Next Generation Sequencing – The Role of New Sequence Technologies in Shaping the Future of Veterinary Science

Hosted by the RCVS Charitable Trust

RCVS charitable trust

UNIVERSITY OF LIVERPOOL

Sequencing pathogens (and their hosts): from large to small, from virulent to benign

THE ROYAL SOCIETY

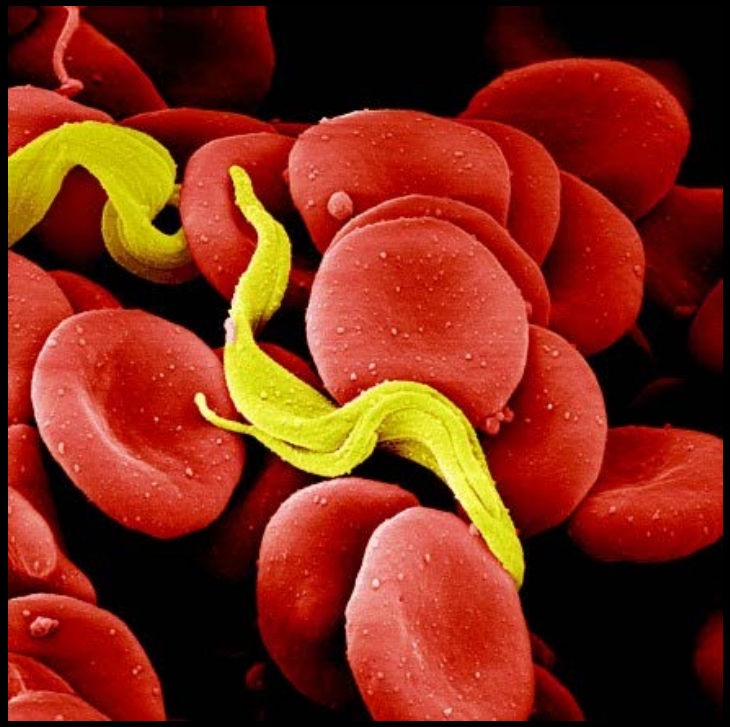
wellcome trust

LSTM

CENTRE for GENOMIC RESEARCH CGR

Trypanosomiasis

- **“Sleeping Sickness” in humans**
  - 500,000 affected annually
  - 37 countries; 22 of which least developed in the world
  - Fatal if untreated
- **“Nagana” in cattle**
  - Roughly 1/3 of African cattle population ‘at risk’
  - Causes cachexia, anaemia and increased susceptibility to secondary disease
  - Losses in livestock and crop production amount to an estimated $5 billion annually



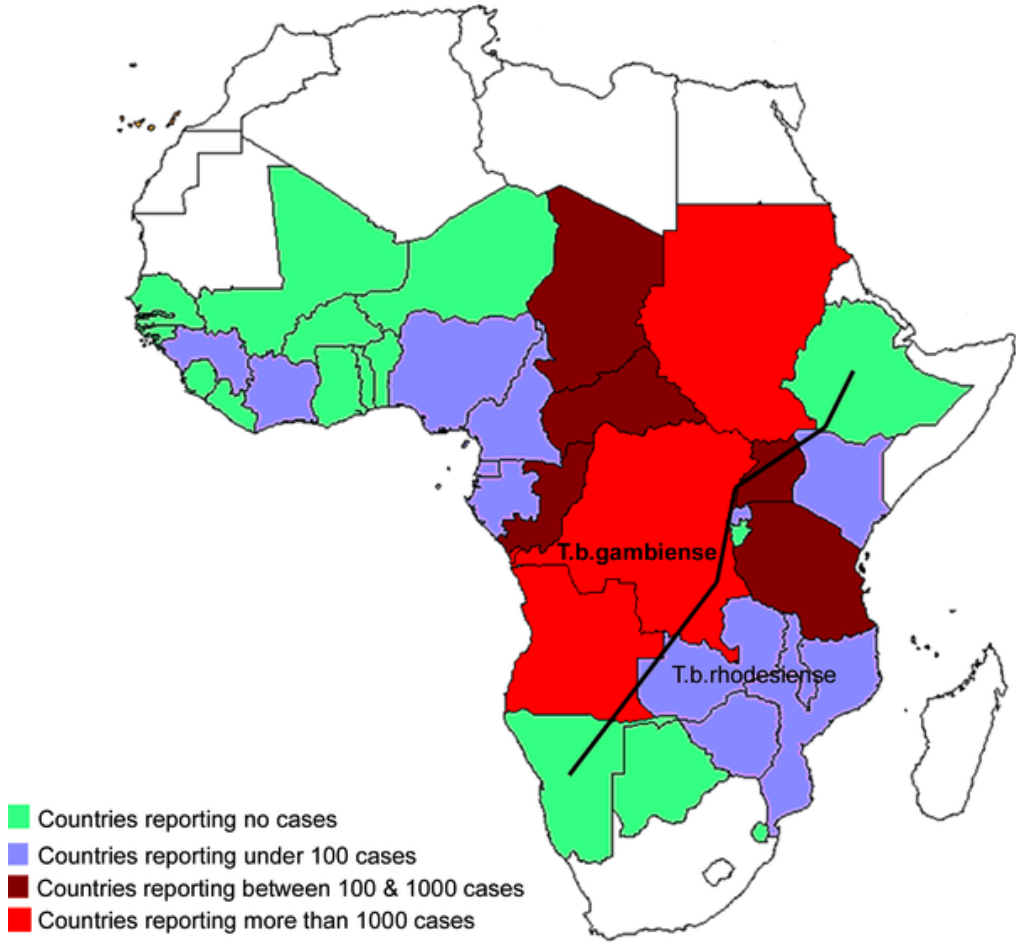
Trypanosomiasis: "Sleeping Sickness" – Human Disease

Affects 500,000 annually across 36 countries
Caused by two species of *T. brucei* parasite:

T. brucei gambiense
- Western Africa
- Chronic disease

T. brucei rhodesiense
- Eastern Africa
- Acute Disease

T. brucei brucei
- widespread
- disease of cattle and other non-human vertebrates

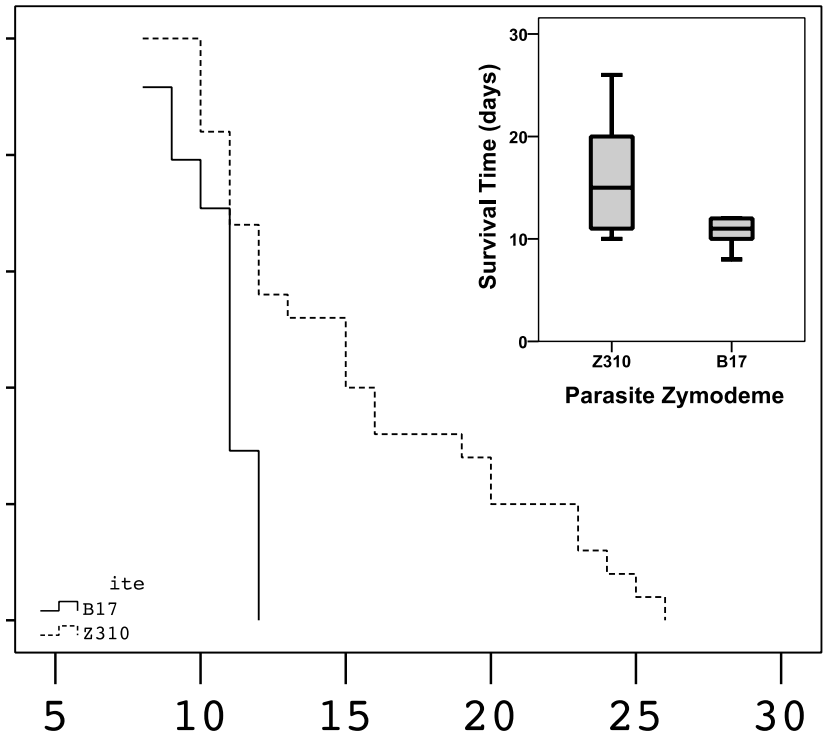


Established Difference in Pathogenicity between strains

- Isolates of *T. b. rhodesiense* from South-Eastern Uganda
- Collected 1989-1992
- Patients reporting showed two broad clinical histories
- Virulence replicated in mice

Course of infection	**Acute** **Mean 4 wks**	**Chronic** **Mean 20 wks**
Stage of disease at first presentation	Early	Late
Chancre	92% of cases	None recalled
Bloodstream Parasitaemia	+++	-
Trypanosomes in CSF	-	+
WBC in CSF	< 5 cells / mm^3	> 5 cells / mm^3
Zymodeme	***Busoga*** **(predominantly *B17)***	***Zambesi*** **(predominantly Z310)**

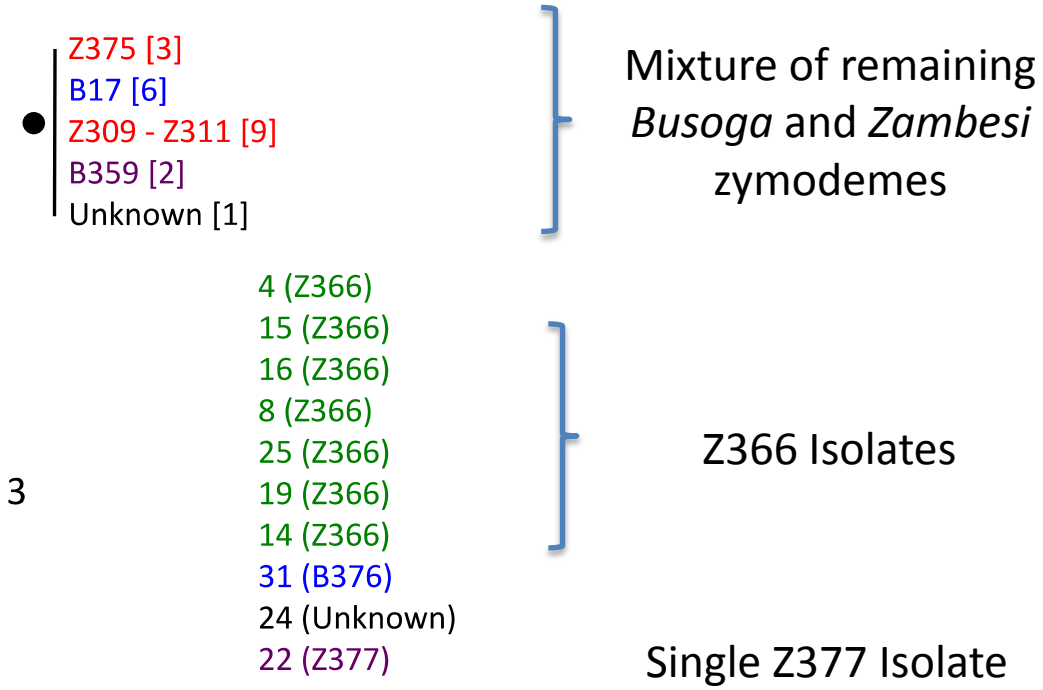
Established Difference in Pathogenicity between strains



- Survival in very susceptible A/J mice corresponds to virulence in humans

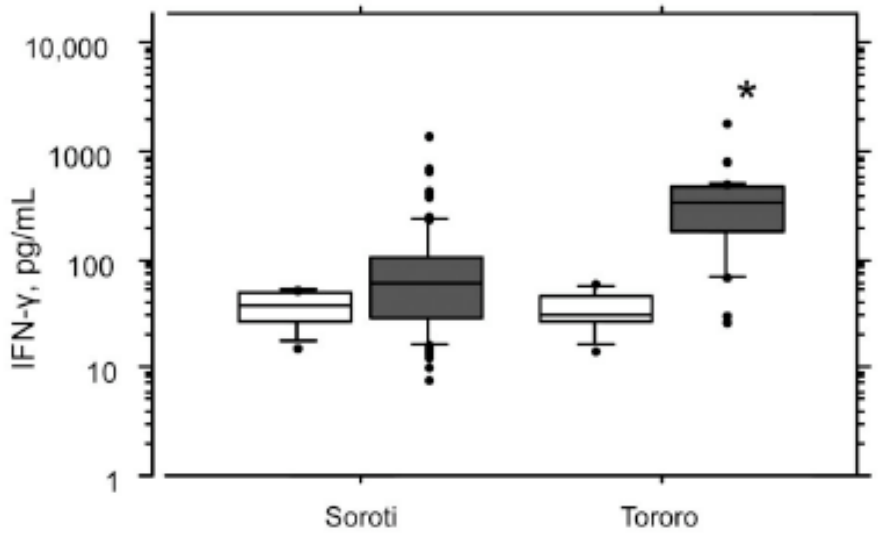
Multilocus genotyping fails to distinguish Z310 and B17 zymodeme isolates

- 31 isolates genotyped at 11 informative microsatellite loci
- Difficult to distinguish Zambesi 310 and Busoga 17 populations
- Strain with intermediate phenotype – Z366 - *is* distinguishable using the same markers
  - All Z366 isolates are from a specific 4-year outbreak in Bugiri region in 1993

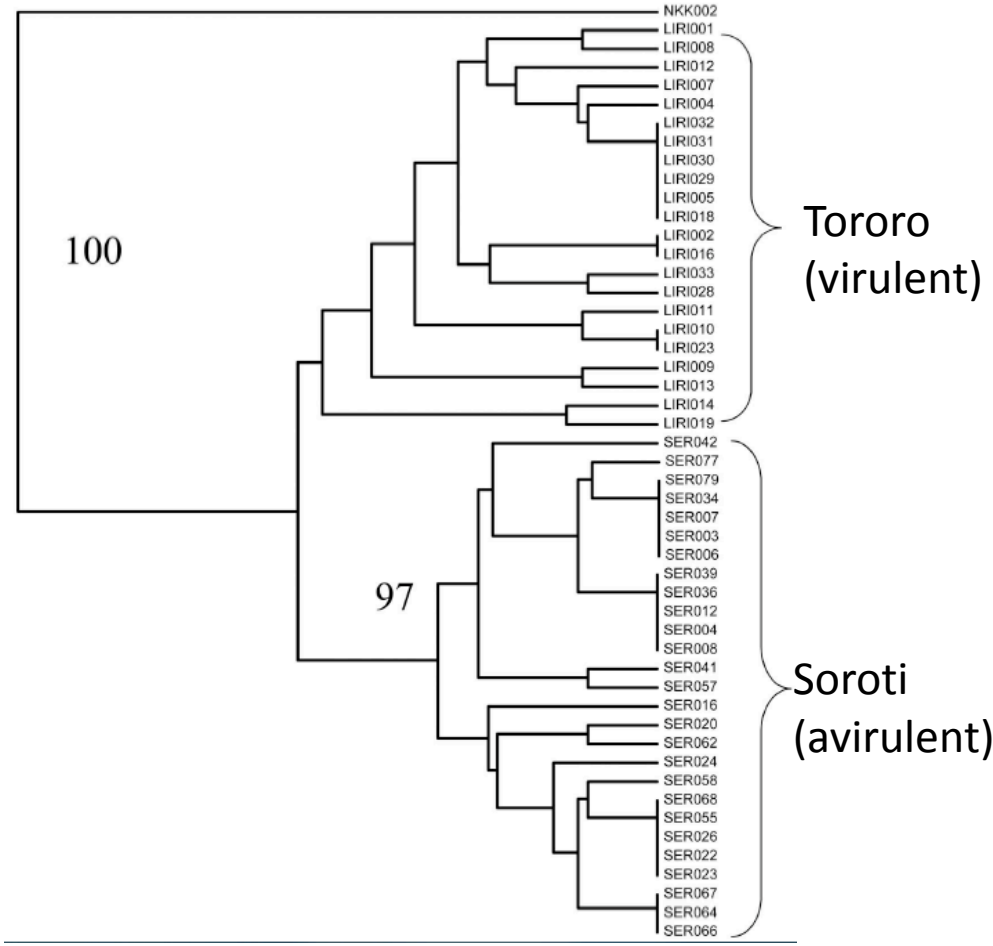


Bootstrapped Dendrogram
(UPGMA; Jaccard's similiarity index)

Spatially and Genetically Distinct African Trypanosome Virulence Variants Defined by Host Interferon-g Response (2002-2004 outbreak)



Human plasma levels of IFN-gamma



Dendrogram showing similarities between trypanosome multilocus microsatellite genotypes.

Maclean, L *et al*. (2007) **The Journal of Infectious Diseases 2007; 196:1620–8**

Hypothesis

- Established clear differences in phenotype in both man and in experimental infections in mice
- Microsatellite genotyping shows that Z310 and B17 are indistinguishable using the eleven markers tested
- Can “next-generation” DNA sequencing provide some clues as to what differences there are between them?

SOLiD Sequencing: Single Nucleotide Polymorphism (SNP) discovery

- Sequenced one isolate each of *Z310* and *B17*
  - Life Technologies SOLiD (v3) platform
  - Single run generated >60X of both genomes!
- Reads aligned against *T. b. brucei* reference genome

- Genomes >97% similar

  Z310 (less virulent)

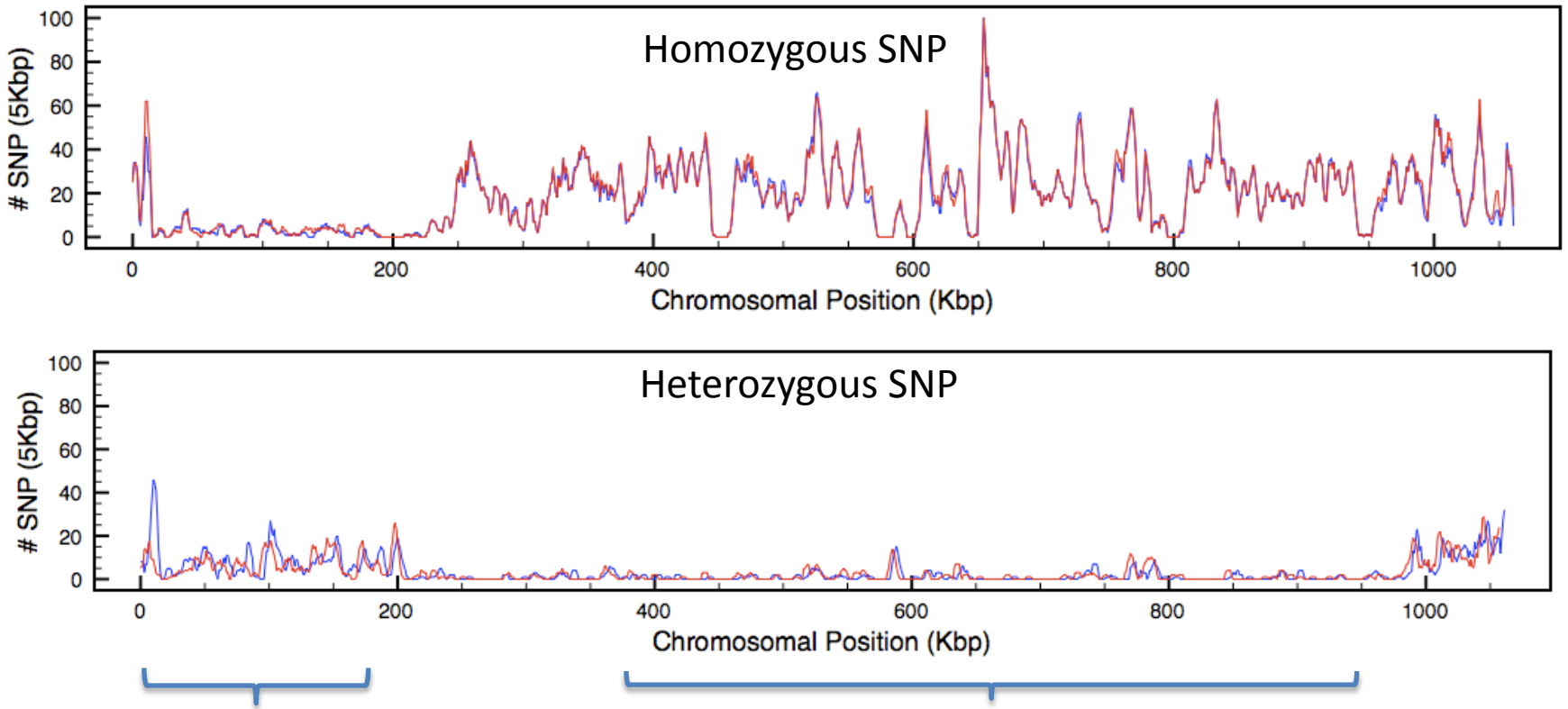
  - 13,716 polymorphisms

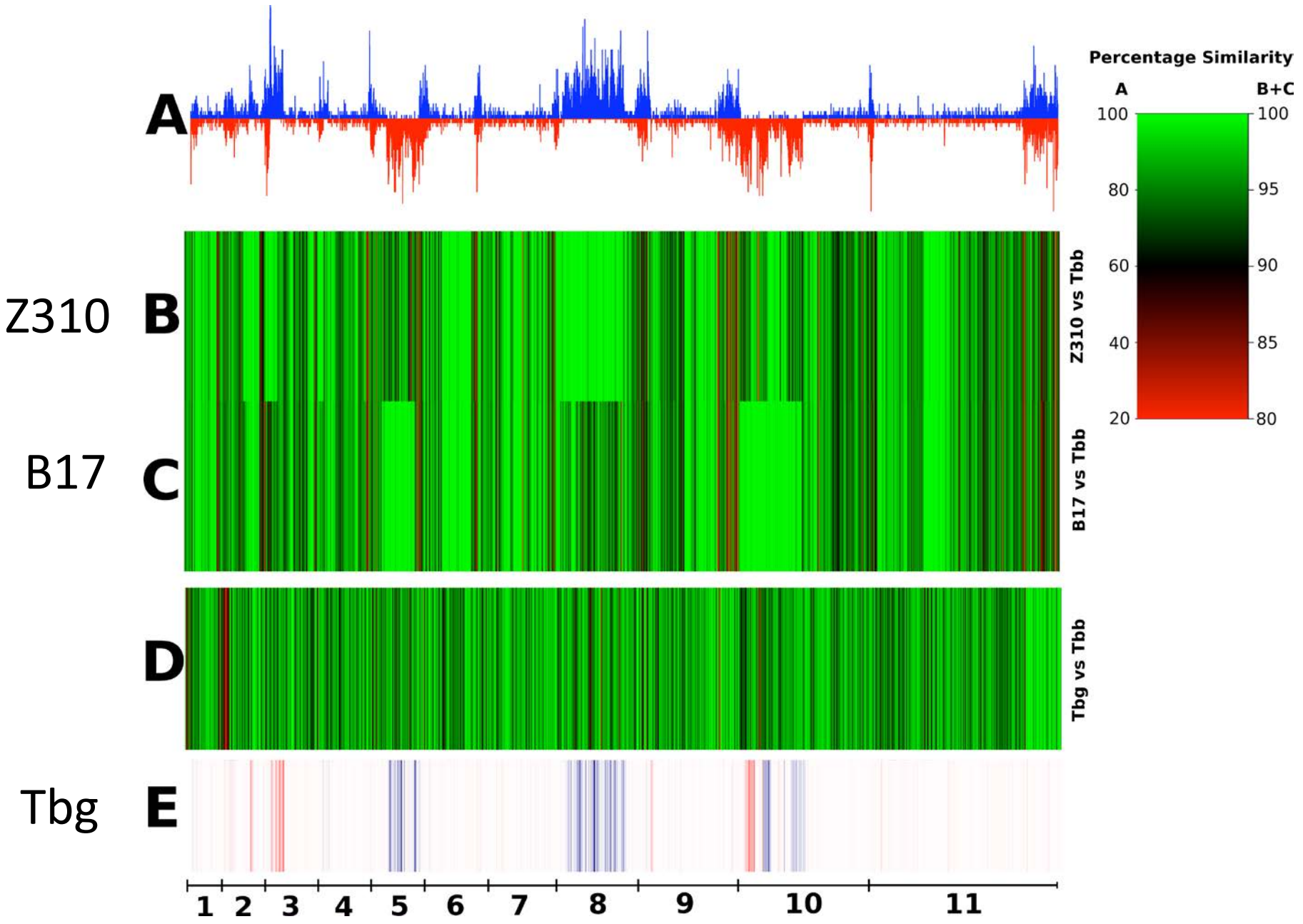
  B17 (more virulent)

  - 18,420 polymorphisms

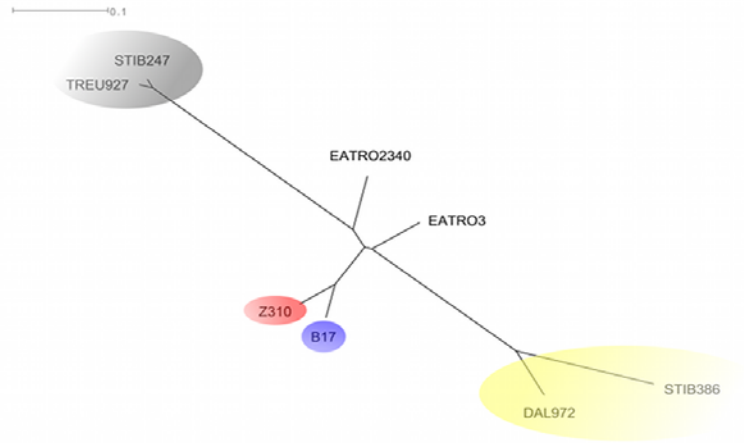
SOLiD sequencing reveals distinct patterns of SNP distribution

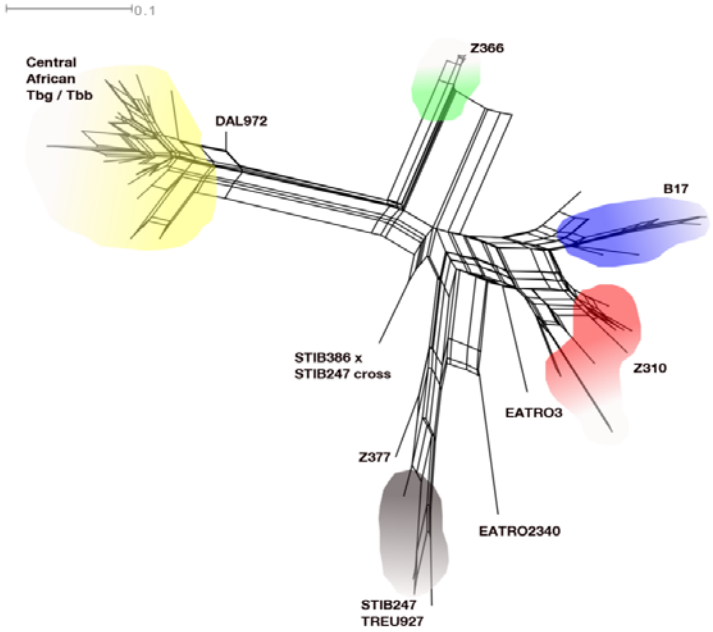
- Plot of 5Kbp 'moving average' of raw SNP (vs. *T. b. brucei* sequence) across each chromosome revealed 'regions' of loss of heterozygosity



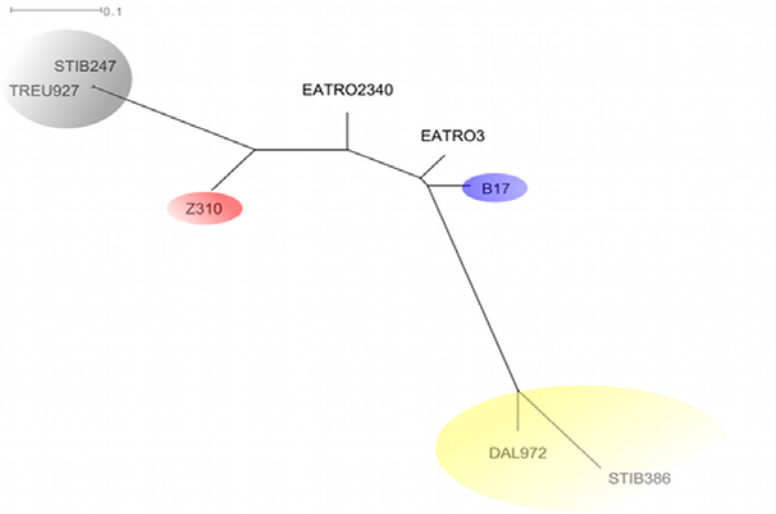

A
B
C
D
E
Z310
B17
Tbg
Percentage Similarity
A
B+C
100
80
60
40
20
100
95
90
85
80
Z310 vs Tbb
B17 vs Tbb
Tbg vs Tbb
1
2
3
4
5
6
7
8
9
10
11

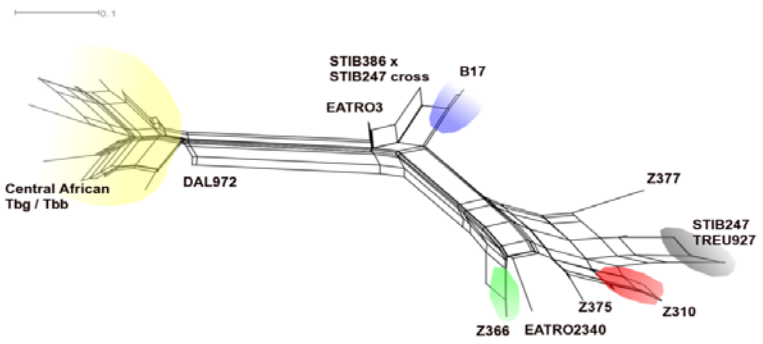
Whole Genome





Chromosome 8





Virluent strain (B17)

Less-virulent strain (Z310)

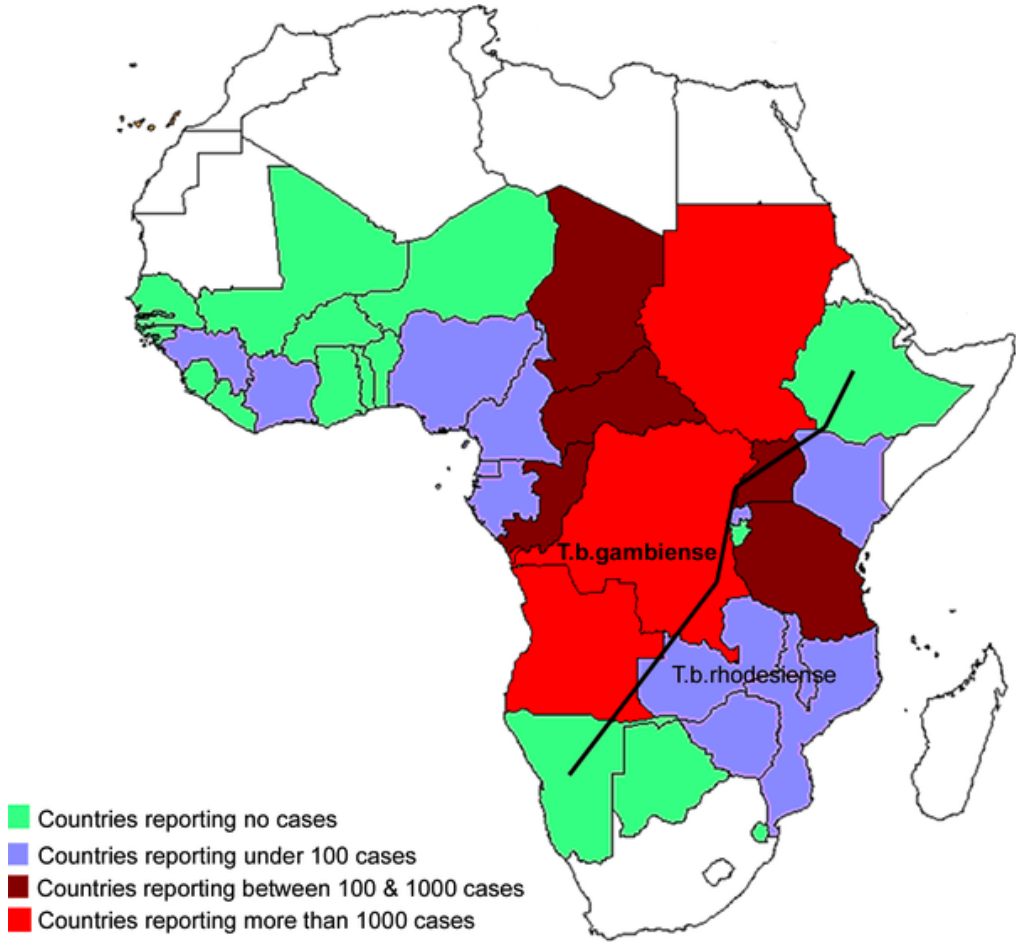
Trypanosomiasis: "Sleeping Sickness" – Human Disease

Affects 500,000 annually across 36 countries
Caused by two species of *T. brucei* parasite:

T. brucei gambiense
- Western Africa
- Chronic disease

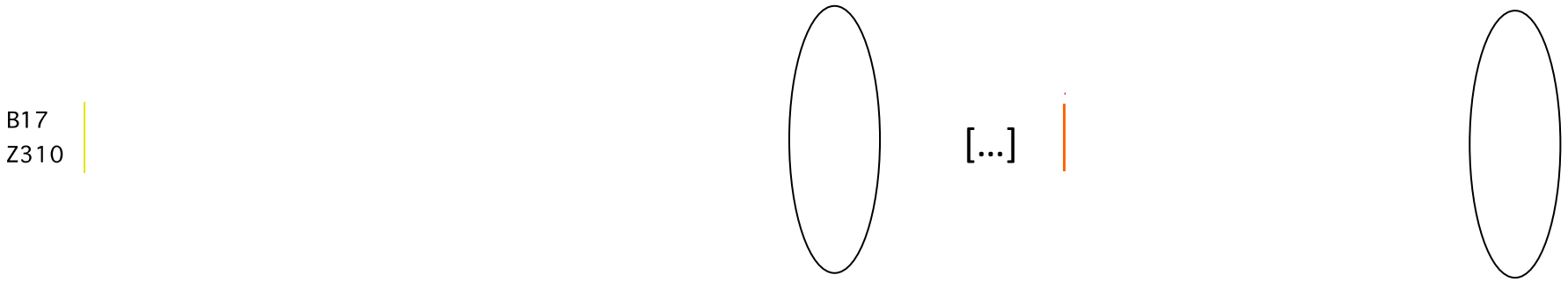
T. brucei rhodesiense
- Eastern Africa
- Acute Disease

T. brucei brucei
- widespread
- disease of cattle and other non-human vertebrates



Identification of possible marker for differences in virulence

- Isocitrate dehydrogenase is the only MLEE enzyme that distinguishes *Zambesi* from *Busoga* isolates
- Three heterozygous non-synonymous SNP predicted within Chromosome 8 copy of ICD gene in SOLiD sequenced *B17* isolate
  - Confirmed by Sanger sequencing on PCR on bloodstream stabilates
  - Found to correlate with zymodeme as per MLEE pattern
- Actually two copies of ICD, None found in chromosome 11 copy of ICD



Alignment of amino acid sequence of the beginning (left) and end (right) of the chromosome 8 copy of Isocitrate dehydrogenase gene

Conclusions

- Strains of *T. b. rhodesiense* have a reproducible difference in virulence phenotype
- Microsatellite data suggests either isolates are very closely related, and/or not sensitive enough to detect variation
- Next generation sequence data reveal high level of genomic similarity between strains (>97%)
- Identified discrete regions of heterozygosity which are different between the two strains that isn't detected by other methods
- Polymorphism(s) on ~80% of chromosome 8 differ between strains may help regulate differences in virulence between *Busoga* and *Zambesi* strains

Trypanosomiasis: "Nagana" – Animal Disease

- Caused largely by *T. congolense* and *T. vivax*
- Symptoms in domestic animals include fever, muscle wastage and anaemia.
- Affects over ten million Km^2
- 30% African cattle are at risk of infection
- Losses in livestock and crop production amount to more than ~$1 billion per annum

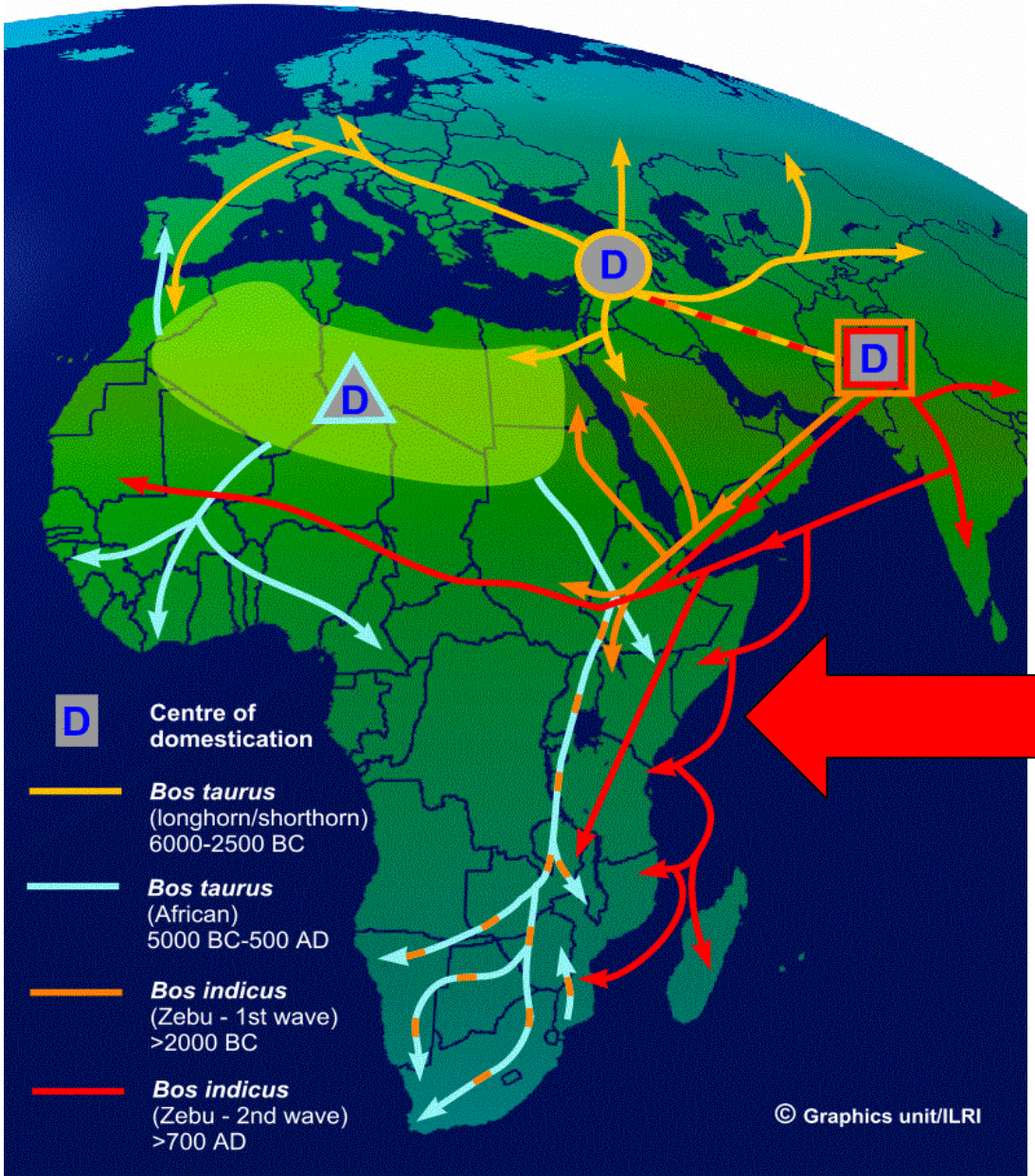
Host Genetics: Trypanotolerance

Resistant: N'dama (*Bos taurus*)

Susceptible: Boran (*Bos indicus*)

Origin and migration routes of domestic cattle in Africa

N'dama (resistant)

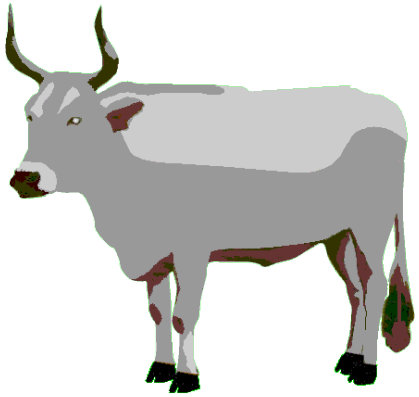


Boran (susceptible)

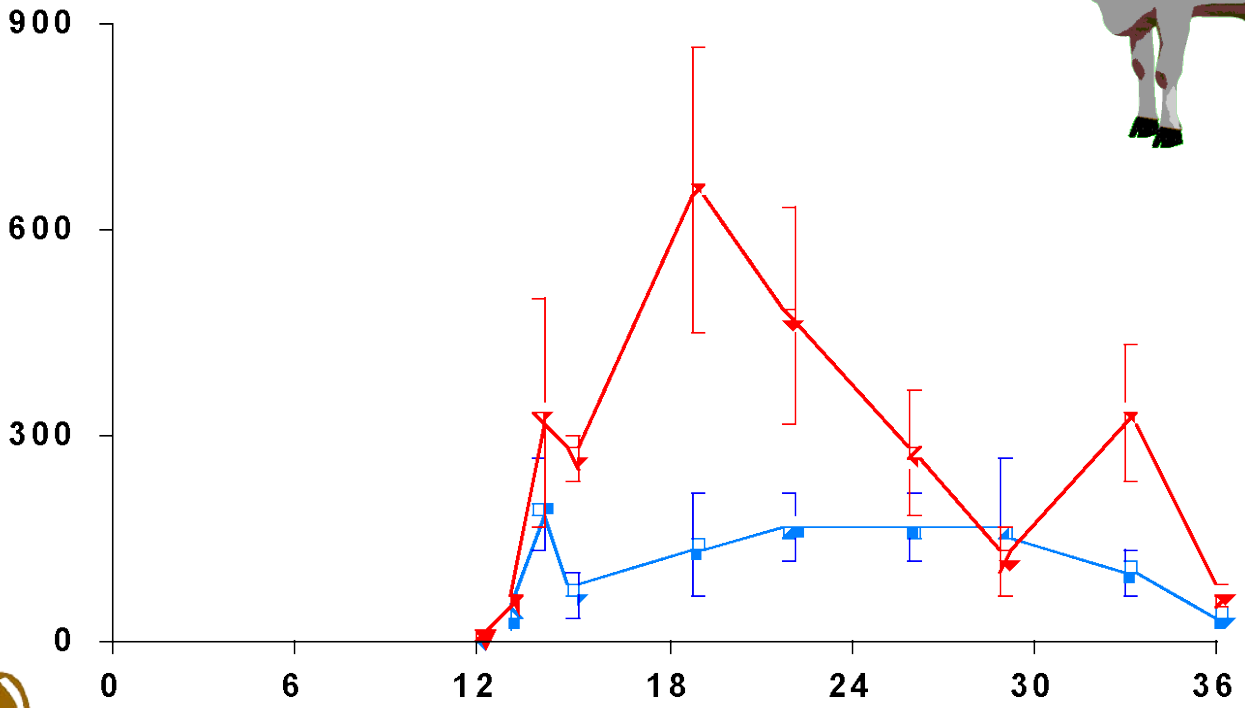
ILRI
INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

Parasitaemia

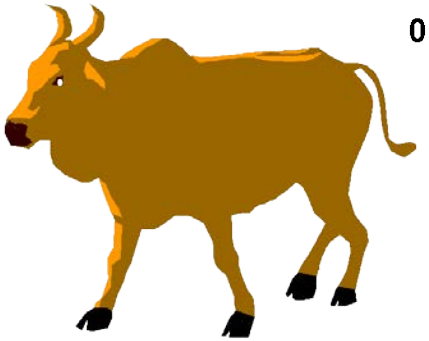
N'Dama

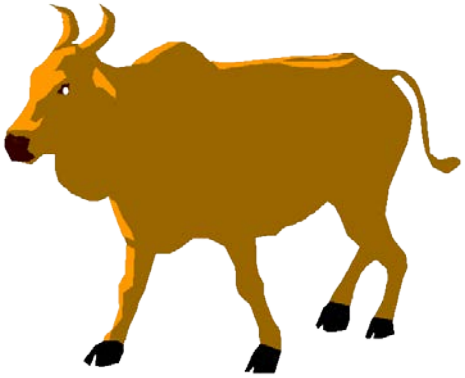


1st wave parasitemia



Boran

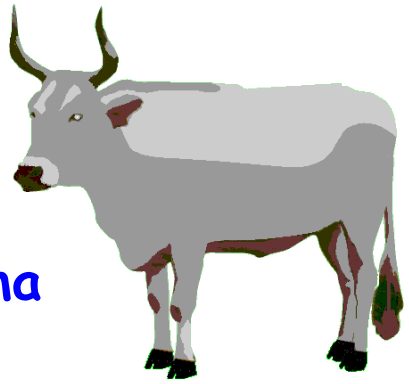




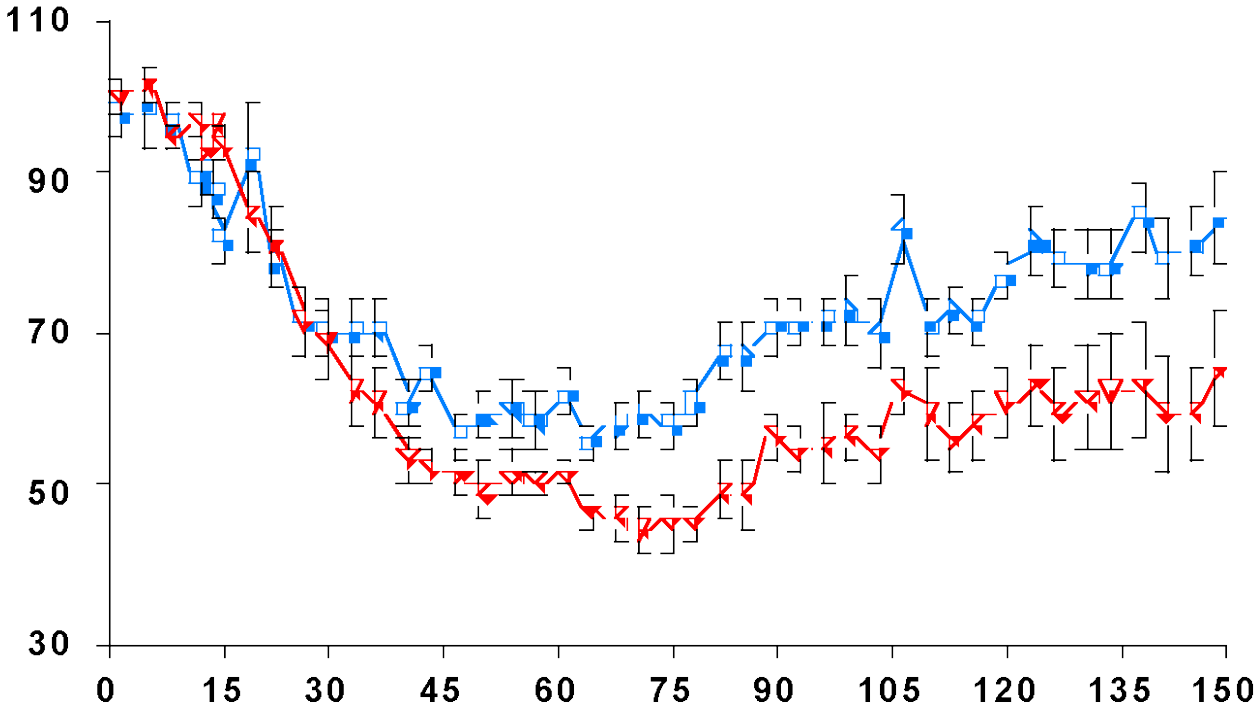
Anaemia (PCV)

Boran

N'Dama



% change in PCV



days after infection

Host Genetics: The mouse model

N'dama
(*Bos taurus*)

Boran
(*Bos indicus*)

Resistant:

C57BL/6 (Black Six) mice

Susceptible:

A/J and BALB/c mice

Tolerance to *T. congolense* infection: Mapping in Inbred Mice

- Tolerance in the mouse model has a major genetic component

- Mapping in genetic crosses has revealed three regions that control resistance / susceptibility

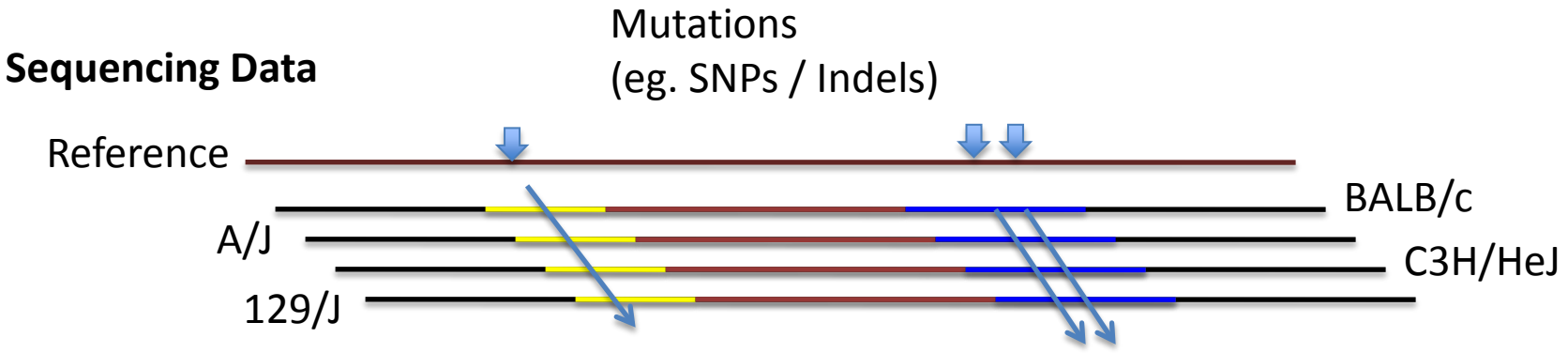
Quantitative trait loci (QTL)

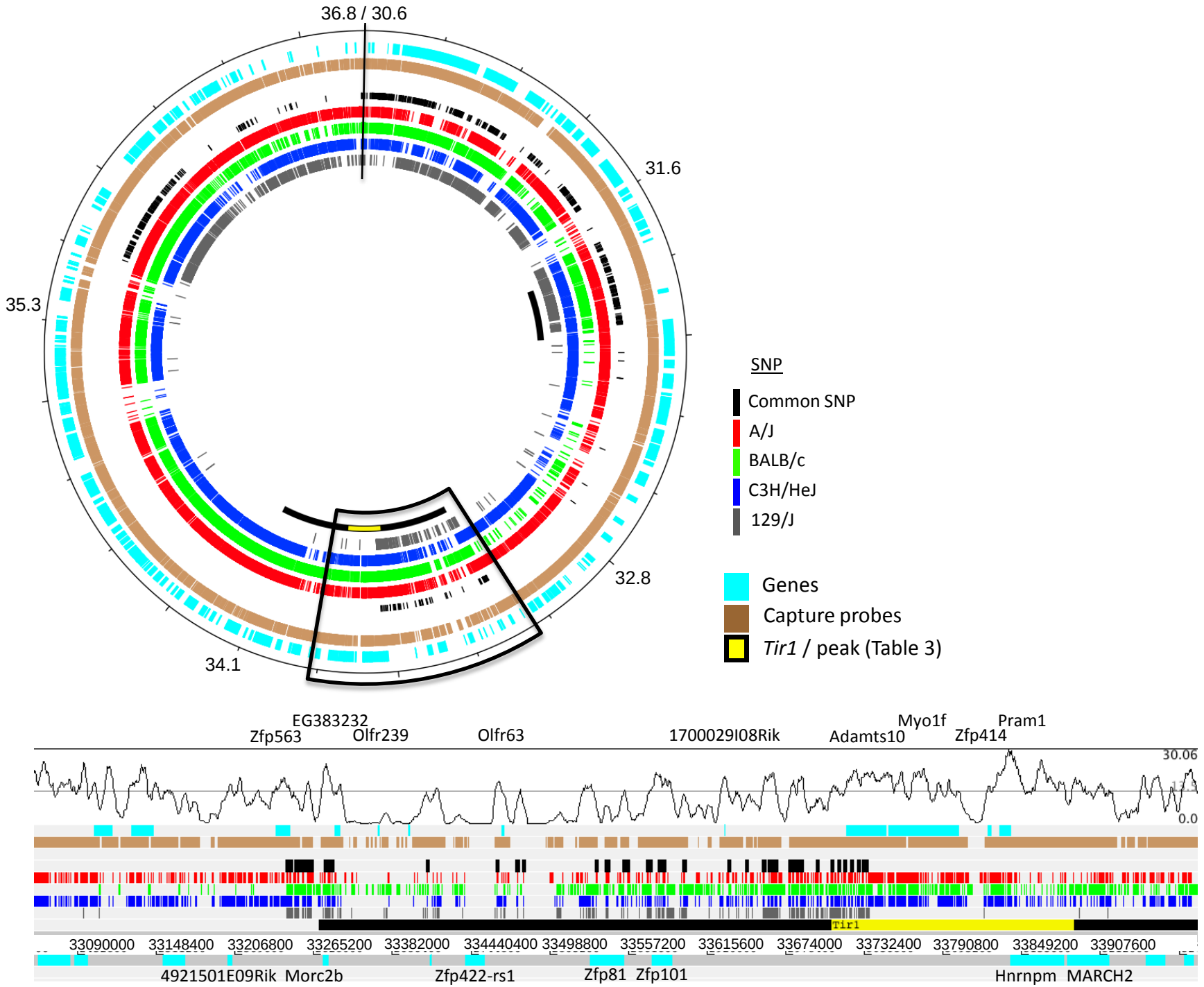
- Chromosome 17 – *Tir1*
- Chromosome 5 – *Tir2*
- Chromosome 1 – *Tir3a-c*

- *Problem – these regions contain hundreds of genes that may be controlling the response to infection*

Sequence Capture of Inbred Mice

- "Capture" region of interest (i.e. *Tir1*)
- Resequence 4 susceptible breeds of mice
- Identify "shared" polymorphisms

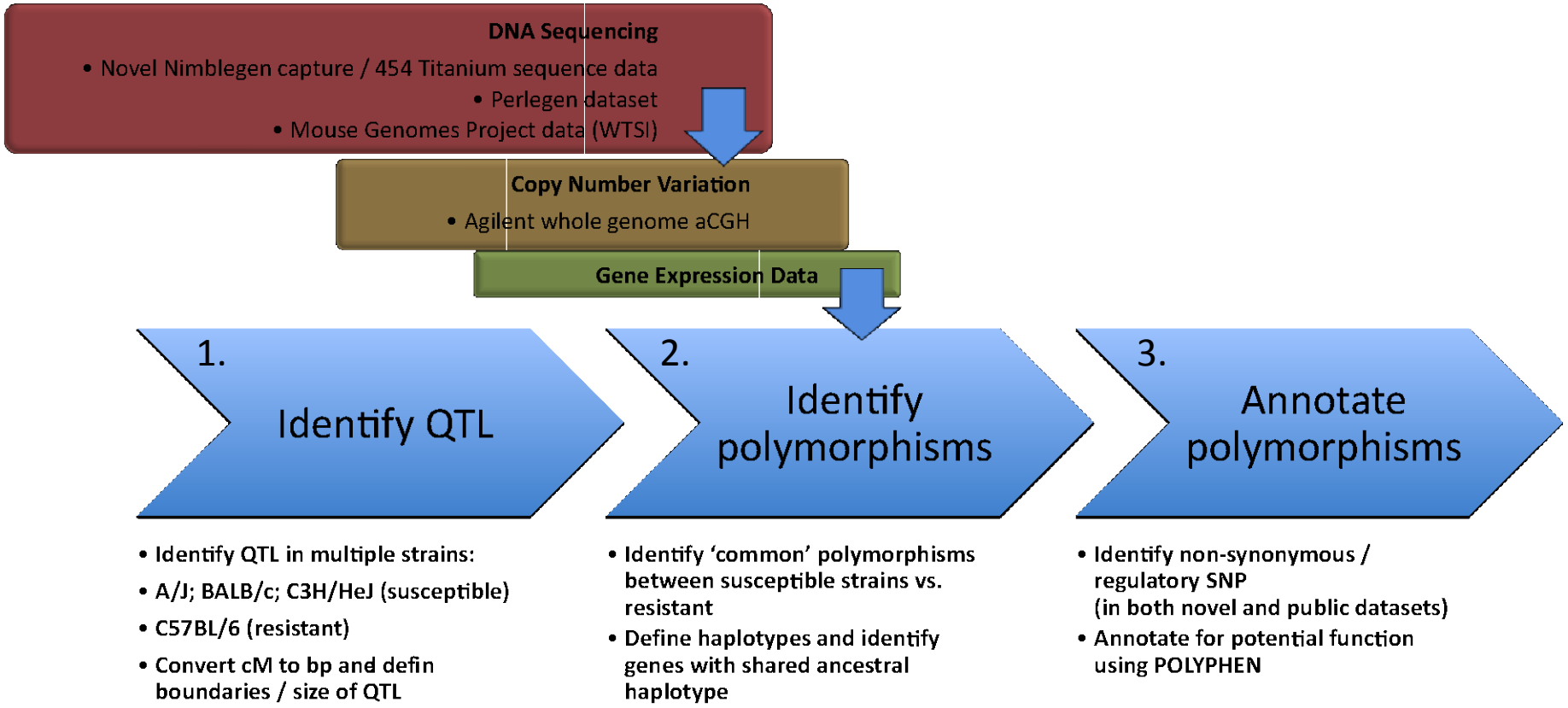



36.8 / 30.6
31.6
32.8
34.1
35.3
SNP
Common SNP
A/J
BALB/c
C3H/HeJ
129/J
Genes
Capture probes
Tir1 / peak (Table 3)
EG383232
Zfp563
Olfr239
Olfr63
1700029I08Rik
Myo1f
Adamts10
Zfp414
Pram1
30.06
0.0
Tir1
33090000
33148400
33206800
33265200
33382000
33444040
33498800
33557200
33615600
33674000
33732400
33790800
33849200
33907600
4921501E09Rik
Morc2b
Zfp422-rs1
Zfp81
Zfp101
Hnrnpm
MARCH2

Targeted Resequencing of *Tir1*:

- Identified one candidate gene with two functional SNP
- One SNP is predicted to affect the function of the protein
  - *Pram1* - PML-retinoic acid receptor alpha regulated adaptor molecule 1
  - Predicted to have an effect on the early- pro-inflammatory response after infection

Utilising Genomic Technologies to reduce the number of candidate genes affecting Trypanotolerance in mice



Trypanotolerance: Number of “candidate genes” under QTL

QTL	Before	After
Tir1	43	27
Tir2	210	14 – 74
Tir3a-c	20 - 650	10 – 355*

* Candidate region revealed later in pipeline

Trypanotolerance Summary:

- Attempted to characterise every Single Nucleotide Polymorphism in a 4.2Mbp region of 4 breeds of inbred mouse.
- Systematically reduced number of candidate genes at *Tir1* from 73 to just 7.

- Functional testing of identified pathways in cattle are now feasible (and underway)

Summary (2)

- Early sequencing work established ‘reference’ genomes on which modern NGS techniques are now based.
- Cheaper, quicker and easier* to produce large numbers of whole genome sequences rapidly to compare
  - Pathogen
  - Host

Acknowledgements

- Neil Hall
- Harry Noyes
- J. Wendi Bailey
- Annette MacLeod
- Paul Capewell
- Tanja Beament
- Mark Taylor
- Michael Chance
- Steve Kemp
- Alan Archibald
- Fuad Iraqi
- Peris Amwayi
- Kevin Ashelford
- Margaret Hughes
- Suzanne Kay
- Jason Dowling
- Jo Sanders
- Pam Pask
- Stephen Johnson
- Wellcome Trust Sanger Institute

Next Generation Sequencing – The Role of New Sequence Technologies in Shaping the Future of Veterinary Science

Hosted by the RCVS Charitable Trust

RCVS charitable trust