HETEROGENEITY AMONG BULLS IN RECOMBINATION FRACTION USING THE GRANDDAUGHTER DESIGN

A. Sabry [1] and L. Gomez-Raya[2]

1- Cell Biology Department, National Research Center, El Tahrir St. 12622, Dokki, Giza , Egypt, 2- Department of Animal Biotechnology, University of Nevada, Reno, Mail Stop 202, Reno, NV, 89557, USA

SUMMARY

Testing heterogeneity of the recombination fraction with the same genotyping data of the granddaughter design is proposed in this paper. Statistical power was computed using a Montecarlo simulation with the breeding structure of Norwegian cattle. A power of 70.8 was obtained when allele frequencies at the flanking marker alleles were 0.1 and the true recombination fraction was 0.20 with a standard deviation of 0.15. Statistical power of detection of heterogeneity decreased markedly with decreasing average recombination fraction among bulls. A genome scam was carried out to test heterogeneity of the recombination fraction using 194 intervals in Norwegian cattle. A total of 12 and 9 intervals were significant at significance levels of 0.05 and 0.01, respectively. These results support the premise that there is heterogeneity of the recombination fraction among Norwegian bulls.

Keywords: linkage analysis, recombination fraction heterogeneity, granddaughter design, dairy cattle

INTRODUCTION

The granddaughter design (GDD) has been extensively used to map QTL in dairy cattle (Georges *et al*., 1995, Olsen *et al.,* 2002, for review see Khatkar *et al.,* 2004). This design is based on following the recombination events from sires to their half-sib sons by genotyping sires and their respective half-sib sons and by recording phenotypes or production in the granddaughters. A test is performed of the association of marker alleles inherited from the bull sire to the sons (and, therefore, to granddaughters) with production traits. One of the assumptions of the granddaughter design is that recombination fraction has a fixed value in all sires. Evidence of variability in recombination fraction has been reported in several species Park *et al.* (1995), Simianer *et al.* (1997), Szyda (1997), Thomsen *et al*. (2001) in cattle, Lien *et al.* (1999) in sheep, Lien *et al.* (2000) in humans. Disregarding the assumption of heterogeneity will narrow the standard error and the confidence interval of the recombination fraction (Yu *et al.,* 1996). In addition, the sample size required to fine map quantitative trait loci (QTL) would also be affected (Lien *et al.,* 1999).

One avenue, not explored yet, is to use the genotyping data of the grand- daughter design in order to test for heterogeneity of the recombination fraction among bulls used in the granddaughter design. However, the dams (mother of the sons) in the

granddaughter design are not genotyped, which can lead to biased estimation of the recombination fraction (Gomez-Raya, 2001). In addition, it is not known if the breeding structure of the grand- daughter design is enough for powerful testing of the variability of the re- combination fraction. The objectives of this work were: 1) to estimate statistical power by Monte Carlo simulation of the granddaughter population structures when used to test heterogeneity of the recombination fraction, and 2) to carry out a genome scan to test for heterogeneity of the recombination fraction in the Norwegian cattle population using a granddaughter design.

MARIAL AND METHODS

Testing heterogeneity of the recombination fraction in the granddaughter design: The maximum likelihood method to estimate the recombination fraction (Gomez-Raya, 2001) will be used for each sire family. This method accounts for the dam contributions and provide asymptotically unbiased estimates of the recombination fraction for large sample sizes (Gomez-Raya, 2001). A general maximum likelihood expression for the estimation of the recombination fraction several grandsire families (k) is

$$L(\theta) = \prod_{i=1}^{k} L_i(\theta) \qquad (1)$$

Where the maximum likelihood estimate of recombination fraction for the ith sire family is given by

$$L_i(\theta) \propto K(\Phi_{AB})^{nAB}(\Phi_{Ab})^{nAb}(\Phi_{aB})^{naB}(\Phi_{ab})^{nab}(\Phi_{Ax})^{nAx}(\Phi_{ax})^{nax}(\Phi_{xB})^{nxB}(\Phi_{xb})^{nxb}(\Phi_{xx})^{nxx}$$

Where K is constant, Φ_i is the gamete probabilities for gamete i
(i = AB, Ab, aB, ab, Ax, ax, xB, xb, xx) and n_i is the number of individuals that inherited the respective gametes. In these notations, n_{Ax} is number of individuals which are informative for marker A and noninformative for marker B, n_{xx} is the number of individuals which are noninformative for both markers.

$\Phi_{AB} = 0.5(1 - \theta)(1 - f_b)(1 - f_a)$,
$\Phi_{Ab} = 0.5(\theta)(1 - f_B)(1 - f_a)$,
$\Phi_{aB} = 0.5(\theta)(1 - f_b)(1 - f_A)$,
$\Phi_{ab} = 0.5(1 - \theta)(1 - f_B)(1 - f_A)$,
$\Phi_{Ax} = 0.5(1 - \theta)(f_A f_b + f_{a'} f_b) + 0.5(\theta)(f_A f_B + f_{a'} f_B)$,
$\Phi_{ax} = 0.5(1 - \theta)(f_a f_B + f_{a'} f_B) + 0.5(\theta)(f_a f_b + f_{a'} f_b)$,
$\Phi_{xB} = 0.5(1 - \theta)(f_a f_B + f_a f_{b'}) + 0.5(\theta)(f_A f_B + f_A f_{b'})$,
$\Phi_{xb} = 0.5(1 - \theta)(f_A f_b + f_A f_{b'}) + 0.5(\theta)(f_a f_b + f_a f_{b'})$,
$\Phi_{xx} = 0.5(1 - \theta)(f_a f_b + f_A f_B) + 0.5(\theta)(f_a f_B + f_A f_b)$

Where θ is the value of the recombination fraction, f is the allele frequency of the respective allele with a' is any different allele from A or a and b' is any other allele different from B or b.

Testing for heterogeneity of the recombination fraction can be carried out using the above method since estimation of the recombination fraction for each individual sire family is asymptotically unbiased. The Morton test (Morton, 1956) can be used

to test for variability of the recombination fraction. This test depends on comparing the lod score of the recombination fraction estimate from each of the families with the joint lod score for all k families combined, namely

$$X = 2\ln(10)\left[\sum_{i=1}^{k} Z_i(\widehat{\theta}_i) - Z(\widehat{\theta})\right] \qquad (2)$$

Where $Z_i(\widehat{\theta}_i)$ is the maximum lod score that occur at the recombination fraction of $(\widehat{\theta}_i)$ for ith family and $Z(\widehat{\theta})$ the maximum lod score at the joint likelihood of all families.   The statistic follows a χ^2 distribution with k-1 degrees of freedom, asymptotically (i.e. with a large family size)

Montecarlo simulation to estimate empirical power of detection of heterogeneity:
The breeding structure of Norwegian cattle was used as the simulating population. The simulation parameters are shown in Table (1). This population consisted of six parental half-sib families with 50 sons per family (average number of sons that used in construction of the linkage map of the Norwegian cattle Våge *et al*. (2000). Each sire was assumed to be heterozygous for two codominant markers. Sire's linkage phase and allele frequencies in dams were assumed to be known without error. Maximum likelihood estimation (MLE) of the recombination fraction was done according to equation (1).

Table 1. Summary of simulation parameters

No. Grandsires	6			
No. Sons/Family	50			
True Recombination Fraction and their standard deviations	0.05(0.025), 0.1(0.05), 0.2(0.1) and 0.2(0.15)			
No. Replicates	10000			
Assumed linkage phase	AB/ab			
	Allele frequency in the dam population			
	Marker A		Marker B	
	A	a	B	b
Allele	0.1	0.8	0.8	0.1
Frequencies	0.8	0.1	0.1	0.8
	0.8	0.1	0.1	0.1
	0.1	0.1	0.1	0.1

The following input parameters were (1) the simulated (true) value of the recombination fraction (θ) (standard deviation) were 0.05(0.025), 0.1(0.05), 0.2(0.1) and 0.2 (0.15).

The generation of data was based on pseudo random numbers, which are uniformly distributed in the interval (0, 1). Simulation was carried out in two steps. Firstly, the flow of sire's recombinant or non-recombinant gametes to sons depends on the simulated (true) value of the recombination fraction. For example, if θ= 0.05

this value is also the probability with which a recombinant gamete is randomly drawn from the sire.

(2) Allele frequency in the dam population: The assumed allelic frequencies are shown in Table 1.

Allele frequencies in the dam population are the probabilities by which a gamete is drawn at random from the dam population. Consider a codominant marker at a heterozygous sire (Aa), whereas alleles segregating in the dam population are A, a, and a' ,where a' is any other allele segregating in the dam population and different from the alleles of the sire. If a random number drawn from the uniform distribution was lower than f_A (frequency of the A allele in the dam population) then the dam transmitted allele A to her offspring. If the drawn number was between f_A and fa (with fa being the frequency of allele a) then the dam transmitted allele a to her offspring. If the value of the random number was higher than $f_A + f_a$ then the dam contributed with any other allele different from A and a to her offspring. The above mentioned two steps were repeated 50 times for each of the six families, and each simulation set was replicated 10000 times.

The variance among replicates (Var), was also estimated according to the following formula:

$$Var = \sum_{i=1}^{N} (\widehat{\theta}_i - \overline{\theta})^2 \ /(N-1) \qquad (3)$$

Where $\widehat{\theta}_i$ is the estimated recombination fraction in the ith replicate, $\overline{\theta}$ is the mean of the recombination fractions among N = 10000 replicates.

The mean square error MSE was estimated by:

$$MSE = \sum_{i=1}^{N} (\widehat{\theta}_i - \theta)^2 / N \qquad (4)$$

The bias was estimated by:

$$Bias = \sum_{i=1}^{N} (\widehat{\theta}_i - \theta)/ N \qquad (5)$$

Where θ is the true (simulated) recombination fraction.

Variability of the recombination fraction among sire families: Under the null hypothesis (homogeneity) the recombination fraction (θ) is identical in all sire families. (H₀: $\theta_1 = \theta_2 = \ldots = \theta_k$), where k is the number of families. Under the alternative hypothesis the recombination fraction is different in different families H₁: ($\theta_i \neq \theta_j$), for at least one pair, i≠ j.

The recombination fraction was simulated assuming a truncated normal distribution. In order to keep θ within the parameter space (0, 0.5), the simulation under the alternative hypothesis was carried out using the truncated normal distribution TN_{ab} (m, σ^2) where a and b are the lower and the upper bounds. This simulation was performed following Devroye (1986)

$\theta = (\mu + \sigma) \ \Phi^{-1} [P_1 - U (P_2 - P_1)] \qquad (6)$

Where Φ^{-1} is the inverse of cumulative distribution function of the normal distribution, U a uniform variate Un(U| 0,1), $P_1 = \phi(\dfrac{a - \mu}{\sigma})$ and $P_2 = \phi(\dfrac{b - \mu}{\sigma})$

Such that the values of a and b will be 0.0 and 0.5 whereas the values of μ and σ differ from one scenario to another (table 1).

Morton test was computed for each of the 10000 replicates, using the same data structure (Table 1), and under the assumptions of the null hypothesis (all families have the same θ). The outcomes of the 10000 Morton tests were sorted in an ascending order. The value representing $(1 - \alpha)$ quantile of such a '*distribution*' was taken as the empirical threshold (t_α), with $\alpha = 0.05$

Genome-wide scan for heterogeneity of the recombination fraction in Norwegian cattle: The Norwegian cattle map (Våge *et al.*, 2000) was based on 6 parental half sib families with a total of 285 sons (The map comprises 261 anonymous microsatellites and 27 coding genes). The genome-wide scanning included 194 intervals distributed along the cattle genome. A genome-wide scan for heterogeneity in the recombination fraction using Morton test was carried out.

RESULTS

Table (2) shows empirical power to detect heterogeneity in the granddaughter design for varying recombination fractions and variances. The larger the distances between markers (in recombination fraction units) the higher the power. These results show that the grand daughter structure of the Norwegian cattle allows testing for heterogeneity when the recombination fraction was 0.2.

Table 2. Empirical power (in percentage) for different values of the recombination fraction and for alternative frequencies in the dam population. (frequencies values represented according to the linkage phase)

Allele freq. in dams	0.05(0.25)	0.1(0.5)	0.2(0.1)	0.2(0.15)
$\begin{array}{cc} \underline{0.8} & \underline{0.8} \\ 0.1 & 0.1 \end{array}$	1.3	7.6	34.2	54.6
$\begin{array}{cc} \underline{0.8} & \underline{0.1} \\ 0.1 & 0.8 \end{array}$	2.6	12.5	39.1	59.3
$\begin{array}{cc} \underline{0.8} & \underline{0.1} \\ 0.1 & 0.1 \end{array}$	3.6	15.3	44.0	64.5
$\begin{array}{cc} \underline{0.1} & \underline{0.1} \\ 0.1 & 0.1 \end{array}$	7.2	21.8	50.8	70.8

Unbiased estimates of the recombination fraction were obtained when simulating both the null (homogeneity) and the alternative hypotheses (heterogeneity). These results are also supported by the equality between the variance and mean square error (Table 3).

Table 3. Estimates of variance among replicates (Var), mean square error (MSE), and bias when simulating heterogeneity of the recombination under a variety of scenarios regarding allele frequencies in the dam population. All values are multiplied by 10^{-3}

Allele freq. in dams		0.05(0.025)			0.1(0.05)			0.2(0.1)			0.2(0.15)		
		Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias
0.8 0.8 / 0.1 0.1	H_0	0.4	0.4	1.3	0.7	0.7	0.2	0.9	0.9	-0.1			
	H_1	0.7	0.5	0.9	1.1	1.1	3.1	2.4	2.4	4.6	3.3	3.6	18.3
0.8 0.1 / 0.1 0.8	H_0	0.3	0.3	0.2	0.9	0.9	-0.1	0.9	0.9	-0.3			
	H_1	0.4	0.4	1.8	0.9	0.9	3.0	2.37	2.3	5.7	3.1	3.5	19.3
0.8 0.1 / 0.1 0.1	H_0	0.3	0.3	0.15	0.5	0.5	-0.1	0.8	0.8	-0.5			
	H_1	0.4	0.4	1.4	0.9	0.9	3.2	2.2	2.2	4.8	3.1	3.4	18.3
0.1 0.1 / 0.1 0.1	H_0	0.2	0.2	0.1	0.4	0.4	0.2	0.7	0.7	-0.3			
	H_1	0.3	0.3	1.4	0.7	0.7	2.8	2.1	2.1	5.7	3.0	3.4	19.5

The results of the genome-wide scan in Norwegian bulls testing for heterogeneity of the recombination fraction are depicted in Table 4. A total of 12 and 9 intervals out of 194 intervals tested for heterogeneity were significant at significance levels of 0.05 and 0.01, respectively. The expected number of significant test just by chance when performing 194 tests is 9.7 and 1.9 at significance levels of 0.05 and 0.01. Therefore, some of the significant intervals might be truly heterogeneous for the recombination fraction in Norwegian bulls.

Table 4. Selected intervals after a genome-wide scan of heterogeneity of the recombination fraction using Morton test (M-test)

BTA	No. families	Markers-interval	M-test
1	2	CSSM32–SSM19	9.90**
1	2	BM148–URB17	6.58*
3	3	FCGRI–INRA3	13.05**
5	5	BL37-AGLA254	13.05*
6	3	BM4528–RM28	9.05**
7	2	BM7160–BM6117	10.76**
7	5	OARAE129–ILSTS6	9.85*
7	4	ILSTS6–INRA53	11.58**
9	4	BM4208–URB24	7.83*
10	2	CSSM39–CSSM46	4.18*
12	3	BM6108–BM6116	6.60*
13	2	BMC1222–BMS1352	9.37**
17	5	ETH185–BM8125	8.39*
17	3	CSSM33–INRA25	12.94**
18	4	BM7109–ILSTS2	11.98**
18	4	ILSTS2–EAC	9.10*
18	2	FCG2R–TGLA227	4.02*
25	2	TGLA40–BP28	5.75*
26	4	ABS12– NOR6	10.41*
27	3	CSSM43–BM1857	6.16*
28	3	IDVGA8–BM2515	10.91**

*$P < 0.05$, ** $P < 0.01$*

DISCUSSION

The purpose of this paper is to show that heterogeneity of the recombination fraction can be tested using the granddaughter design. The family size in the Norwegian cattle population was used to compute power is rather small. Testing for heterogeneity will be powerful in populations with a larger family sizes like the

American dairy cattle where most of grandsires have more than 100 sons each. The simulation.

Results also showed that power of detection is sufficient in Norwegian cattle as long as the length of the interval is at least 20 cM. This is normally, the distance used in genome scan searching for QTL in cattle populations, and therefore, testing of heterogeneity of the recombination fraction can be carried out using the same genotyping data. At low genetic distances, the differences between recombination fractions among bulls are small due to the low chances of recombination events which make difficult the detection of heterogeneity.

The maximum likelihood estimation of the recombination fraction was found to be unbiased for both the null hypothesis (homogeneity) and the alternative hypothesis (heterogeneity). This method assumes that dam allele frequencies are known without error. In practice, estimation of allele frequencies is usually carried out from the same data set (Gomez-Raya, 2001). For large family sizes, maximum likelihood estimates of allele frequencies in the dam population should have a small impact on the estimates of the recombination fraction for each individual family since maximum likelihood estimates are asymptotically unbiased when the sample size is large. More research is needed to understand the impact of using estimates of the dam allele frequencies on testing for heterogeneity when the family size is small. The results of the genome-wide scan of the Norwegian cattle for heterogeneity of the recombination fraction showed that heterogeneity of the recombination fraction might occur among Norwegian bulls. The number of interval significant for heterogeneity was higher than the expected by chance alone. However, and because of multiple testing we can not know which intervals among those significant are truly heterogeneous. Other methods to test heterogeneity could be performed in those significant intervals such as single sperm typing. This technique allows for estimating bulls recombination fraction by typing a number of sperm cells from each bull. Simianer *et al*. (1997) and Szyda (1997) performed single sperm typing for six intervals in three chromosomes of Norwegian cattle. They found evidences of heterogeneity in one interval located in the sexual chromosomes.

Sexual chromosomes were not tested in our study. In addition, Park *et al.* (1995) and Thomsen *et al*. (2001) reported detection of heterogeneity of the recombination fraction. Thomsen *et al*. (2001) reported 22 significant intervals on 15 chromosomes, of which chromosomes 1, 6, 7, 12, 13, 27 and 28 were also reported in the present study, where the intervals were not exactly the same but overlapped. With respect to BTA23 which was reported by both Park *et al.* (1995) and Thomsen *et al*. (2001) no significant intervals were found in the present study. This disagreement may be due to the use of different breeds in both studies, where only the Norwegian cattle were used in our study. The conclusion of the present work is that the granddaughter design can be used to test for heterogeneity of the recombination fraction as long as individual recombination fractions are estimated accounting for the dams allele contributions. Evidences of heterogeneity of the recombination fraction among Norwegian bulls were also shown.

REFERENCES

Devroye L., 1986 Non–Uniform Random Variate Generation. pp. 1-23. Springer–Verlag, NewYork.

Georges M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A.T. Pasquino, L.S. Sargeant.,A. Sorensen, M.R. Steele, and X. Zhao, 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139, 907–920.

Gomez-Raya L. 2001. Biased estimation of the recombination fraction using half–sib families and informative offspring. Genetics 157, 1357–1367.

Khatkar M S., P. C. Thomson, I. Tammen, and H.W. Raadsma, 2004. Quantitative trait loci mapping in dairy cattle: review and meta analysis. Genetics, Selection and Evolution 36, 163–190.

Lien S., N. E. Cockett.H. Klungland , N, Arnheim,M. Georges, and L. Gomez-Raya, 1999. High–resolution gemetic map of the sheep(callipyge) region : linkage heterogeneity among rams detected by sperm typing. Animal Genetics. 30, 42–46.

Lien, S., J. Szyda, N. Arnheim, B. Schechinger, G. Rappold,and N. Arnheim, 2000. Evidence for heterogeneity in recombination in the human Pseudoautosomal region: high resolution analysis by sperm typing radiation–hybrid mapping. American. Journal of Human Genetics 66, 557–566.

Morton N.E., 1956. The detection and estimation of linkage between the genes for elliptocytosis and RH blood type. American .Journal of Human Genetics 8, 80–96.

Olsen H.G., L. Gomez-Raya, D. I. Våge, I. OLsaker, H. Klungland, T. Ådnøy , A. Sabry, G. Klemetsdal, N. Schulman, W. Krämer, G. Thaller, K. Rønning and S. Lien, 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. Journal of Dairy Science, 85, 3124–3130.

Park C., I. Russ, Y. Da, and H.A. Lewin, 1995. Genetic mapping of F13A to BTA23 by sperm typing: difference in recombination rate between bulls in the DYA-PRL interval. Genomics 27, 113–118.

Simianer, H., J., Szyda, Ramon, G., and Lien, S. 1997. Evidence for individual and between-family variability of recombination rate in cattle. Mammalian. Genome 8, 830–835.

Szyda, J. 1997. Variability of the recombination rate in cattle. Ph.D. the- sis, Aus dem institut fur tierhaltung und Tierzuchtung der universtitat hohenheim.

Thomsen H., N. Reinsch, N. Xu, J. Bennewitz, C. Looft, S. Grupe, C. Kühn, G.A. Brockmann, M. Schwerin, B. Leyhe- Horn, S. Hiendleder, G. Erhardt, I. Medjugorac ,I. Russ, M. Forster, B. Brenig, F. Reinhardt, R. Reents, J. Blumel, G. Averdun., and E. Kalm, 2001. A whole genome scan for differences in recombination rates among three Bos taurus breeds. Mammalian. Genome 12, 724–728.

Våge D.I., I. Olsaker, H. Klungland, L. Gomez-Raya, and S. Lien, 2000. A male geneticmap designed for quantitative trait loci mapping in Norwegian cattle. Acta Agriculture Scandinavica-section A-Animal science 50, 56–63.

Yu J., Lazzeroni L., Qin J., Huang M-M., Navidi W., Erlich H., and Arnheim N. 1996. Individual variation in recombination among human males. American .Journal of Human Genetics 59, 1186–1192.

الإختلافات بين الطلائق فى نسبة الإتحادات الجديدة بإستخدام تصميم الحفيدة

أيمن محمود صبرى¹، لويس جوميز-رايا²

١- قسم بيولوجيا الخلية، المركز القومى للبحوث، شارع التحرير ١٢٦٢٢، الدقى، الجيزة، ٢- قسم بيوتكنولوجيا الحيوان، جامعة نيفادا، رنو ص. ب. ٢٠٢، رنو، ن فى ٨٩٥٥٧، الولايات المُتحدة الأمريكية

في هذا البحث تم اختبار الاختلافات في نسبة الإتحادات الجديدة Recombination fraction لنفس البيانات الوراثية في تصميم الحفيدة Granddaughter Design. حيث تم حساب القوه الإحصائية Statistical power بإستخدام محاكاة مونت كارلو للتركيب التربوي للماشية النرويجية. تم الحصول على قوة إحصائية قدرها ٧٠.٨% عندما كان تكرار أليلات الواسمات المحيطة ٠.١ . وذلك عندما كانت القيمة الحقيقية لنسبه الإتحادات الجديدة 0.2 بإنحراف معياري قدره ٠.١٥ كما أوضح البحث أن القوة الإحصائية إنخفضت بدرجة ملحوظة مع انخفاض القيمة الحقيقية لمتوسط الإتحادات الجديدة . كما تم عمل مسح جينومى لإختبار الاختلافات في نسبة الإتحادات الجديدة باستخدام ١٩٤ مقطعاً في الماشية النرويجية حيث كان إجمالي المقاطع المعنوية ١٢، ٩ مقطعاً معنوياً عند مستوى معنوية ٠.٠٥ و ٠.٠١.

أظهرت النتائج الحالية أن هناك تباين واضح لوجود إختلافات وراثية بين نسب الإتحادات الجديدة بين الطلائق النرويجية.