Haplotype Sharing for Fine Mapping Quantitative Trait Loci Controlling Trypanotolerance in Mice

Amwayi, P. W.[1,2], *Limoh, M. K.*[1], *Agaba, M.*[2], *Kemp, S. J.*[3], *Gibson, J. P.*[4] *and Iraqi, F. A.*[2]

[1]*Department of Biochemistry and Molecular Biology, Egerton University P. O. Box 536, Egerton, Kenya*

[2]*International Livestock Research Institute, P. O. Box 30709, Nairobi, Kenya*

[3]*School of Biological Sciences, University of Liverpool L69 7ZD, UK*

[4]*Institute for Genetics and Bioinformatics, Armidale, NSW 235, Australia*

Abstract

Quantitative trait loci (QTL) mapping and fine mapping in mouse models demonstrates the possibility of localizing genes that determine genetic variations in inbred strains. Previously, three quantitative trait loci (QTL), *Tir1*, *Tir2* and *Tir3* on chromosome 17, 5 and 1 respectively associated with resistance to trypanosomosis, were mapped in two F2 resource populations, (C57BL/6J x A/J) and (C57BL/6J x BALB/cJ). The QTL were mapped within 10-40cM genomic intervals (CI). Subsequently, using F6 advanced intercross lines (AILs), the QTL were fine mapped to a smaller Confidence interval (CI), but not sufficient for positional cloning. C3H/HeJ and 129/J mouse strains are relatively susceptible, however, it is not known if this is due to 'susceptible' alleles at the previously identified QTL. To determine this, an F2 cross (C57BL/6J x C3H/HeJ) was developed and challenged with *Trypanosoma congolense* and the response and survival time monitored. Interval mapping identified significant QTL on chromosomes 17 and 1, however, *Tir2* was not confirmed. Following the confirmation of the QTL on chromosome 1 and 17, the conserved susceptible/resistance (QTLs) regions between A/J, BALB/cJ, C3H/HeJ, 129/J and C57BL/6J were explored using single nucleotide polymorphism (SNP) haplotypes for fine mapping these QTL. The QTL precision was increased significantly from 30-40cM to less than 1cM which is now adequate for positional cloning.

Key Words: *Trypanosoma congolense*, Quantitative trait loci, Single nucleotide polymorphism, Haplotypes, Confidence interval

Introduction

Traits that show continuous variation in a population are referred to as complex traits or quantitative traits. Quantitative trait loci mapping involves the use of evenly spaced polymorphic DNA markers to correlate between marker alleles and the phenotype variation (Schork *et al*., 1996). This work has attracted considerable research interest for several years and efforts are being undertaken to map genes that determine quantitative genetic variation with little success. This is mostly as a result of current detection methods which place quantitative trait loci (QTLs) within very large interval not adequate for positional cloning. Recently, an in silico SNP haplotype mapping strategy was proposed to accelerate the identification of genes associated with complex traits (Grupe *et al*., 2001). With the availability of various web-accessible murine SNP databases, the chromosomal regions that most likely contribute to trypanotolerance could be narrowed down thus the potential candidate gene list. Evaluation of genetic variation patterns has been reported in two recent studies (Frazier *et al*., 2004; Yalcin *et al*., 2004) based on fine resolution haplotype structure across multiple strains. These findings suggest that a high resolution SNP map is required to obtain an accurate description of the genetic variations in the laboratory mouse genome. Therefore, a detailed exploration of common ancestral regions that lie between strains can hasten QTL mapping by identification of shared regions for consideration as candidate loci (Wade *et al*., 2002). Examination of the haplotype structure across the QTL candidate region might reveal regions that segregate appropriately with the phenotype of the strains.

In a recent study, three loci, *Tir1*, *Tir2* and *Tir3* were identified and mapped on chromosome 17, 5 and 1 respectively, with confidence intervals (CIs) in the range 10-40cM that control significant genetic differences between resistant and susceptible mice after linkage mapping studies on two F2 crosses (C57BL/6 x A/J and C57BL/6 x BALB/c strains). However, the confidence interval report was too large to facilitate positional cloning. A subsequent fine mapping of the *Tir1*, *Tir2* and *Tir3* loci was carried out using advanced intercross lines (AIL) created by crossing the C57BL/6 strain with the A/J and BALB/c strains respectively. Darvasi and Soller (1995) introduced AIL approach where random mating over a number of generations from the F2 is used to accumulate meiosis for the purpose of high resolution mapping of QTL by the time F6 to F8 generation is attained. Consequently, *Tir1*, *Tir2* and *Tir3* loci showed significant improved resolution revealing a single region in each, however, the AIL analyses revealed a degree of

complexity at the *Tir3* which appeared to resolve into three distinct regions (Clapcott, 1998, Iraqi *et al.*, 2000). The foregoing highlights the need to utilize haplotype mapping to achieve very fine localization of the QTL that will enhance positional cloning or position candidate gene identification. In this study, single nucleotide polymorphisms (SNP) database were used to predict variation in the mouse genome, where the recently assembled genome sequence of C57BL/6J (resistant) strain is aligned with sample sequences of other susceptible strains. The existence of common haplotype patterns in the four susceptible strains, A/J, BALB/cJ, C3H/HeJ, and 129/J will be a reflection of recent evolutionary origins, thus representing ancestral haplotypes. The identification of haplotypes can be effective in reducing QTL intervals to sizes amenable to analysis of the candidate gene (Wiltshire *et al.*, 2003).

Materials and Methods

Generation of F2 Resource Population

Parental lines C57BL/6J (trypanosomosis resistant), C3H/HeJ (susceptible) were obtained from Harlan Ltd, UK. From each cross, 120 mice of FI (C3H/HeJ X C57BL/6J) were developed by mating 20 (10 males and 10 females) C3H/HeJ mice with 20 (10 males and 10 females) C57BL/6J mice. Sixty breeding pairs of the F1 generation were intercrossed to produce 345 F2 in each cross (C3H/HeJ X C57BL/6J) generation and used in this study. Breeding of the F2 population and subsequent phenotyping was done at the small animal unit, International Livestock Research Institute (ILRI), Kenya.

Trypanosomosis Challenge, Phenotyping and Genotyping

The F2 (C3H/HeJ X C57BL/6J) resource populations together with the control parental mice were challenged at 12 weeks of age by intraperitoneal inoculation of 4 x 10^4 blood stream forms of *T. congolense* clone IL1180 (Masake *et al*, 1983; Nantulya *et al,* 1984). In the following 14 days, blood samples were collected daily from the tail tip of all challenged mice and examined microscopically. Phenotypic data was defined as survival time in days following the day of challenge. The first group to succumb were taken as the most susceptible (S), while the last one to succumb to infection were presumed to be resistant (R), and those ones in between were the intermediate group (I). Non- parasitemic mice were excluded from further

analysis. DNA was extracted from mouse tail and genotyped with microsatellite markers, which previously confirmed to be informative between C57BL/6J and C3H/HeJ mouse strains and located within the six previously, mapped (trypanotolerance) QTL intervals. A selective genotyping approach was used in this experiment (Ronin *et al.*, 2003 and Darvasi, 1997).

Linkage Analysis and QTL Mapping

Genotype frequencies in resistant and susceptible groups of mice were checked against Hard-Weinberg equilibrium (HWE) (Deng *et al.*, 2000, 2003; Deng and Chen 2000). Multipoint analysis was performed with MAPMAKER/EXP version 3.0 (Lincoln *et al.*, 1992a), and map distances were calculated with the Haldane function. QTL interval mapping analysis were performed with the maximum likelihood (ML) approach of MAPMAKER/QTL version 1.1 (Lincoln *et al.*, 1992b) and with the least Square (LS) approach of QTL express (Seaton *et al.*, 2002). The significant LOD score in Mapmaker/QTL was defined as the interval above the two LOD scores; however in QTL express permutation test was run 1000 times randomly across the data and significant F value (LOD score) was defined by the software. QTL express was used with marker orders and distances from the sequenced mouse genome (Waterston *et al.*, 2002). The markers used were first tested for linkage analysis as described (Lincoln and Lander 1992). The QTL position and significance was confirmed by maximum likelihood estimation method using Mapmaker/QTL programs by incorporating marker order (Lander *et al.*, 1987).

Single Nucleotide Polymorphism (SNP) Haplotype Analysis

Markers flanking each QTL were identified by the linkage analysis and subsequently mapped on the mouse genome sequence database. These markers were used to identify the reference SNP positions within the assembled genome sequences for database query. The mapped QTL region in five different mouse strains (i.e. 129/J, BALB/cJ, A/J, C3H/HeJ and C57BL/6J were aligned to sequenced mouse genome to identify the shared single nucleotide polymorphism (SNP) for the purpose of fine mapping and possible candidate gene(s) identification. The mouse SNP database http://mousesnp.roche.com/cgi-bin/msnp_public.pl and the Jackson Laboratory database http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn-docs/home was screened for SNPs within each QTL for total SNPs available,

and then filtered to keep only the SNPs for 129/J, A/J, BALB/cJ, C3H/HeJ and C57BL/6J.

Results and Discussion

Phenotyping

The survival time of F2 (C3H/HeJ x C57BL/6J) cross showed the highest survival rate compared with the parental strains. In addition, the resistance/ susceptibility status of C57BL/6J and C3H/HeJ were confirmed phenotypically in this study where C57Bl/6J had a higher survival time than C3H/HeJ and A/J mouse strains (Morrison *et al.*, 1978). The mean survival times in days were 53, 63, 87 and 97 for A/J, C3H/HeJ, C57BL/6J and F2 respectively. None of the C3H/HeJ and A/J mice survived the challenge.

Linkage Analysis

The QTL analysis revealed presence of loci that influence survival of mice under trypanosome challenge on chromosome 1 and 17 (Figure 1), agreeing with earlier reports by Kemp *et al.*, (1997). The QTLs *Tir1* and *Tir3* mapped on Chr 17 and 1 QTLs comprised a single locus with LOD scores of 4.864 and 4.59 in C3H/HeJ x C57BL/6J respectively. The threshold obtained with least square method was comparable to the two LOD score level of significance assigned to the mapmaker results. The summary of the QTLs mapped is as indicated in Table 1. Mapped QTLs confirmed that C3H/HeJ mouse strain possessed the susceptible alleles at *Tir1* and *3* that had been previously mapped in A/J and BALB/cJ mouse strains (Kemp *et al.*, 1997).

Chr5 (*Tir2*) was not mapped in F2 (C3H/HeJ X C57BL/6J) crosses as previously reported (Kemp *et al.*, 1997). Earlier studies have showed no evidence of chromosome 5 QTL in (BALB/cJ x C57BL/6) F6 population and it was thought that it could have been due to loss of the allele resulting from many recombination events during the development of the AIL (Iraqi *et al.*, 2000). In this study, this is unlikely since at F2 level, recombinations are very limited (Darvasi and Soller, 1995).

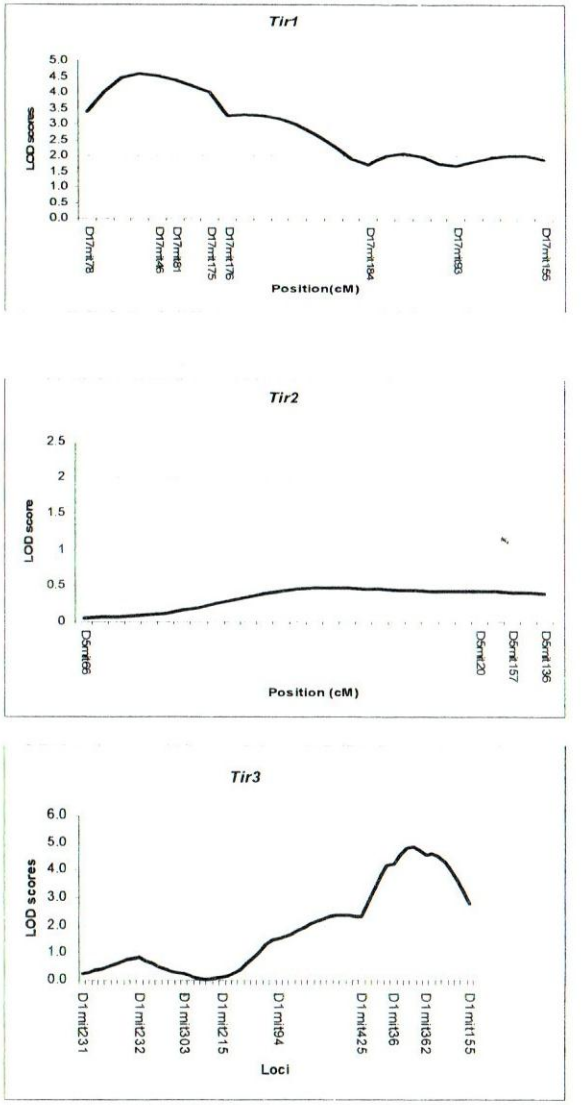

Figure 1. LOD scores of putative QTLs for trypanotolerance. (A) *Tir1* (Chr 17 QTL), (B) *Tir2* (Chr 5 QTL) and (C) *Tir3* (Chr1 QTL).

Lack of confirmation of chromosome 5 QTL on the *Tir2* region might be due to small genetic variation between 129/J, C3H/HeJ and C57BL/6J hence resulting to less significant LOD scores, hence the QTL could have been too weak to be detected. In addition, this apparent loss of *Tir2* from F2 (C3H/HeJ X C57BL/6J) crosses may have been due to an allele in chromosome 5 within C57BL/6, C3H/HeJ and 129J mouse strains having the same function, meaning chromosome 5 in 129/J and C3H/HeJ does not carry the susceptible allele of trypanotolerance.

This was postulated that it might have been inherited from either one of three wild mouse strains where the inbred lines were developed from and or that an allele in C57BL/6 in this particular locus is not yet completely developed.

Table 1. Locations and statistics for putative QTL in F2 C3H/HeJxC57BL/6J cross

Chr	F-value	$P<0.05$	$P<0.01$	LOD score*	LOD score**	Position (cM)	Flanking marker
17	11.4	4.815	6.724	4.38	4.59	10cM	D17Mit78-D17Mit184
5	1.58	5.623	7.534	0.506	0.419	-	-
1	12.44	5.238	7.383	4.733	4.864	92.3cM	D1Mit425-D1Mit155

*Least square analysis; ** Maximum likelihood

Chromosomes 2, 3 and 15 did not show any significant QTL in both crosses (Table 2). The LOD scores were below the threshold of LOD Score 2, though these QTLs had been confirmed (McLeod *et al.*, 2002) using the FDR approach on F2 and F6 Mouse populations. From this, it is then possible that the genome-wide (maximum-likelihood) approach used in the analysis was too restrictive whereby the LOD score of 2 in the analysis was too high to detect QTL with very small effects (LOD score below two).

Table 2. Chromosome 2, 3 and 15 least square QTL analysis results

Chr	F-value	LOD score	$P<0.05$	$P<0.01$	Significance
2	0.5	0.216	4.763	7.188	Not significant
3	3.51	1.467	5.209	8.078	Not significant
15	1.7	0.724	4.814	7.564	Not significant

Haplotype Scans

This method was predicted by Grupe *et al.*, 2001 where murine single nucleotide polymorphism (SNP) database are scanned and on the basis of known mice phenotypes and genotypes; one can predict the chromosomal regions that most likely contribute to complex traits. The QTL on chromosome 17 and 1 were confirmed in C3H/HeJ and 129/J study and were found at similar positions as those previously mapped in A/J and BALB/cJ. Shared haplotypes were defined by identifying the longest regions of contiguous strain pair identity and also by taking all other strains with same allele and colouring them as shared haplotypes within that region. The conserved haplotype regions obtained in A/J, BALB/cJ, C3H/HeJ, 129/J comparing with C57BL/6J was used to refine these QTL. The location of the gene underlying this QTL was narrowed down, however, the regions resolved into many sub regions where the resistant and susceptible strains differed. This study is consistent with Bonhomme *et al.*, 1987 where, the genome of laboratory inbred mice were predicted to be a 'mosaic' of regions with origins in the different subspecies, but a clear description of this variation has remained largely elusive due to lack of high resolution data across the genome. This work, therefore, indicates that the use of haplotype mapping approach as a high resolution mapping tool increasing the resolution of the QTLs leading to consideration of possible candidate genes. Thus, the decreased pool of positional candidate genes potentially represents the genes controlling resistance to trypanosomosis.

Conclusions

This study was undertaken to refine the position of trypanotolerance QTL, *Tir1*, *Tir2* and *Tir3* mapped previously (Kemp *et al.*, 1997; Iraqi *et al.*, 2000). This study indicates that in silico SNP haplotype analysis might be a useful strategy for mapping complex traits. Although a controversial idea (Chesler *et al* 2001; Darvasi 2001), combining the developing of mouse SNP databases with experimental crosses has provided an important tool in narrowing the large list of potential candidate genes to a handful of genes for further analysis. However, these genes still need to be examined further, using different techniques and strategies to reduce the number of high priority candidate genes.

A survey of literature for genes lying in the haplotype regions that are likely to be involved in the pathology of trypanosomosis need to be undertaken.

This study further shows us that genetic approaches are still the best optimism of identifying genes for trypanotolerance to allow MAS and MAI breeding schemes to improve livestock productivity. The only difficulty it faces is reducing the list of candidate genes to the smallest number possible and quickly identifying those that have the greatest likelihood of influencing the phenotype being studied.

References

Bonhomme, F., Guénet, J. L., Dod, B., Moriwaki, K. and Bulfield, G. (1987). The polyphyletic origin of laboratory inbred mice and their rate of evolution. J. Linn. Soc. 30:51-58.

Clapcott, S.J. (1998). Localisation of genes controlling resistance to trypanosomiasis in mice (*Mus musculus*). Ph.D. Thesis. University of Liverpool, UK.

Chelser, E.J., Rodriguez-Zas, S.L. and Mogil, J.S. (2001). In silico mapping of mouse quantitative trait loci. Science 294:2423.

Darvasi, A. (2001). In silico mapping of mouse quantitative trait loci. Science 294:2423.

Darvasi, A. and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207.

Darvasi, A. (1997). The effect of selective genotyping on QTL mapping accuracy. Mammalian Genome 8(1):67-68.

Deng, H.W. and Chen, W.M. (2000). QTL fine mapping, in extreme samples of finite populations, for complex traits with familial correlation due to polygenes. American Journal of Human Genetics 67(1):259-262.

Deng, H.W., Chen, M.W. and Recker, R.R. (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. American Journal of Human Genetics 66(3):1027-45

Deng, H.W., Li, Y.M. Li, M.X. and Liu, P.Y. (2003). Robust indices of Hardy-Weinberg disequilibrium for QTL fine mapping. Human Heredity 56(4):160-165.

Frazer, K.A., Wade, C.M., Hinds, D.A., Patil, N. and Cox, D.R. (2004). Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6mb of mouse genome. Genome Researxh 14:1493-1500.

Grupe, A., Germer, S., Usuka, J., Aud, D., Belknap, J.K., Klein, R.F., Ahluwalia, M.K., Higuchi, R. and Peltz, G. (2001). In silico mapping of complex disease-related traits in mice. Science 292:1915–1918.

Haley, S.C. and Knott, A.S. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324.

Iraqi, F., Clapcott, S.J., Kumari, P., Haley, C.S., Kemp, S. and Teale, A.L. (2000). Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines. Mammalian Genome 11:645-648.

Kemp, S.J., Iraqi, F., Darvasi, A., Soller, M. and Teale, A.J. (1997). Localization of genes controlling resistance to trypanosomiasis in mice. Nat. Genet. 16:193-196.

Lander, E., Green, P., Abrahamson, J., Barlow, A. and Darley, M. (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Lincoln, S.E and Lander, E.S (1992). Systematic detection of errors in genetic linkage data. Genomics 14: 604-610.

Lincoln, S.E, Daly, M. and Lander, E.S. (1992a). Constructing genetic maps with MAPMAKER/EXP 3.0 Whitehead Institute Technical Report, Cambridge, Mass.

Lincoln, S.E, Daly, M, and Lander E.S. (1992b). Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1 Whitehead Institute Technical Report, Cambridge, Mass.

Macleod, K.A., Iraqi, F., Haley, S.C. and Gibson, P.J. (2002). Using the false discovery rate in the detection of QTL for trypanosomosis resistance in mice. Proceeding for the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France. 19-23. August. 32:753-756.

Masake, R.A., Musoke, A.J. and Nantulya, V.M. (1983). Specific antibody responses to the variable surface glycoproteins of *Trypanosoma congolense* in infected cattle. Parasite Immunology (oxf) 5:345-355.

Morrison, W.I., Relates, G.E., Mayor-Withey, K.S. and Murray, M. (1978). Susceptibility of inbred strains of mice to *Trypanosoma congolense*: correlation with changes in spleen lymphocyte population. Clin.Exp. Immunol. 32:25-40.

Nantulya, V.M., Musoke, A.J., Rurangirwa, F.R. and Moloo, S.K. (1984). Resistance of cattle to tsetse-transmitted challenge with *Trypanosoma brucei* or *Trypanosoma congolense* after spontaneous recovery from syringe passed infections. Infection and Immunity 43:735-738.

Ronin, Y., Korol, A., Shtemberg, M. Nevo, E. and Soller, M. (2003). High-resolution mapping of quantitative trait loci by selective recombinant genotyping. Genetics 164(4):1657-66.

Schork, N.J., Nath, S.P., Lindpaintner, K. and Jacob, H.J. (1996). Extensions to quantitative trait locus mapping in experimental organisms. Hypertension 28:1104-1111.

Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M. and Visscher P.M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics 18:339–340.

Wade, C.M., Kulbokas III, E.J., Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K. and Daly, M.J. (2002). The mosaic structure of variation in the laboratory mouse genome. Nature 420:574-578.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R. and Alexandersson, M. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520-562.

Wiltshire, T., Pletcher, M.T., Batalov, S., Barnes, S.W., Tarantino, L.M., Cooke, M.P., Wu, H., Smylie, K., Santrosyan, A., Copeland, N.G., Jenkins, N.A., Kalush, F., Murai, R.J., Glynne, R.J., Kay, S.A., Adams, M.D. and Fletcher, C.F. (2003). Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. Proc. Natl. Acad. Sci. 100(6):3380-3385.

Yalcin, B., Fullerton, J., Miller, S. Keays, D.A., and Brady, S. (2004). Unexpected complexity in the haplotypes of commonly used inbred strains of laboratory mice. Proc. Natl. Acad. Sci. 101:9734-9739.

Eger. J. Spec. Issue (2007) 7(S):90-101