# Identifying candidate genes for the regulation of the response to *Trypanosoma congolense* infection


Noyes HA[1]

Agaba M[2]

Ogugo M[2]

Brass A[3]

Anderson S[4]

Archibald A[4]

Hulme H[3]

Kemp SJ[1]

[1]School of Biological Sciences
University of Liverpool
Crown Street
Liverpool
L69 7ZB

[2]International Livestock Research Institute (ILRI)
P O Box 30709
Nairobi 00100
Kenya

[3]Department of Computer Science
University of Manchester
Oxford Road
Manchester
M13 9PL

[4]Roslin Institute
Roslin BioCentre
Midlothian
EH25 9PS


## Introduction

African cattle breeds differ significantly in their ability to survive low to moderate levels of challenge with *Trypanosoma congolense*. Similarly the survival times of inbred mouse strains vary substantially after infection. We have previously identified three regions of the mouse genome that regulate survival after infection in two crosses (A/J x C57BL/6 and Balb/c x C57BL/6) (Kemp et al 1997 Nature Genetics). These have been designated *Tir1*, *Tir2* and *Tir3* for Trypanosoma infection response loci on mouse chromosomes 17, 5 and 1 respectively. We have now used two strategies to reduce the size of the regions that appear to be regulating survival. Firstly, congenic mice lines carrying defined regions of the C57BL/6 genome on an A/J background were developed to identify physical boundaries of the regions and confirm its effect. Secondly the response to infection has been mapped in an additional mouse strain (129J). The mapping data has been combined with haplotype maps to identify a 70kb high priority region containing just 5 strong candidates for the causative gene for resistance to trypanosomiasis in mice on chromosome 17.

## Congenic Mice

Congenic lines are created by crossing resistant C57BL/6 with the susceptible A/J mice. At each generation the offspring are genotyped to identify those animals that are carrying alleles from C57BL/6 in the target region of interest and these mice are then selected to be backcrossed to the recipient genome. After seven generations of backcrossing to A/J, heterozygous carriers of the C57BL/6 donor region of interest were intercrossed and homozygous carriers were selected as the congenic line and were designated TirnCC in this study.
Homozygous carriers of the A/J alleles were selected as the control line and are designated TirnAA. The creation of these mouse lines makes it possible to study the effect of each locus in isolation from the other loci and most background effects.
The positions of the regions of the C57BL/6 genome that were introgressed into the A/J background were determined by genotyping the three lines with the Illumina 1536 SNP marker panel (Figure 1).
The introgressed regions on chromosome 5 and 17 had a significant effect on survival (Figure 2) but the C57BL/6 region on chromosome 1 had no effect on survival indicating that the genes regulating the response to infection are elsewhere on this chromosome.

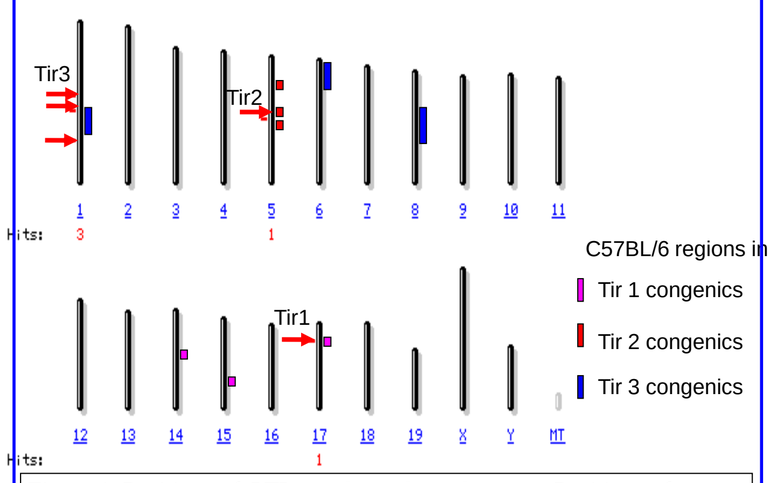



Figure 1. Positions of QTL are shown by red arrows. Positions of regions of C57BL/6 genome introgressed into the A/J genome are shown by the coloured blocks

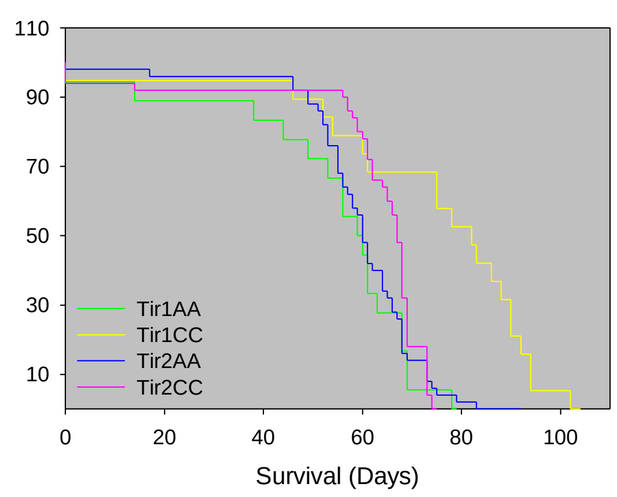
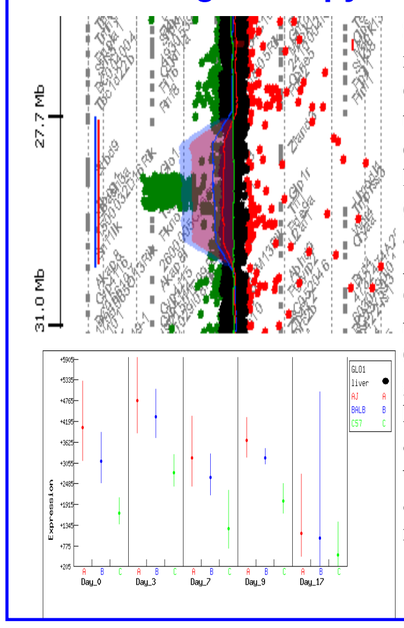



Figure 2. Survival times. The plots show that the Tir1CC mice carrying C57BL/6 DNA at the chromosome 17 locus survived longer than the Tir1AA mice carrying A/J DNA at this locus. Also the Tir2CC mice carrying C57BL/6 DNA at the chromosome 5 locus survived longer than the Tir2AA mice carrying A/J DNA at this locus. There was no difference in survival of the Tir3CC and Tir3AA mice (not shown) indicating that the genes carrying resistance alleles on chromosome 1 are outside this region.

## Effect of gene copy number on expression




Single nucleotide poly-morphsims are not the only form of genetic variation. Regions of the mouse genome can be duplicated or deleted and are known as copy number variants (CNV). CNV between C57BL/6 and A/J mice were detected using Agilent 240k whole genome CGH arrays. An amplifcation of the Glo1 gene within the Tir1 QTL on chromosome 17. This gene was not identified by the haplotype analysis but expression analysis indicated that the gene was more highly expressed in A/J and Balb/c consistent with copy number.

## Mapping loci controlling the response to infection in 129J mice

Identifying the regions controlling the response to infection in additional mouse strains make it possible refine the list of candidate genes that might regulate survival time.
F2 C57BL/6 x 129J were bred by crossing 129J and C57BL/6 mice to create and F1 generation and then by intercrossing their offspring to create an F2.
135 F2 C57BL/6 x 129J mice were genotyped with the Illumina 384 SNP mapping panel and the data was analysed with the JQTL package (Figure 3).
The data confirmed the presence of QTL on chromosomes 1 and 17 but there was no evidence of a QTL on chromosome 5 suggesting that 129J might carry the C57BL/6 allele at this locus.
The locus on chromosome 17 was significantly associated with survival. The locus on distal chromosome 1 was not significant after correction for multiple testing, but since it was supported by multiple SNP markers and it coincides with previously identified loci in A/J and Balb/c mice, it may be real but the effect size of the alleles may be insufficient to show a significant effect in this relatively small panel of mice.

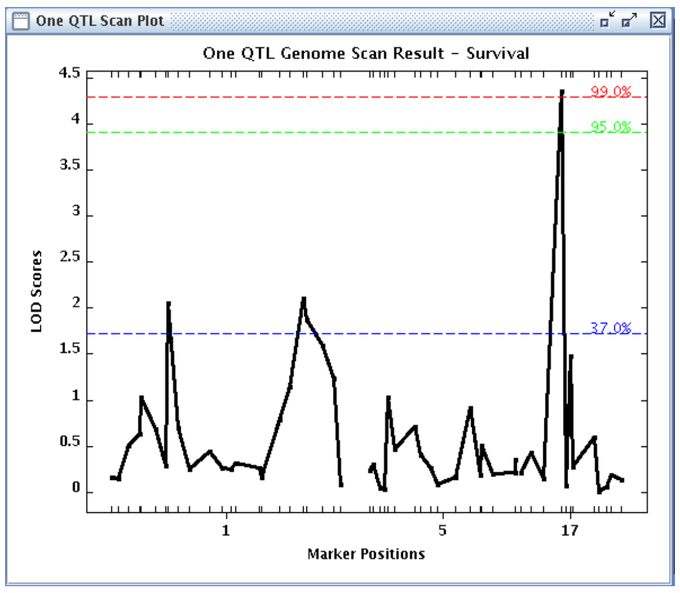



Figure 3. Plot of LOD scores for markers on chromosomes 1, 5 and 17. Showing evidence of QTL on chromosomes 1 and 17 but not on chromosome 5.

## Haplotype analysis

The availability of mapping data from three pairs of mouse strains makes it possible to look for correlation between haplotypes across the QTL regions and response to infection. The genome of inbred mouse strains is believed to be composed of a mosaic of regions derived from three or four ancestral strains. Given that we have observed the some QTL in multiple pairs of mouse strains it is likely that the polymorphisms that make C57BL/6 more resistant are derived from one of those ancestral strains. If the ancestral strain from which gene is inherited can be identified then it should be possible to produce a short list of candidate genes for which C57BL/6 inherits its allele from one strain and A/J, Balb/c and 129J inherit their alleles from a different strain.

The recent publication of 8 million SNP in the fifteen most widely used mouse strains makes it possible to identify the haplotype of origin of most regions of the mouse genome (Frazer et al. 2007 Nature). http://mouse.perlegen.com/mouse/mousehap.html
We have used this data to assign each gene to an ancestral haplotype (Figure 4). Only five genes were on the same ancestral haplotype in 129J, A/J and Balb/c and on a different haplotype in C57BL/6. These genes are therefore strong candidates for the gene that regulates the difference in survival after infection, Quantative Trait Genes (QTG). An additional 19 genes were on different haplotypes in 129J, A/J and Balb/c but also all different from C57BL/6. These genes are also possible candidate QTG.

It is as important to exclude genes as include them and the addition of data from the 129J x C57BL/6 cross makes it possible to exclude the class 1 and 2 MHC genes. Since the chromosome 17 QTL overlaps the the MHC region these were natural candidate genes. 129J and C57BL/6 share the same MHC haplotype (b) as can be seen in figure 4. Consequently it seems unlikely that the classical MHC genes are responsible for the large difference in survival time associated with this locus.

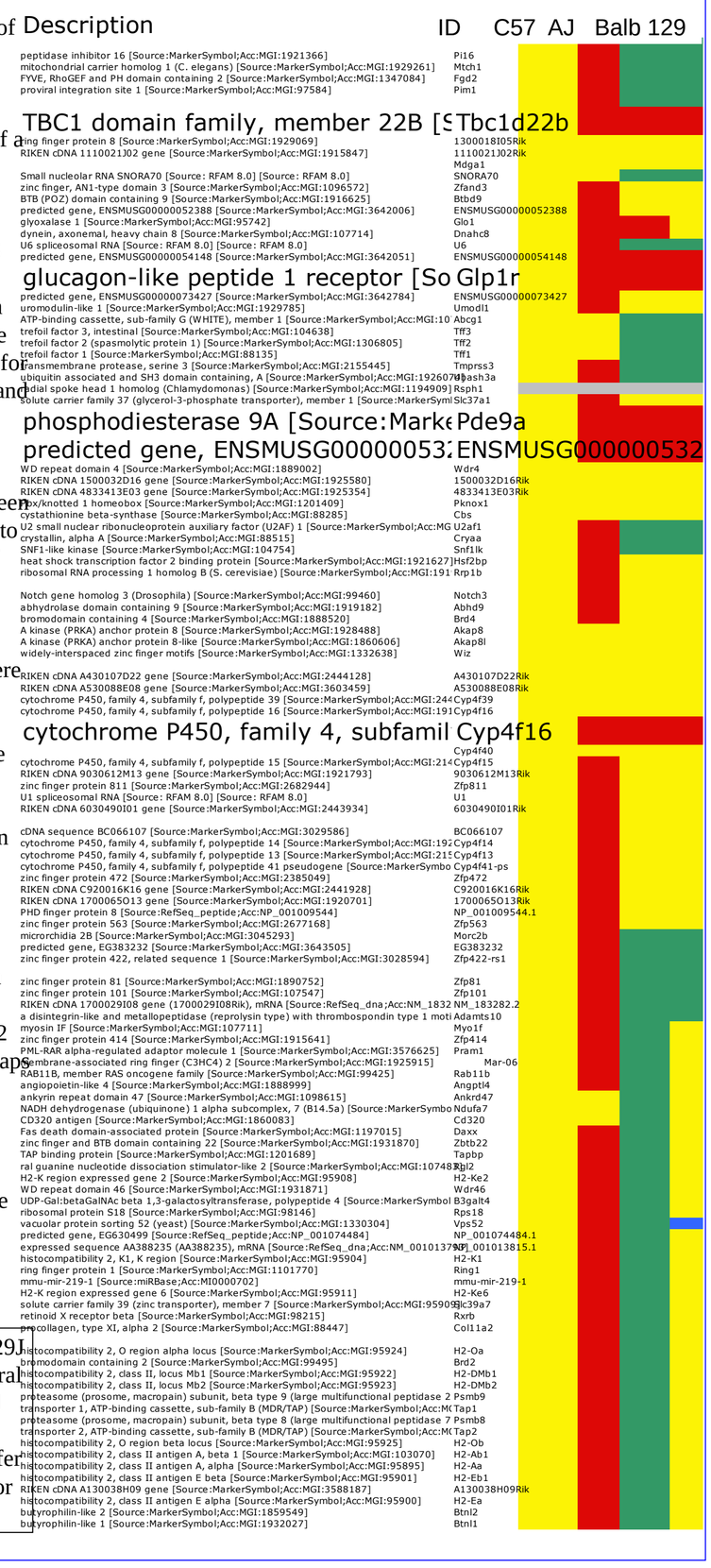



Figure 4. Haplotypes of C57BL/6, AJ, Balb/c and 129J across the Tir1 QTL on chromosome 17. The ancestral haplotype of each strain is indicated by the coloured blocks to the right of the gene names. C57BL/6 is always shown in yellow. Where the other strains differ in ancestral haplotype they are shown in red, green or blue.

## Conclusions

The combination of congenic mice, additional mapping data and discovery of copy number variations within the QTL regions has made it possible to identify a short list of genes that might regulate the response to infection with *T. congolense*. This list is now sufficiently short that it is practicable to undertake detailed studies on the role of individual genes in the response to infection and hence determine whether they cause the difference in survival time after infection. The identification of these genes is expected to give an insight into the pathways that regulate the response to infection and may lead to new approaches to treatment.


Acknowledgements: We thank John Wambugu, Moses Ogugo and Leanne Wardlesworth for technical assistance. These studies were funded by the Wellcome Trust