Chromosomal sex via epigenetic modification?

Zhang, X., Wagner, S. E., Holleley, C. E., Graves, J. A. M. and Georges, A.
Institute for Applied Ecology, University of Canberra, School of Life Sciences, La Trobe University

Abstract:
The central bearded dragon (Pogona vitticeps) has a ZZ male/ZW female sex chromosome system. The Z and W are tiny microchromosomes, but they are easily distinguishable microscopically by the regions of repetitive sequence on the larger W. We searched the sex chromosome sequences for candidate sex determining genes, but we detected no single copy W-specific genomic sequence. The most promising candidate gene was nr5a1, which codes for steroidogenic factor SFI, a protein essential for mammalian sex determination. This gene has alleles on both the Z and W (Z-nr5a1 and W-nr5a1), which have the same genomic sequence, and are subject to recombination. Three isoforms were detected in Z-nr5a1 transcripts from gonads of adult ZZ males, two of which would translate into intact protein. However, the Z-nr5a1 and W-nr5a1 alleles of ZW females produced sixteen isoforms, most of which contained chain terminating sequences. This suggests that the W-borne allele produces transcripts that are differentially spliced and produce truncated polypeptides. Structures of these truncated polypeptides confirm that the DNA binding domain is intact, so that they could act as competitive inhibitors of the full length intact protein. We propose that an altered configuration of the W chromosomes affects the splicing to generate inhibitory W-borne isoforms that suppress tests determination. Thus this GSD system may be controlled, not genetically but epigenetically, by the sex chromosomes.

1. Mapping nr5a1 sequences in Z and W microchromosomes in Pogona

The protein encoded by nr5a1 is SFI, commonly contain a N-terminal DNA-binding domain (DBD) and a C-terminal Ligand binding domain (LBD), linked by a hinge region (Figure 2A). The DBD functions in binding the target genes, and the LBD in ligand-mediated transcription activity, both are highly conserved. The hinge region displays high diversity between species. For most species, the hinge region is encoded within a large exon in the middle of the gene, but the model RefSeq transcript generated by NCBI for Pogona splits it into three exons (Exon 4 to 6) and defines an unresolved gap (Figure 2B). We resolve the genomic sequence of the region by Sanger sequencing. The region has a high GC content harboring four microsatellites (Figure 2C), including a mononmeric (CCCCCC) sub, a trinucleotide (CAC) sub, a mononmeric(GGGG) sub and trinucleotide (GGT/A) sub repeats. The monomeric C and G STRs were absent from other species, so appear to be Pogona specific.

2. The hinge region of Pogona nr5a1 gene contains GC-rich microsatellite repeats (STR)

The protein encoded by nr5a1 is SFI, commonly contain a N-terminal DNA-binding domain (DBD) and a C-terminal Ligand binding domain (LBD), linked by a hinge region (Figure 2A). The DBD functions in binding the target genes, and the LBD in ligand-mediated transcription activity, both are highly conserved. The hinge region displays high diversity between species. For most species, the hinge region is encoded within a large exon in the middle of the gene, but the model RefSeq transcript generated by NCBI for Pogona splits it into three exons (Exon 4 to 6) and defines an unresolved gap (Figure 2B). We resolve the genomic sequence of the region by Sanger sequencing. The region has a high GC content harboring four microsatellites (Figure 2C), including a mononmeric (CCCCCC) sub, a trinucleotide (CAC) sub, a mononmeric(GGGG) sub and trinucleotide (GGT/A) sub repeats. The monomeric C and G STRs were absent from other species, so appear to be Pogona specific.

3. No difference in gDNA sequences nor expression levels detected between male and female individuals

Figure 3. Genetic instability

We sequenced STR region of nr5a1 in 33 individuals of P. vitticeps and assigned alleles to the Z or W chromosome by pedigree analysis. The STR region appears genetically unstable and the W-nr5a1 has a higher mutation rate (Figure 3A). Mitotic recombination still occurs between the Z and W chromosome in nr5a1 although the frequency of recombination showed lower than that between two Z in ZZ individuals (Figure 3B). RNA-seq data showed that high levels of nr5a1 mRNA were only found in gonads, consistent with its significant role in gonadal development and maintenance (Figure 4A). We observed no significant difference in nr5a1 expression level between adult ZZ tests and ZZ ovaries (Figure 4B). RNA-seq and 3’RACE both detected a long 3’UTR tail of 3,895 nt in addition to a 2,210 bp 3’UTR (Figure 4C).

4. Sex differential alternative splicing in nr5a1 STR region

Table 1. nr5a1 DNA variants identified in male and female gonads

<table>
<thead>
<tr>
<th>STR</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pit_005005003443</td>
<td>Pit_001003344962</td>
<td>Pit_001003344020</td>
</tr>
<tr>
<td>Pit_005005003448</td>
<td>Pit_001003344862</td>
<td>Pit_001003344020</td>
</tr>
</tbody>
</table>

RT-PCR detected three nr5a1 cDNA variants from five ZZ male individuals and sixteen nr5a1 DNA variants from four ZW female individuals (Table 1). The cDNA variants result from alternative splicing in the STR region (Figure 5A and B).

5. Putative function of polypeptides encoded by alternative transcripts

Figure 5

Two of the three cDNA isoforms common to ZZ male and ZW female (nr5a1-al and nr5a1-br) and one only from ZW females (nr5a1-c) would translate into intact protein (Figure 5C and D). Thirteen of cDNA isoforms in ZW female contain a translation frame shift that introduces a premature stop codon. They could translate into polypeptide containing the DBD but not the LBD (Figure 5C and D), and retain the function of binding to the target genes.

6. Hypothesis

Figure 6. Hypothesis of the epigenetic control of sex determination by post-transcriptional control of nr5a1. The Z and W-borne nr5a1 have the same base sequence, but the W is held in an altered conformation. This makes the STR region in the middle of W-nr5a1 genetically unstable, and affects transcription and secondary structure of the pre-mRNA. This affects splicing, leading to many isoforms that are translated into truncated SFI polypeptides that act as competitive inhibitors of SFI function. Suppression of SFI leads to female development of ZW animals.

Future work: It will be exciting to test this hypothesis by phased long-read sequencing of Z and W chromosomes, Hi-C and recombination analysis across the entirety of both sex chromosomes, genome wide CHIP as well as detailed studies of transcripts early in embryonic development, and the effects of temperature on differential splicing.

Acknowledgement: We thank Jianwei Deakin, Kazumi Matsukura, Ira W. Deverson, Denis O’Neill, Harriit R. Patel, Teri Eser, Zhao Li, Chenmu Wang, Helen J. Edwards, Wendy Ruiter, Jacqui Richardson, Andrew Sinclair, Sarah Whiteley, Meghan Castelli, Shayer Alame and Foyez Shams for all sorts of help.

This work is funded by an Australian Research Council Discovery Grant (DP170101473) awarded to AG (LEAD), CER, WO, JAMG, TE, SS, LS, and PW.