Note

Survival Quantitative Trait Locus Fine Mapping by Measuring and Testing for Hardy–Weinberg and Linkage Disequilibrium

J. Casellas¹

Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona,
08193 Bellaterra, Barcelona, Spain

Manuscript received October 24, 2006

Accepted for publication February 25, 2007

ABSTRACT

I show that fine-scale localization of a survival-related locus can be accomplished on the basis of deviations from Hardy–Weinberg equilibrium and linkage disequilibrium at closely linked marker loci. The method is based on χ^2 -tests and they can be performed for age-specific samples of alive (or dead) individuals, as for combined samples of alive and dead individuals.

CONVENTIONAL tools for the analysis of QTL can locate loci underlying the variation of continuous quantitative traits to a genomic region of ~ 30 cM (DENG *et al.* 2000). Fine-scale mapping (~ 1 cM) is required to reduce the range of these candidate genomic regions, and some appropriate techniques have been developed for complex diseases and quantitative traits under Gaussian distributions (SPIELMAN *et al.* 1993; FEDER *et al.* 1996; NIELSEN *et al.* 1998; DENG *et al.* 2000, 2003). Although survival has become an emergent research field in human health (PUCA *et al.* 2001) and animal breeding (KLEINBAUM 1996), there are not appropriate fine-mapping techniques for survival traits. The objective of this article is to adapt DENG's *et al.* (2000) QTL fine-mapping method to survival data.

Take as starting point a survival-related QTL locus with two alleles, A_1 and A_2 , and allelic frequencies p and $q = 1 - p$, respectively. Under the proportional hazards framework (Cox 1972), $S(t | A_1 A_1) = S_0(t)^{\exp(a)} = \phi_{11}$ is the survival probability at time t for an individual with genotype $A_1 A_1$, where $S_0(t)$ is the baseline survival function and a is the genotypic value of the $A_1 A_1$ genotype. In a similar way, we define $S(t | A_1 A_2) = S_0(t)^{\exp(d)} = \phi_{12}$ and $S(t | A_2 A_2) = S_0(t)^{\exp(-a)} = \phi_{22}$, d and $-a$ being the genotypic values for the $A_1 A_2$ and $A_2 A_2$ genotypes, respectively. Without loss of generality, we can assume that $S_0(t)$ represents a random variable for the combined effects of all the rest of the polymorphic loci and all random environmental effects. As in the original research

of FEDER *et al.* (1996), NIELSEN *et al.* (1998), and DENG *et al.* (2000), a large population under random mating is assumed and thus Hardy–Weinberg (HW) equilibrium holds in each generation of individuals at birth. The proportion of survivors at time t (π_t) is stated as $\pi_t = p^2 \phi_{11} + 2pq \phi_{12} + q^2 \phi_{22}$, $P_{A_1 | \text{ALIVE}_t} = (p^2 \phi_{11} + pq \phi_{12}) / \pi_t$ and $P_{A_1 A_1 | \text{ALIVE}_t} = p^2 \phi_{11} / \pi_t$ being the allelic (A_1) and genotypic ($A_1 A_1$) frequency within the group of alive individuals at time t (ALIVE_{*t*}), respectively (the remaining frequencies can be easily derived following DENG *et al.* 2000). Deviation from HW equilibrium at the survival QTL can be measured by the disequilibrium coefficient $D_{A_1 A_1}$ by WEIR (1996), $D_{A_1 A_1} = P_{A_1 A_1 | \text{ALIVE}_t} - P_{A_1 | \text{ALIVE}_t}^2 = p^2 q^2 (\phi_{11} \phi_{22} - \phi_{12}^2) / \pi_t^2$ or, following DENG *et al.* (2000), by the function between observed and expected homozygosities, $F_{A_1} = pq(\phi_{11} \phi_{22} - \phi_{12}^2) / \pi_t^2$.

Previous derivations can be easily adapted to a marker locus closely located near the survival QTL, with alleles M_1 and M_2 , and allelic frequencies r and $s = 1 - r$. As in DENG *et al.* (2000), $P_{M_1 | \text{ALIVE}_t} = r + D_{A_1 M_1} [p \phi_{11} + (q - p) \phi_{12} - q \phi_{22}] / \pi_t$ is the allelic frequency of M_1 and $P_{M_1 M_1 | \text{ALIVE}_t} = (P_{A_1 M_1}^2 \phi_{11} + 2P_{A_1 M_1} P_{A_2 M_1} \phi_{12} + P_{A_2 M_1}^2 \phi_{22}) / \pi_t$ is the genotypic frequency of $M_1 M_1$, where $D_{A_1 M_1} = P_{A_1 M_1} - P_{A_1} P_{M_1}$ is the linkage disequilibrium (LD) measure between A_1 and M_1 (CROW and KIMURA 1970) and $P_{A_1 M_1}$ is the frequency of haplotypes carrying both A_1 and M_1 . According to DENG *et al.* (2000), the HW disequilibrium among ALIVE_{*t*} individuals at the marker locus is $D_{M_1 M_1 | \text{ALIVE}_t} = P_{M_1 M_1 | \text{ALIVE}_t} - P_{M_1 | \text{ALIVE}_t}^2 = (\phi_{11} \phi_{22} - \phi_{12}^2) D_{A_1 M_1}^2 / \pi_t$, it being nonzero when $(\phi_{11} \phi_{22} - \phi_{12}^2) \neq 0$ and $D_{A_1 M_1}^2 \neq 0$. A wide range of combinations of ϕ_{11} , ϕ_{12} , and ϕ_{22} provide a value different from zero and, in practice, the HW disequilibrium at

¹Address for correspondence: Genètica i Millora Animal, IRTA-Lleida, 25198 Lleida, Spain. E-mail: joaquim.casellas@irta.es

FIGURE 1.—Comparison of the statistical power of the χ^2_{HW} -test (open boxes) and the χ^2_{LD} -test (solid boxes) under various genetic models: (A) additive ($a = -0.69$, $-a = 0.69$, $d = 0$), (B) dominant ($a = -0.69$, $-a = 0.69$, $d = a$), (C) recessive ($a = -0.69$, $-a = 0.69$, $d = -a$), (D) partial dominant ($a = -0.69$, $-a = 0.69$, $d = a/2$), and (E) partial recessive ($a = -0.69$, $-a = 0.69$, $d = -a/2$). The bottom and top edges of the boxes represent the sample 25th and 75th percentiles; the whiskers extend the range of the results. Under a specific set of simulation parameters, the population started at generation zero, with complete association between allele A_1 ($p = 0.5$) at the QTL and marker allele M_1 ($r = 0.5$), and evolved for 50 generations, under random mating and genetic drift. A set of dense marker loci positioned at 0.25-cM intervals and that span 0–2 cM of the QTL were simulated, with the recombination rate obtained from Haldane's map function (OTT 1991). The first 100 populations for which the difference in p at the start and at the end of evolution do not differ by $>5\%$ were retained. The effective population size per generation was 15,000. Survival records were simulated under a proportional hazard model (Cox 1972), assuming a Weibull distribution with parameters $\rho = 1.5$ and $\lambda = 0.001$ for the baseline survival function ($S_0(t) = \exp(-\lambda^\rho t^\rho)$), and the threshold between alive and dead individuals was assumed at $t = 1500$. In each case, 5000 appropriate samples of 200 individuals (sampling with replacement) from each of the 100 simulated populations were sampled and the statistical power was calculated as the percentage of times that the null hypothesis of no disequilibrium was rejected.

the marker locus solely reflects the LD in the whole generation (DENG *et al.* 2000). In a similar way, HW disequilibrium for the marker locus among alive individuals can be derived as $F_{M_1|\text{ALIVE}_t} = D_{A_1 M_1}^4 F_{A_1|\text{ALIVE}_t} / (p_{qrs})^2$, as described by FEDER *et al.* (1996) and NIELSEN *et al.* (1998) for affected individuals of complex traits. Both F_{M_1} and $D_{M_1 M_1}$ statistics converge to the key point that HW disequilibrium at a marker locus corresponds

to the whole-generation LD between the marker locus and the QTL (DENG *et al.* 2000). Alternatively, one could use a direct measure of LD like the p_{excess} statistic proposed by BENGTSOON and THOMSON (1981). For a survival QTL, p_{excess} becomes $p_{\text{excess}} = P_{M_1|\text{ALIVE}_t} - P_{M_1|\text{DEAD}_t} / 1 - P_{M_1|\text{DEAD}_t}$, where $P_{M_1|\text{DEAD}_t}$ was the allelic frequency of M_1 within the group of dead individuals at time t (DEAD_t). Therefore, p_{excess} is proportional to

FIGURE 2.—Comparison of type I error of the χ^2_{HW} -test (open square) and the χ^2_{LD} -test (solid square) under the five inheritance models: additive, dominant, recessive, partial dominant, and partial recessive. Type I error for both tests was calculated as the percentage of times that the null hypothesis of no disequilibrium was rejected when the simulations were performed under the null hypothesis of no linkage disequilibrium. As is described in Figure 1, simulation parameters were $p = 0.5$, $2n = 200$, $t = 1500$, $\lambda = 0.001$, and $\rho = 1.5$. The square represents the average value and the whiskers extend the range of the results.

$D_{A_1M_1}$ and reaches its maximum at the marker with the greatest LD with the QTL (NIELSEN *et al.* 1998).

To test for the statistical significance of the HW disequilibrium measures ($D_{M_1M_2|ALIVE_t}$ and $F_{M_1|ALIVE_t}$) and the LD measure (p_{excess}), two χ^2 -tests can be easily applied. Following DENG *et al.* (2000), the χ^2 -test statistic for HW disequilibrium is derived as

FIGURE 3.—Average power under different sample sizes ($2n$) for the (A) χ^2_{HW} - and (B) χ^2_{LD} -tests (the simulation process is described in Figure 1).

FIGURE 4.—Comparison of statistical power for various temporal cut points (t). The average results for the (A) χ^2_{HW} - and (B) χ^2_{LD} -tests are presented (the simulation process is described in Figure 1).

$$\chi^2_{HW} = 2n \left[\frac{(\tilde{P}_{M_1M_2|ALIVE_t} - 2\tilde{P}_{M_1|ALIVE_t}\tilde{P}_{M_2|ALIVE_t})^2}{2\tilde{P}_{M_1|ALIVE_t}\tilde{P}_{M_2|ALIVE_t}} + \sum_{i=1}^2 \frac{(\tilde{P}_{M_iM_1|ALIVE_t} - \tilde{P}_{M_i|ALIVE_t}^2)^2}{\tilde{P}_{M_i|ALIVE_t}^2} \right],$$

where the tilde ($\sim$) denotes an estimated value from the sample and $2n$ is the total sample size of individuals. The χ^2_{HW} test has $[m(m-1)/2] = 1$ d.f., m being the number of alleles at the marker locus being tested ($k = 2$). On the other hand, the χ^2 for p_{excess} (WEIR 1996; DENG *et al.* 2000) is stated as

$$\chi^2_{LD} = 2n \sum_{i=1}^2 \frac{(\tilde{P}_{M_i|ALIVE_t} - \tilde{P}_{M_i|DEAD_t})^2}{\tilde{P}_{M_i|ALIVE_t} + \tilde{P}_{M_i|DEAD_t}}$$

with $m - 1 = 1$ d.f.

To illustrate the tests outlined above, extensive computer simulations were performed for a biallelic survival QTL and several biallelic markers. These computer simulations were carried out under a wide range of inheritance models (additive, dominant, recessive, partial dominant, and partial recessive), sample sizes, and ages and under a Weibull assumption for the baseline survival function (DUCROCQ *et al.* 1988a,b; IBRAHIM *et al.* 2001). For the five genetic models, both tests showed reduced power at greater distances between the QTL and the marker loci, although the power decayed more quickly for χ^2_{HW} than for χ^2_{LD} , and it was higher for χ^2_{LD}

than for χ^2_{HW} (Figure 1). These results agree with previous QTL fine-mapping research (NIELSEN *et al.* 1998; DENG *et al.* 2000) and they are not surprising because, in models where both the survival QTL and the marker locus have only two alleles, HW disequilibrium is proportional to the square of LD (NIELSEN *et al.* 1998). Whereas χ^2_{LD} provided a similar power for all genetic models, χ^2_{HW} showed substantial discrepancies. Within this context, χ^2_{LD} seemed preferable if samples of both alive and dead individuals were accessible, although they could be unavailable if the study was not previously scheduled. On the other hand, the average type I error of both tests was close to the expected level of 0.05, slightly higher for χ^2_{HW} than for χ^2_{LD} (Figure 2). This larger variation in χ^2_{HW} was consistent with previous research (NIELSEN *et al.* 1998; DENG *et al.* 2000). As was expected, the average power of both tests at the different marker positions increased as the amount of available information increased (*e.g.*, the number of sampled individuals increases; Figure 3) or the selection criteria became more strict (*e.g.*, elderly ages to differentiate between alive and dead individuals; Figure 4). These results agreed with those of DENG *et al.* (2000).

In conclusion, LD is captured and magnified in extreme samples of elderly individuals, where QTL genotypes and alleles are disproportionately represented. The disequilibrium must be the highest at the QTL locus, since it is the underlying factor that determines the selection criterion, and it decreases as the degree of linkage between the QTL and the markers decreases. This relation between the HW equilibrium and/or LD and the physical distance between a panel of linked marker loci and a QTL is the key point that provides a straightforward basis for QTL fine mapping with use of the peaks of the disequilibrium measures and/or test statistics.

LITERATURE CITED

- BENGTTSSON, B. O., and G. THOMSON, 1981 Measuring the strength of association between HLA antigens and diseases. *Tissue Antigens* **18**: 356–363.
- COX, D. R., 1972 Regression models and life-tables. *J. R. Stat. Soc. Ser. B* **34**: 187–220.
- CROW, L., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- DENG, H.-W., W.-M. CHEN and R. R. RECKER, 2000 QTL fine-mapping by measuring and testing for Hardy-Weinberg and linkage disequilibrium at a series of linked marker loci in extreme samples of populations. *Am. J. Hum. Genet.* **66**: 1027–1045.
- DENG, H.-W., Y.-M. LI, M.-X. LI and P.-Y. LIU, 2003 Robust indices of Hardy-Weinberg disequilibrium for QTL fine mapping. *Hum. Hered.* **56**: 160–165.
- DUCROCQ, V., R. L. QUAAS, E. J. POLLAK and G. CASELLA, 1988a Length of productive life of dairy cows. 1. Justification of a Weibull model. *J. Dairy Sci.* **71**: 3061–3070.
- DUCROCQ, V., R. L. QUAAS, E. J. POLLAK and G. CASELLA, 1988b Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation. *J. Dairy Sci.* **71**: 3071–3079.
- FEDER, J. N., A. GNIRKE, W. THOMAS and Z. TSUCHIHASHI, 1996 A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat. Genet.* **13**: 399–408.
- IBRAHIM, J. G., M. CHEN and D. SINHA, 2001 *Bayesian Survival Analysis*. Springer, New York.
- KLEINBAUM, D. G., 1996 *Survival Analysis: A Self-learning Text*. Springer, New York.
- NIELSEN, D. M., M. G. EHM and B. S. WEIR, 1998 Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.* **63**: 1531–1540.
- OTT, J., 1991 *Analysis of Human Genetic Linkage*. Johns Hopkins University Press, Baltimore.
- PUCA, A. A., M. J. DALY, S. J. BREWSTER, T. C. MATISE, J. BARRETT *et al.*, 2001 A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc. Natl. Acad. Sci. USA* **98**: 10505–10508.
- SPIELMAN, R. S., R. E. MCGINNIS and W. J. EWENS, 1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**: 506–516.
- WEIR, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Communicating editor: J. B. WALSH