The sequence of a male-specific genome region containing the sex determination switch in *Aedes aegypti*

Joe Turner [1,2] *, Ritesh Krishna [1], Arjen E. van 't Hof [1], Elizabeth R. Sutton [2,3], Kelly Matzen [2], Alistair C. Darby [1]

1. Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool
2. Oxitec Ltd., Oxford
3. Department of Zoology, University of Oxford

*joe.turner2@liverpool.ac.uk **@joe_turner177**

Summary

- The Y chromosome-like sex determining region of the *Aedes aegypti* genome (the **M locus**) is significant both as a case study of the evolution of sex chromosomes, and because of its applicability to disease control programmes
- The male-determining M locus gene *Nix* contains an **extraordinarily long intron**
- The *Nix* intron is **primarily composed of repetitive DNA**
- The M locus, despite its small size, shows features in common with large male-specific regions of the Y-chromosomes of other species

Introduction

- >2.5 billion people are at risk from diseases vectored by *Ae. aegypti.*
- Future mosquito control strategies may include the **release of sterile or transgenic "self-limiting" mosquitoes** (Alphey, 2014), which could be improved through **manipulation of sex determination** (Adelman & Tu, 2016).
- In 2015 an **M locus gene named *Nix*** was identified, which is expressed in the early embryo and acts as the upstream male-determining switch (Hall *et al,* 2015).
- So far only the cDNA of *Nix* has been characterised. In this study we aimed to reveal its **full gene structure** and genomic context.

Methods

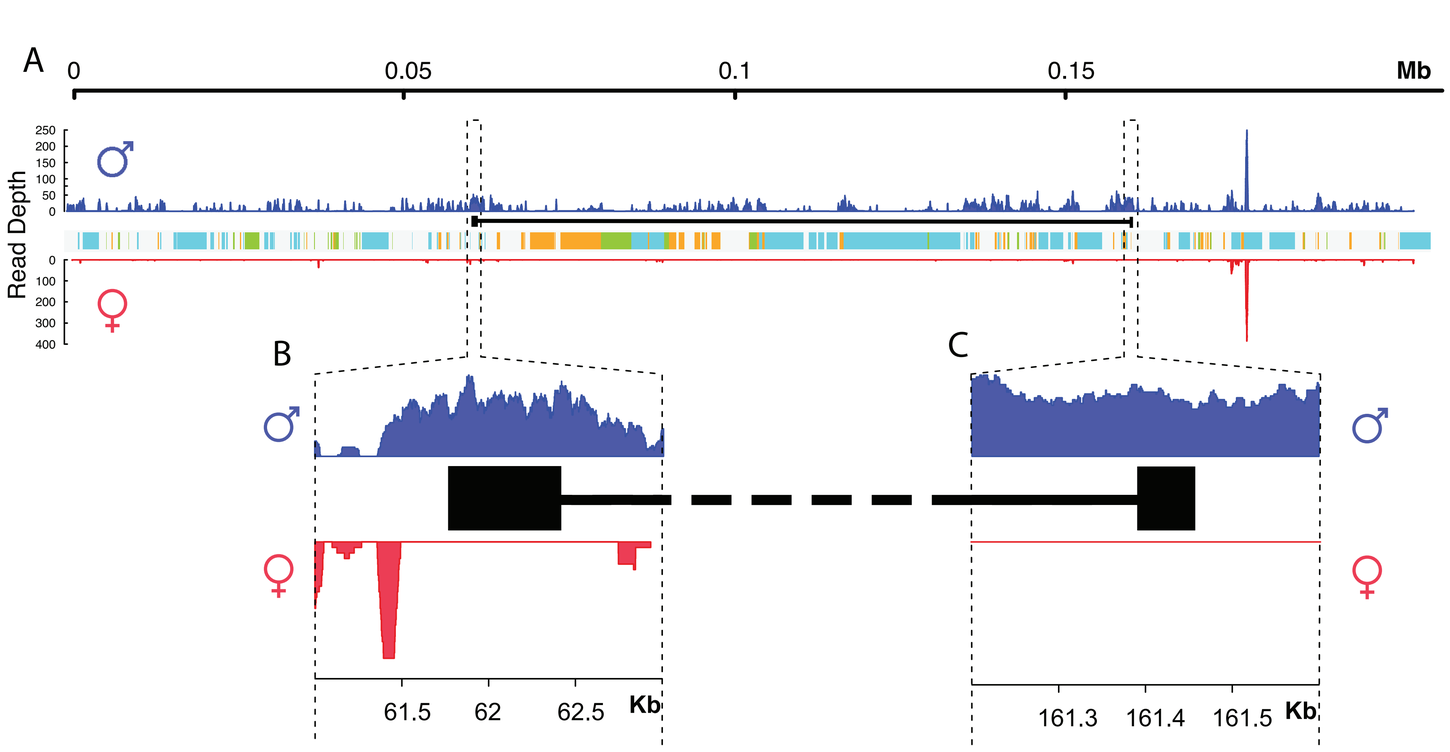
BAC library sequencing

- A BAC library was constructed from a DNA pool of Asian wild type sibling males for an estimated coverage of 5x (2.5x for sex specific regions)
- The BAC library was PCR screened using primers based on the complete coding sequence of *Nix.*
- Four positive clones were sequenced using the PacBio RSII platform.
- The sequence data was assembled with the CANU v1 assembler, followed by sequence polishing with QUIVER.

Data analysis

- Male and female gDNA and RNA-Seq reads were mapped to the sequence with BOWTIE 2.2.1 and TOPHAT 2.1.1, respectively.
- RNA-Seq data was processed using the CUFFLINKS 2.2.1 pipeline to look for potential genes and male/female specific expression from the region.
- Genes were predicted with AUGUSTUS using the *Ae. aegypti* model.
- Repetitive regions were described using REPEATMASKER 4.0.6 and the *Ae. aegypti* repeat database.

Gene structure of the sex determination switch, *Nix*



Structure and gene expression of the ~207 kb genomic region containing the *Nix* gene
Nix is shown as two black boxes representing the exons, joined by a black line representing the intron. Colours on the central track of **A** represent the classes of repetitive elements (orange: DNA transposons; cyan: Gypsy LTRs; green: Ty1/Copia LTRs). Blue histograms represent the coverage of RNA-Seq reads from male samples on the *y* axis; red histograms represent the coverage from female samples. **B** and **C** show enlargements of the first and second exons of *Nix* in the dotted regions in **A**, respectively.

Repeat content of *Nix*

Types and abundance of repeats in the 207kb assembled M locus region and 99kb *Nix* intron, identified with RepeatMasker using the *Aedes aegypti* repeat library.

	Entire region		*Nix* intron region	
Repeat Type	Number of elements	Percentage of sequence	Number of elements	Percentage of sequence
Retroelements	105	42.1%	49	51.0%
SINEs	8	0.81%	5	1.11%
Penelope	3	0.08%	2	0.20%
LINEs	24	5.43%	6	6.85%
L2/CR1/Rex	4	0.13%	0	0%
R1/L0A/Jockey	13	3.87%	3	6.60%
RTE/Bov-B	3	1.33%	0	0%
L1/CIN4	1	0.02%	1	0.05%
LTR Elements	73	35.8%	38	43.0%
BEL/Pao	9	0.71%	3	0.87%
Ty1/Copia	16	11.3%	14	19.2%
Gypsy/DIRS1	48	23.8%	21	23.0%
DNA transposons	97	11.7%	69	20.1%
Tc1-IS630-Pogo	11	3.87%	11	9.04%
Other (Mirage, P-element)	1	0.06%	0	0%
Unclassified	6	0.48%	3	0.22%
Small RNA	8	0.81%	5	1.11%
Satellites	1	0.75%	0	0%
Simple repeats	19	0.34%	7	0.24%
Low complexity	3	0.07%	1	0.04%
Total repeats		55.4%		71.6%

Results and Conclusions

- ***Nix* is ~100 kb in length** - exceptionally long for a gene expressed in the early embryo, and one of the longest in the mosquito genome.
- This length is because **the M locus has accumulated repeats in between protein-coding DNA** in a manner characteristic of a sex-limited chromosome, which are prone to degeneration by Muller's ratchet due to the lack of recombination (Bachtrog, 2013)
- **This is similar to other insect Y chromosomes** - e.g. repetitive sequences comprise almost the entire *Anopheles gambiae* Y chromosome (Hall *et al*, 2016), while some *Dropsophila* Y genes contain gigantic repetitive introns (Carvalho *et al*, 2001)

This study

Turner, J. *et al*. The sequence of a male-specific genome region containing the sex determination switch in *Aedes aegypti*. *bioRxiv*. doi: 10.1101/122804

References

Adelman ZN & Tu Z Control of mosquito-borne infectious diseases: sex and gene drive. *Trends Parasitol.* **32,** 219-229 (2016)
Alphey L Genetic control of mosquitoes. *Annu. Rev. Entomol.* **59,** 205-224 (2014)
Bachtrog D Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14,** 113-124 (2013)
Carvalho AB *et al* Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *PNAS* **98,** 13225-13230 (2001)
Hall AB *et al* A male-determining factor in the mosquito *Aedes aegypti*. *Science* **348,** 1268-70 (2015)
Hall AB *et al* Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *PNAS.* **113,** 201525164 (2016)

Acknowledgements

This work was funded by BBSRC PhD training grant BB/M503460/1 and a BBSRC grant BB/M001512/1. PacBio sequencing was conducted at the Centre for Genomics Research, University of Liverpool with the assistance of Dr Margaret Hughes and Dr John Kenny.