underlying (usually unknown) genetic model, e.g., allele frequencies or

Received: September 1998.
Correspondence: J. SZYDA, Department of Animal Genetics, Agricultural University of Wrocław, ul. Kożuchowska 7, 51-631 Wrocław, Poland.

mode of inheritance. Usually such statistics are more robust than model-based tests performed under false assumptions (ELSTON 1995). Additionally, calculations based on a model-free approach are easy and fast to compute, as they do not require estimation of many parameters.

The prior goal of the paper is to investigate statistical properties of TDT for quantitative traits and data structure typical for dairy cattle populations. This is followed by the application of TDT to a sample of cows of the German Holstein breed.

Material and methods

Transmission-disequilibrium test

In its original form TDT compares the number of affected offspring who received a given marker allele from heterozygous parents to the number of affected offspring who received the other parental marker allele. For quantitative trait data, average trait values for offspring receiving different parental marker alleles are compared (ALLISON 1997, MARTIN et al. 1997; Table 1). The test statistic is given by:

$$TDT=\frac{m-1}{m}\sum_{i=1}^{m}\frac{(\bar{y}_{i.}-\bar{y}_{.i})^2}{\left(\frac{1}{n_{i.}}+\frac{1}{n_{.i}}\right)S_i^2}\sim\chi^2_{m-1}$$

where: m is the number of marker alleles, $\bar{y}$ is the average trait value, e.g.,

$$\bar{y}_{i.}=\frac{\sum_{j=1}^{n_i}y_{i.j}}{n_{i.}},$$

n is the number of offspring, subscript "$i.$" denotes that allele i was transmitted from a heterozygous parent to the offspring, subscript "$.i$" denotes that an allele other than i was transmitted from a heterozygous parent to the offspring, S_i^2 is the estimator of pooled variance of $y_{i.}$ and $y_{.i}$:

$$\frac{\left[\sum_{j=1}^{n_{i.}}y_{i.j}^2-\frac{\left(\sum_{j=1}^{n_{i.}}y_{i.j}\right)^2}{n_{i.}}\right]+\left[\sum_{j=1}^{n_{.i}}y_{.ij}^2-\frac{\left(\sum_{j=1}^{n_{.i}}y_{.ij}\right)^2}{n_{.i}}\right]}{n_{i.}+n_{.i}-2}.$$

The transmission-disequilibrium test provides a simultaneous test for linkage and linkage disequilibrium between a marker and a quantitative trait locus (QTL). The null hypothesis H_0: $\delta(1-2\theta) = 0$, is rejected only when both linkage ($\theta < 0.5$) and linkage disequilibrium ($\delta \neq 0$) exist in the analysed population (ELSTON 1995). Similar hypotheses can also be tested by other commonly known statistics. Using a *t*-test – as a comparison between average yields of daughters receiving different marker alleles from their sires or an *F*-test – as an across sire regression fitting the effect of paternally transmitted marker alleles.

Simulated data

In order to investigate statistical properties of TDT, data under a daughter design – structure which is most common for dairy cattle populations, were generated. Simulated data comprised marker allele transmissions from sires to daughters and daughters quantitative yields. For each constellation of parameters, describing population structure and traits genetic background, 2000 replications of data were generated. Through all simulations the type I error was set to 5%.

Phenotypes were simulated according to the following linear model:

$$y = X\beta + e$$

where: **y** is a vector of quantitative yields, β is a vector of fixed effects, X is a design matrix for β, e is a vector of random errors.

Under the null hypothesis (H_0) no QTL was assumed so that: $\beta' = [\mu, s]$, where μ is the overall mean, and s is the sire effect. The alternative (H_1) models a quantitative trait controlled by a biallelic QTL assuming: $\beta' = [\mu, s, G]$ where μ, s are as above, and G is the effect of a QTL genotype.

For simplicity, sire effects (s) were set to zero in simulations under both hypotheses. Following this model, quantitative phenotypes were simulated from the univariate normal distribution under H_0 and from the mixture of normal distributions, with mixing proportions determined by QTL allele frequencies (p, 1 – p) and means dependent on QTL genotypes:

$p^2\ N(\mu_{QQ} = a, \sigma_e^2)$, $2p(1-p)\ N(\mu_{Qq} = d, \sigma_e^2)$, $(1-p)^2\ N(\mu_{qq} = -a, \sigma_e^2)$ under H_1.

Symbols a and d are, respectively, additive and dominant effects at the QTL, QQ, Qq and qq represent the three possible genotypes.

The design of simulated data covered various sample structures (i.e., number of sires ranging from 10 to 100, and number of daughters per sire from 10 to 100), as well as different variants of genetic composition of a quantitative trait, expressed in terms of a recombination rate (θ) and linkage disequilibrium (δ) between a marker and a QTL. Values of recombination rate ranged from 0.05

to 0.30, linkage disequilibrium was equal to 0.1 or 0.2. Under the alternative, an additive genetic effect of a QTL, expressed in phenotypic standard deviation units, was simulated as 0.1, 0.3 or 0.5, while no dominance ($d = 0$) was assumed, i.e. $\mu_{Qq} = 0$. The number of alleles simulated at marker loci was 2, 5 or 10.

Real data

The real data set available for TDT analysis was collected from the German Holstein population. It consists of 671 sires and their 3059 daughters. For each daughter genotypes at three milk protein markers and milk production data were available. Chromosome 6 is represented by two very closely linked markers – κ-casein (3 alleles) and β-casein (4 alleles). The third marker – β-lactoglobulin (2 alleles), maps to chromosome 11. Production data comprise 305-day milk, fat, and protein yields, together with fat and protein contents. All traits were preadjusted by subtracting the GLS (Generalised Least Squares) estimates of herd-year-season of calving effects from individual yields (HENDERSON 1984).

The data were not collected especially for TDT, but for the purpose of major gene detection using likelihood based approach (WILD 1997). For this reason, the data had to be adapted accordingly prior to the analysis. First, sires' marker genotypes – originally unknown, were estimated using marker information from daughters. Only heterozygous sires with marker genotypes assigned with probability at least 0.975 were chosen. Secondly, dams' marker genotypes were also unknown, so that only transmission from sires to daughters could be considered and those sire-daughter pairs for which the transmission probability of a given allele was 0 or 1 were included in the analysis. Finally, trait values were additionally corrected for the sire effect, so that TDT statistic remains a valid test also in case of an apparent (due to a sampling process) or a true distortion of marker allele segregation from sires to offspring (see Appendix 1). Thus, the final data set available for model-free analysis with TDT is much smaller than the original sample and consists of 17 heterozygous sires with an average of 44 daughters per sire for κ-casein, 24 heterozygous sires with an average of 43 daughters per sire for β-casein, and 18 sires with an average of 32 daughters per sire for β-lactoglobulin.

Results

H_0 distribution

The empirical H_0 distribution of TDT was constructed for different numbers of marker alleles, sires and daughters per sire. As shown in Figure 1, under

Table 1. Partitioning of the sample information for TDT

Parental marker allele		Not transmitted to offspring			
		1	2	m	sum
Transmitted to offspring	1		$n_{12}\,\bar{y}_{12}$	$n_{1m}\,\bar{y}_{1m}$	$n_{1.}\,\bar{y}_{1.}$
	2	$n_{21}\,\bar{y}_{21}$		$n_{2m}\,\bar{y}_{2m}$	$n_{2.}\,\bar{y}_{2.}$
	m	$n_{m1}\,\bar{y}_{m1}$	$n_{m2}\,\bar{y}_{m2}$		$n_{m.}\,\bar{y}_{m.}$
	sum	$n_{.1}\,\bar{y}_{.1}$	$n_{.2}\,\bar{y}_{.2}$	$n_{.m}\,\bar{y}_{.m}$	

n – number of offspring, $\bar{y}$ – average trait value

all considered constellations of parameters, the test statistic remains in good agreement with the χ^2 distribution, with only a slight tendency of conservativeness for a large number of daughters per sire and a large number of halfsib families.

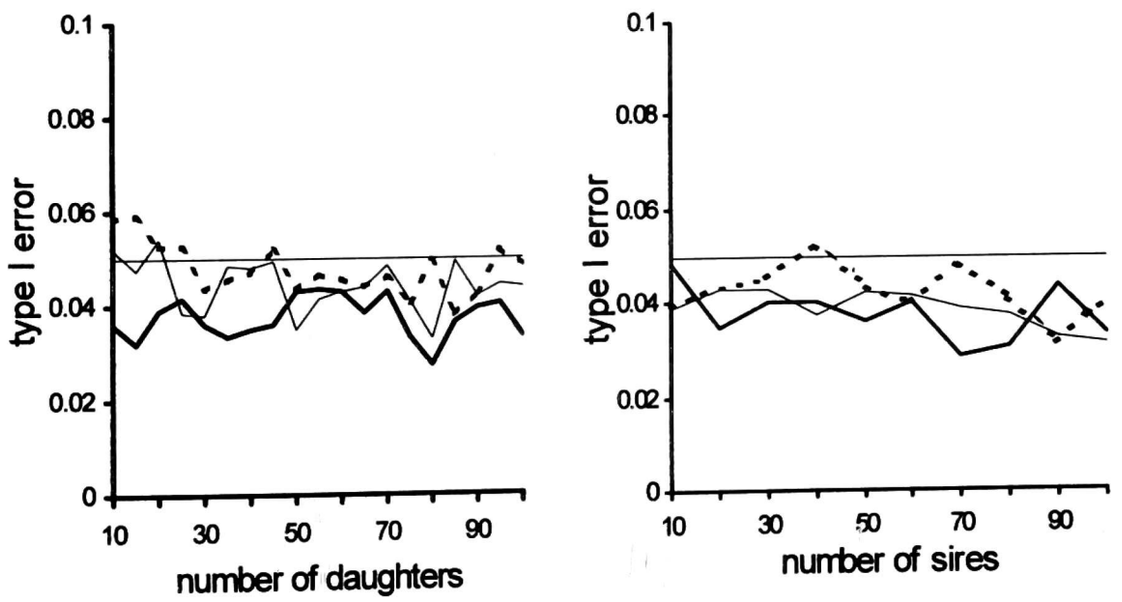

Figure 1. Empirical H_0 distribution of TDT based on simulations for 2 (—), 5 (—), and 10 (---) alleles at a marker locus, and $\alpha = 0.05$

left graph – 20 sires, right graph – 50 daughters per sire, α – type I error based on the χ^2 distribution

Power

Recombination rate and linkage disequilibrium between a QTL and a marker locus are the quantities determining the hypotheses tested by TDT. The rate of recombination operates on a halfsib family level determining the proportion of daughters which obtain a favourable QTL allele together with a given marker allele from a (heterozygous) sire. Linkage disequilibrium plays a key role across halfsib families, determining the proportion of sires with the same

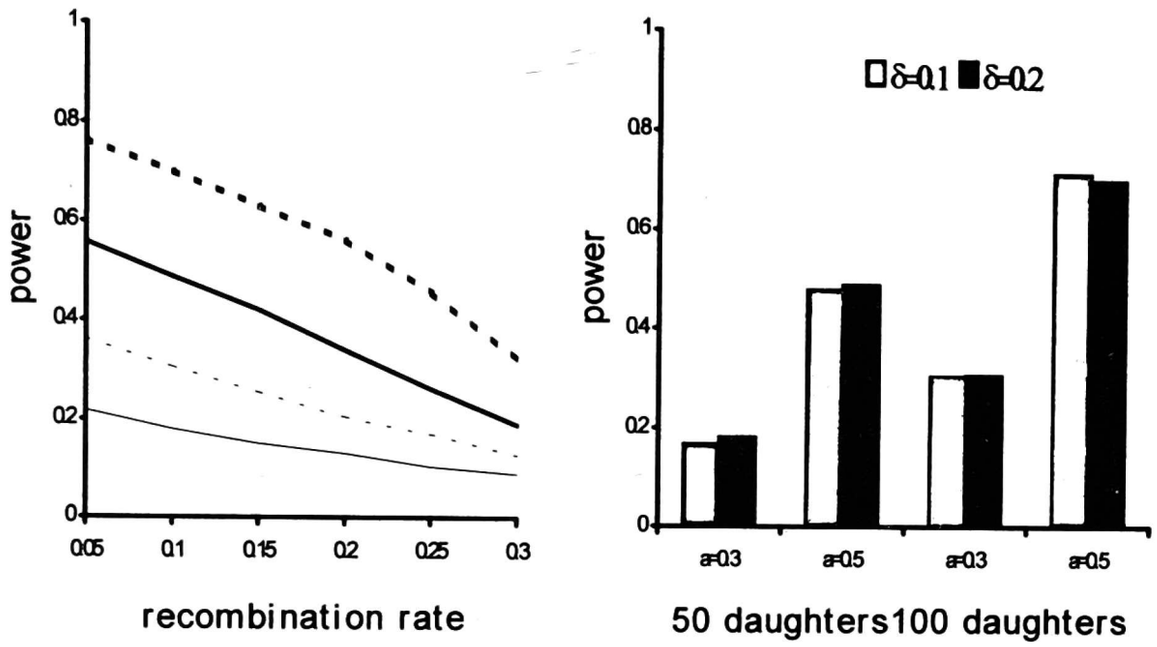

Figure 2. Power of TDT for 20 sires, 5 alleles at a marker locus, and $\alpha = 0.05$
left graph – $\delta = 0.2$ (—) 50 daughters per sire, a = 0.3; (----) 100 daughters per sire, a = 0.3; (—) 50 daughters per sire, a = 0.5; (---) 100 daughters per sire, a = 0.5
right graph – $\theta = 0.1$ a – additive genetic effect of a QTL, θ – recombination rate, δ – linkage disequilibrium, α – type I error based on the χ^2 distribution

linkage phase at quantitative trait and marker loci. Their impact on power is summarised in Figure 2. It shows that power is strongly dependent on recombination rate. Tight linkage is a prerequisite for TDT, so that markers having more than 20% of recombination with a QTL do not provide satisfactory power,

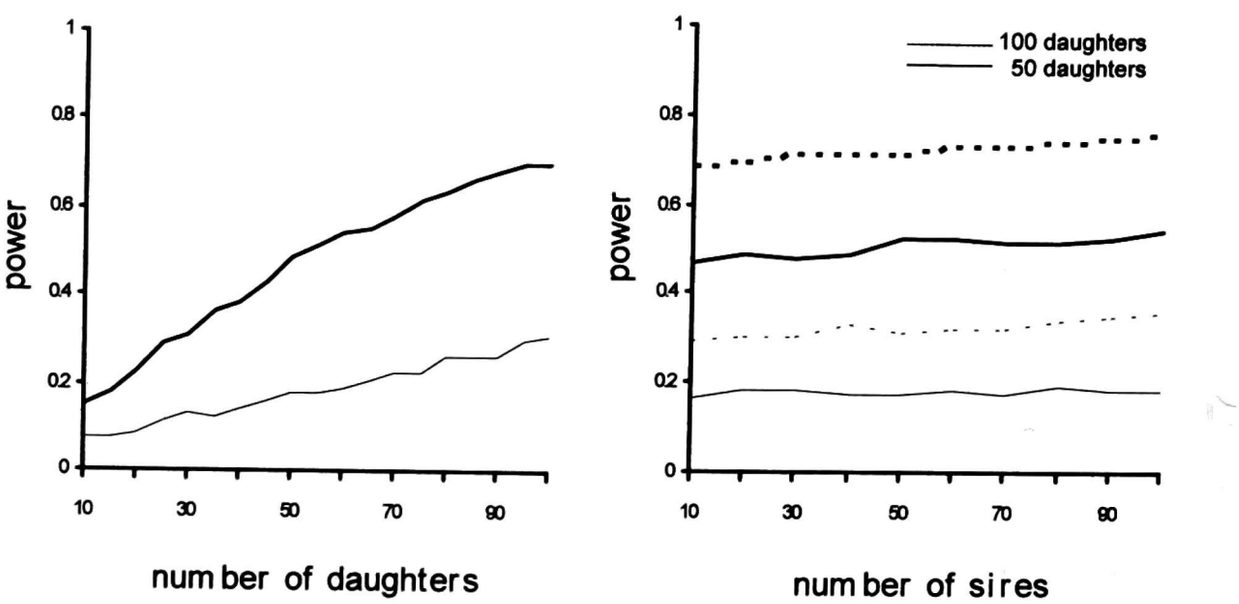

Figure 3. Power of TDT for a = 0.3 (—), and a = 0.5 (—), $\theta = 0.1$, $\delta = 0.2$, 5 alleles at a marker locus, $\alpha = 0.05$
left graph – 20 sires, right graph – 50, and 100 daughters per sire, a – additive genetic effect of a QTL, θ – recombination rate, δ – linkage disequilibrium, α – type I error based on the χ^2 distribution

even in large halfsib families. On the contrary, the amount of linkage disequilibrium has only negligible effect on power, so that increasing linkage disequilibrium from 0.1 to 0.2 resulted in less than 1% gain in power for most of the experimental designs considered. It should be kept in mind that on a population level, recombination rate and linkage disequilibrium are confounded so that for an ideal population (i.e., undergoing random mating, no mutation and no migration) linkage disequilibrium in generation *t* is stronger for closely linked than for loosely linked loci (see Appendix 2).

Table 2. Power of TDT for 20 sires, 100 daughters per sire, $\delta = 0.2$, 5 alleles at a marker locus; $\alpha = 0.05$

QTL additive effect[a]	Recombination rate	
	0.05	0.1
0.1	0.069	0.064
0.3	0.362	0.305
0.5	0.763	0.699

a – expressed in phenotypic standard deviation units, δ – linkage disequilibrium, α – type I error based on the χ^2 distribution

Another factor influencing the power of hypothesis testing is sample structure. Power of TDT is strongly dependent on the number of daughters per sire, whereas increasing the number of sires improves testing performance only marginally (Figure 3). To map a QTL with a large effect (a = 0.5), power of 60% requires over 55 daughters per sire when a marker locus has 5% recombination with a QTL, and over 70 when recombination increases to 10%.

For halfsib family sizes commonly available in dairy cattle, TDT enables mapping of genes with large effects on the variation of a quantitative trait, explaining over 30% of the total variance. For up to 100 daughters per sire, the power of the detection of genes with relatively small effects is low, even if a gene is very closely linked to the available marker (Table 2).

Table 3 . Empirical P-values of TDT for milk protein marker loci and yield traits

Trait	Marker locus		
	κ-casein	β-casein	β-lactoglobulin
Milk yield	0.897	0.804	0.508
Fat yield	0.996	0.910	0.186
Fat (%)	0.733	0.868	0.019
Protein yield	0.950	0.597	0.195
Protein (%)	0.807	0.162	0.164

Application to real data

Results from the real data set of German Holstein cattle are given in Table 3 in form of empirical P-values obtained by permutation of individual yields among daughters (DOERGE, CHURCHILL 1996). For casein loci no evidence of a cosegregating QTL was found, while at the β-lactoglobulin marker significant transmission disequilibrium for fat content (P = 0.019) was identified.

Discussion

Results of simulations show that regardless of the experimental design, the null hypothesis distribution of the test stays in good agreement with the χ^2 distribution. Under the alternative hypothesis, tight linkage is a prerequisite for identifying a QTL. Power can also be enhanced by increasing the number of daughters per family, while increasing the number of families (e.g. sires) does not affect power. Unless there is complete linkage disequilibrium between a marker and a QTL, a certain percent of families from the analysed population contribute data from the opposite linkage phase, so that increasing the number of families also increases the amount of data under this phase.

The application of TDT to the real data set from the German Holstein cattle does not give evidence of a QTL cosegregating with casein markers. This remains in agreement with results of WILD (1997), obtained from the same data set which provided a sample for TDT. Recombination rates estimated in this study are about 30% and estimates of additive QTL effects for different milk production traits range between 0.057 and 0.250. These estimates are given in form of means from marginal posterior distribution of parameters, obtained using the Gibbs Sampling algorithm. Also LIU (1994) found no evidence for a QTL in the neighbourhood of casein loci in another sample from German Holstein cattle population. However, in Dutch and U.S. Holstein families, a QTL responsible for milk production was identified around the proximal part of chromosome 6 (BOVENHUIS, WELLER 1994, SPELMAN 1996, ZHANG et al. 1998). In the current study a significant result indicates possible location of a QTL responsible for fat content near or at the β-lactoglobulin locus on chromosome 11. Significant direct effects of this locus were also reported by (BOVENHUIS, WELLER 1994, LIU 1994, WILD 1997).

Conclusions

TDT does not require intensive computations, is not time consuming and can be obtained using commonly available statistical software. Another advantage of TDT is that, regardless of the experimental design, the null hypothesis distribution of the test stays in good agreement with the χ^2 distribution, so there is no need for the calculation of empirical significance levels. The main disadvantage of the test is its high requirements to reach satisfactory power, i.e., large halfsib families, tight linkage and a QTL with large effect. Providing a satisfactory number of offspring per family does not seem to be a major problem for animal breeding applications, but the prerequisite of tight linkage may often be difficult to obtain, as marker maps of livestock species do not yet provide high density markers all over the genome.

As a result of all these features, TDT is not recommended to search for putative quantitative trait loci with unknown chromosome localisation. However, this statistic can be very useful for fine positioning of already detected genes, as it is commonly applied in human genetics (see, e.g., CEROSALETTI et al. 1998, or JENISCH et al. 1998). In the future, it can also prove its usefulness in monitoring linkage disequilibrium in populations subjected to marker assisted selection.

Acknowledgements. Analysed data set was collected with the financial support of the German Research Foundation. Computer simulations were partially carried out at the Institute of Biophysics at the Agricultural University of Wrocław. The editor and reviewers are thanked for helpful comments.

REFERENCES

ALLISON D.B. (1997). Transmission-disequilibrium tests for quantitative traits. Am. J. Hum. Genet. 60: 676-690.

BOVENHUIS H., WELLER J.I. (1994). Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Genetics 137: 267-280.

CEROSALETTI K.M., LANGE E., STRINGHAM H.M., WEEMAES C.M.R., SMEETS D., SÖLDER B., BELOHRADSKY B.H., TAYLOR A.M.R., KARNES P., ELLIOTT A., KOMATSU K., GATTI R.A., BOEHNKE M., CONCANNON P. (1998). Fine localization of the Nijmegen Breakage Syndrome gene to 8q21: evidence for a common founder haplotype. Am. J. Hum. Genet. 63: 125-134.

DOERGE R., CHURCHILL G.A. (1996). Permutation test for multiple loci affecting a quantitative character. Genetics 142: 285-294.

ELSTON R.C. (1995). Linkage and association to genetic markers. Exp. Clin. Immunogenet. 12: 129-140.

HENDERSON C.R. (1984). Application of linear models in animal breeding. University of Guelph Press, Guelph, Canada.

JENISCH S., HENSELER T., NAIR R.P., GUO S.W., WESTPHAL E., STUART P., KRÖNKE M., VOORHEES J.J., CHRISTOPHERS E., ELDER J.T. (1998). Linkage analysis of Human Leukocyte Antigen (HLA) markers in Familial Psoriasis: strong disequilibrium effects provide evidence for a major determinant in the HLA-B/-C region. Am. J. Hum. Genet. 63: 191-199.

LIU Z. (1994). Marker assisted complex segregation analysis of milk production traits in dairy cattle. PhD Thesis, Hohenheim.

MARTIN B.R., WEIR B.S., KAPLAN N.L. (1997). Transmission/disequilibrium tests for identifying quantitative trait loci. Am. J. Hum. Genet. 60, Sup 1659, A294.

SPELMAN R.J., COPPIETERS W., KARIM L., van ARENDONK J.A., BOVENHUIS H. (1996). Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics 144: 1799-1808.

SPIELMAN R.S., MCGINNIS R.E., EWENS W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent Diabetes Mellitus (IDDM). Am. J. Hum. Genet. 52: 506-516.

WILD V. (1997). Marker assisted complex segregation analysis of milk production traits in dairy cattle with Gibbs Sampling. PhD Thesis, Hohenheim.

ZHANG Q., BOICHARD D., HOESCHELE I., ERNST C., EGGEN A., MURKVE B., PFISTER-GENSKOW M., WITTE L.A., GRIGNOLA F.E., UIMARI P., THALLER G., BISHOP M.D. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149: 1959-1973.

Appendix 1

Let the statistical model for a quantitative yield of j-th daughter of i-th sire, under the null hypothesis of no QTL, be: $y_{ij} = \mu + s_i + e_{ij}$. Following this model and assuming two alleles at a marker locus, the numerator of the TDT $\bar{y}_{12} - \bar{y}_{21}$ can be expressed as:

$$\frac{\sum_{i=1}^{ns}\sum_{j=1}^{n_{12i}}(\mu + s_i + e_{ij})}{n_{12.}} - \frac{\sum_{i=1}^{ns}\sum_{j'=1}^{n_{21i}}(\mu + s_i + e_{ij'})}{n_{21.}} = \frac{n_{12.}\,n_{21.}\mu - n_{12.}\,n_{21.}\mu}{n_{12.}\,n_{21.}} +$$

$$+\frac{n_{21.}\sum_{i=1}^{ns} n_{12i}\,s_i - n_{12.}\sum_{i=1}^{ns} n_{21i}\,s_i}{n_{12.}\,n_{21.}} + \frac{\sum_{i=1}^{ns}\sum_{j=1}^{n_{12i}} e_{ij} - \sum_{i=1}^{ns}\sum_{j'=1}^{n_{21i}} e_{ij'}}{n_{12.}\,n_{21.}}$$

where, ns is the number of sires,

$n_{12.}$ and $n_{21.}$ are $\sum_{i=1}^{ns} n_{12i}$ and $\sum_{i=1}^{ns} n_{21i}$, respectively.

Three components of the numerator are:

$$\frac{n_{12.}\, n_{21.}\mu - n_{12.}\, n_{21.}\mu}{n_{12.}\, n_{21.}} = 0,$$

$$\frac{\sum_{i=1}^{ns}\sum_{j=1}^{n_{12i}} e_{ij} - \sum_{i=1}^{ns}\sum_{j'=1}^{n_{21i}} e_{ij'}}{n_{12.}\, n_{21.}} \approx 0,$$

$$\frac{n_{21.}\sum_{i=1}^{ns} n_{12i}\, s_i - n_{12.}\sum_{i=1}^{ns} n_{21i}\, s_i}{n_{12.} n_{21.}} = 0 \Leftrightarrow (\forall\, i \in \{1, ..., ns\} n_{12i} = n_{21i})$$

$\vee\, (\forall\, i,i' \in \{1, ..., ns\} s_i = s_{i'})$.

Thus, the preadjustment of sire effects guarantees that TDT remains a valid test also if a segregation distortion of marker alleles from sires to their offspring is observed in a sample (i.e. $n_{12i} \neq n_{21i}$).

Appendix 2

For two loci A and B, let linkage disequilibrium (δ) be defined as: $\delta = P(A_1B_1) - P(A_1)P(B_1)$, where $P(A_1\, B_1)$ is the frequency of haplotype 1–1, $P(A_1)$ is the frequency of allele 1 at locus A, $P(B_1)$ is the frequency of allele 1 at locus B, and θ is a recombination rate between loci A and B.

i) Linkage disequilibrium in generation 1 is: $\delta_1 = P_1(A_1B_1) - P(A_1)P(B_1)$, and the frequency of a coupling phase haplotype 1–1 at generation 1 is: $P_1(A_1B_1) = \frac{1-\theta}{2}$.

ii) Linkage disequilibrium in generation t is: $\delta_t = P_t(A_1B_1) - P(A_1)P(B_1)$, and the frequency of a coupling phase haplotype 1–1 at generation t is: $P_t(A_1B_1) = \frac{(1-\theta)^t}{2}$.

Thus, for an ideal population, the decrease in linkage disequilibrium throughout t generations ($\Delta\delta^t$) depends only on recombination rate (θ) and is equal to:

$$\Delta\delta^t = \delta_t - \delta_1 = P_t(A_1B_1) - P(A_1)P(B_1) - P_1(A_1B_1) + P(A_1)P(B_1) = (1-\theta)^t - (1-\theta).$$