Sex-specific splicing of Z- and W-borne nr5a1 alleles suggests sex determination is controlled by chromosome conformation

Xiwen Zhang1, Susan Wagner1, Clare E. Holleley2, b Janine E. Deakin3, Kazumi Matsubara1,2, Ira W. Deveson4, d
Denis O’Meally1,3, Hardip R. Patel5,6, Tariq Ezaza7, Zhao Li8, Chexu Wang9, Melanie Edwards5,4,
Jennifer A. Marshall Gravesa, f,5,6, and Arthur Georgesa,5,6

Pogona vitticeps has female heterogametgy (ZZ/ZW), but the master sex-determining gene is unknown, as it is for all reptiles. We show that nr5a1 (Nuclear Receptor Subfamily 5 Group A Member 1), a gene that is essential in mammalian sex determination, has alleles on the Z and W chromosomes (Z-nr5a1 and W-nr5a1), which are both expressed and can recombine. Three transcript isoforms of Z-nr5a1 were detected in gonads of adult ZZ males, two of which encode a functional protein. However, ZW females produced 16 isoforms, most of which contained premature stop codons. The array of transcripts produced by the W-borne allele (W-nr5a1) is likely to produce truncated polypeptides that contain a structurally normal DNA-binding domain and could act as a competitive inhibitor to the full-length intact protein. We hypothesize that an altered configuration of the W chromosome affects the conformation of the primary transcript generating inhibitory W-borne isoforms that suppress testis determination. Under this hypothesis, the genetic sex determination (GSD) system of P. vitticeps is a W-borne dominant female-determining gene that may be controlled epigenetically.

Sex in many vertebrate species is determined genotypically (genetic sex determination [GSD]) by sex chromosomes. Sex chromosomes are defined by sex-determining genes, which may differ between lineages, but all trigger a sex differentiation network that is highly conserved among vertebrates. Mammals normally have an XX female:XY male system, in which a male-dominant gene, SRY (1, 2), is located on the Y chromosome. It directs differentiation of a bipotential gonad down a male trajectory by up-regulating SOX9 and repressing genes that initiate the female trajectory; ultimately, this leads to the formation of functional testes in the embryo. In the absence of SRY, SOX9 is not up-regulated, the WNT/β-Catenin signaling pathway is not repressed, and XX embryos develop as females (3).

The vertebrate sex differentiation network involves many conserved genes, including NR5A1, which encodes the steroidogenic factor steroidogenic factor 1 (SF1). This gene plays a central role in male development of mammals (4). SF1 is important for the initial formation of the bipotential gonadal ridge in both sexes, but it later becomes differentially expressed and integrally involved in the establishment of the male phenotype. One central role of SF1 in mammals is in supporting the self-up-regulation of SOX9 and possibly, cross-suppression of female developmental processes, ultimately leading to the formation of a testis (5). Deletion of Nr5a1 in mice results in gonadal agenesis and a female phenotype, and SF1 loss of function mutations in humans lead to feminization (6).

Birds have a ZW female:ZZ male system in which the dosage of the Z-borne DMRT1 gene initiates male or female development (7, 8). In Xenopus laevis, a truncated copy of dmrt1 lies on a W chromosome. The product of the dmw lacks a transactivating domain, competes with the product of an autosomal dmrt1, and acts as a female dominant gene (9). Several master

Significance

Reptiles have an extraordinary variety of mechanisms to determine sex. The best candidate sex-determining gene in our model reptile (the Australian central bearded dragon) is the key vertebrate sex gene nr5a1 (coding for the steroidogenic factor 1). There are no sex-specific sequence differences between nr5a1 alleles on the sex chromosomes, but the Z- and W-borne alleles are transcribed into remarkably different alternative transcripts. We propose that altered configuration of the repeat-laden W chromosome affects the conformation of the primary transcript to generate more diverse and potentially inhibitory W-borne isoforms that suppress testis determination. This is a mechanism for vertebrate sex determination, in which epigenetic control regulates the action of a gene present on both sex chromosomes.


Reviewers: M.S., University of Würzburg Biocenter; C.S., Yellow Sea Fisheries Research Institute; and A.S., Murdoch Children’s Research Institute and ARC Centre of Excellence in Biotechnology and Development.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1X.Z. and S.W. contributed equally to this work.
2Present address: Department of Environmental Biology, Chubu University, Kasugai, Aichi 487-8501, Japan.
3Present address: Beckman Research Institute, City of Hope Medical Center, Duarte, CA 91010.
4Present address: Australian Wildlife Services, Deakin, ACT 2600, Australia.
5J.A.M.G. and A.G. contributed equally to this work.
6To whom correspondence may be addressed. Email: j.graves@latrobe.edu.au or georges@aerg.canberra.edu.au.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116475119/-/DCSupplemental.

Published January 24, 2022.
Sex-determining genes have now been isolated, largely from fish. Most encode transcription factors that are part of the complex sex differentiation network common to all vertebrates (10), involving dmyt1 and sax3 as well as amh, which is a glycoprotein hormone. No master sex-determining gene has yet been identified for any reptile species, although wt1 is a candidate in the wood turtle (Glyptemys insculpta) (11) and nr5a1 and its repressor foxl2 are candidates in the softshell turtle Apalone spinifera (12).

The task of discovering the master sex-determining gene in any species is formidable. It is expected to reside on the sex chromosomes, be differentially expressed in male and female embryos early in development, and promote sex reversal when manipulated experimentally (1, 2, 7). Sex-dominant genes must reside on a Y or W, whereas dosage-sensitive genes are expected to lie on the X/Z. Thus, a reasonable first step in the process of identifying the master sex-determining gene is to identify which genes reside on sex chromosomes.

Genome sequencing has yielded good assemblies of X or Z chromosomes in many species, but obtaining the sequence for the Y or W chromosome is technically challenging because the nonrecombining regions are replete with repeats (13). Assembly of the euchromatic male-specific regions of mammal Y chromosomes (human, chimpanzee, rhesus macaque, mouse, and pig) was achieved by painstaking sequencing and assembly of bacterial artificial chromosome (BAC) contigs (14–18). Recently, alternative approaches, such as flow sorting, microdissection, and in silico methods, have identified sequences unique to the heterogametic sex (19). Restriction site–associated DNA sequencing (20, 21) and in silico genome subtraction (22) have proven useful for identifying sex-linked markers that can then be mapped back to a reference sequence to identify genes on the Y or W chromosome. Development of long-read technologies promises phased sex chromosomes with chromosome-sized scaffolds (23–25).

We have undertaken a search for sex-determining genes in the model reptile, the central bearded dragon (Pogona vitticeps; Squamata:Agamidae). P. vitticeps has ZZ male/ZW female sex chromosomes. Both sex chromosomes are in the microchromosome range, but the W chromosome is slightly longer than the Z and is readily distinguishable by C banding and repetitive sequence probes (26).

We, therefore, looked for sequences that differed quantitatively or qualitatively between ZZ males and ZW females and followed this search by identifying genes located on the Z and W chromosomes. The most prominent sex-determining candidate was nr5a1, which encodes the SFI whose action is required for male development in all mammals. We found copies of nr5a1 on the Z and W that encode functional SFI. Both alleles were transcribed at similar levels in the gonad of both sexes. ZZ male and ZW female gonads produced transcripts that can be translated into functional SFI, but ZW females also produced many transcript variants with premature stop codons whose truncated polypeptide products could act as competitive inhibitors of SFI and cause female development.

Results

Screening for W-Specific Single-Copy Sequences. Based on the obvious cytogenetic differences between the sex chromosomes in P. vitticeps (26), we expected that the sequence of the Z and W chromosomes would be sufficiently different to use genome subtraction to detect regions of the W chromosome that carried female-specific genes or alleles. Representational difference analysis–differential subtraction chain (RDA-DSC) (27) was, therefore, applied to a pool of 10 male ZZ (driver) and 10 female ZW (tester) dragons. This identified W-specific repetitive sequences but identified no coding regions specific to the W chromosome. (1)

Sequences from SI Appendix, Supplementary Text a male ZZ, a female ZW, and a sex-reversed female ZZ dragon (28) were used for an in silico subtraction. This strategy detected many polymorphisms in wild populations of P. vitticeps, with 89 large indels (1,028 to 6,210 bp) occurring frequently (allele frequency range: 0.03 to 0.65) (Dataset S3). The presence of so many autosomal polymorphisms produced many false positives that made isolating sex-specific differences challenging. Despite a robust bioinformatics pipeline and empirical validation approach to identify sex-specific sequences, we detected no single-copy W-specific genomic sequence (SI Appendix, Supplementary Text and Dataset S4).

We also used the proprietary reduced representation approach DArT-seq™ (Diversity Arrays Technology Pty Ltd.) (29, 30) (SI Appendix, Supplementary Text) to detect sex-specific single-nucleotide polymorphisms (SNPs). A total of 18 individuals (3 × ZW and 15 × ZZ) were individually genotyped using DArT-seq. However, analysis of the 62,222 SNP dataset revealed only a single locus with a putatively sex-linked pattern: TGCAGCTTGTITTTACGAAACAAAAAGTITTTGTTGCGTGGCCCTATCTTTAGACAG[TW1G2]GTGTITGTCCAAA (clone identification 4061001). This sequence occurs on scaffold NW_018163204 of the P. vitticeps genome, which was physically mapped to autosomal microchromosome 10 and not to the sex chromosomes (31).

Thus, three molecular approaches to identify sex-linked sequences in P. vitticeps revealed no differences in genomic coding sequences on the Z and W chromosomes. Although we cannot exclude the presence of a single-copy gene that eluded our detection, this suggests that the differences between the Z and W chromosomes that are obvious in cytological preparations are largely or entirely restricted to repetitive sequence.

Search for Candidate Sex-Determining Genes. We sequenced the genome of a male P. vitticeps, annotating 19,406 protein-coding genes in the genome assembly Pvi1.1 by de novo prediction and alignment to the green anole, chicken, and human genomes (28).

Sex chromosomes in P. vitticeps are microchromosomes (26). Therefore, we scanned the dragon genome assembly for genes that map to scaffold contigs or BAC clones assigned to microchromosomes in P. vitticeps (31) and have a conserved role in vertebrate sex determination. Three candidate genes were identified—nr5a1, wnt4, and cyp19a (aromatase). Of these, only scaffold NW_018150817 containing nr5a1 mapped to the Z chromosome (31). The Nr5a1 coding sequence was not detected in the incompletely sequenced microdissected W chromosome that was previously used to characterize gene content of the P. vitticeps Z chromosome. However, when we isolated a BAC clone (116G15) containing the nr5a1 gene and hybridized this together with previously confirmed sex chromosome BAC clone 3L7 (32) onto metaphase chromosomes from a confirmed genotypic ZW female, signal was observed on the W chromosome as well as the Z chromosome (Fig. 1 and SI Appendix, Supplementary Text).

We found that the read depth of nr5a1 in the sequenced genome of ZZ and ZW individuals was well within the range of a single-copy gene, and the read depth of nr5a1 in ZZ was comparable with that of ZW (SI Appendix, Fig. S1). We concluded that P. vitticeps harbors a single copy of nr5a1 on both the Z and W chromosomes. We will refer to the copies on the Z and W chromosomes as the Z-nr5a1 and W-nr5a1 alleles, respectively.

Characterization of nr5a1 in P. vitticeps. Nr5a1 and its homologs were originally identified as fushi tarazu factor 1 box (Ftz-F1), expressed early in Drosophila embryogenesis (33), and have
since been cloned from vertebrates ranging from humans to fish (34–44). Mammals and chicken possess two members of this family, nr5a1 and nr5a2 (37), as does P. vitticeps too, and several homologs occur in fish. The coding regions of P. vitticeps nr5a1 and nr5a2 share 70 to 80% sequence identity.

The protein encoded by nr5a1 is SF1. SF1 belongs to a family of transcription factors that commonly contain an N-terminal DNA-binding domain (DBD) consisting of two zinc finger modules and Ftz-F1, a C-terminal ligand-binding domain (LBD), two functional activation domains AF-1 and AF-2, and a hinge region joining the DBD and the LBD (Fig. 2A). The DBD domain is involved in specific recognition and interaction with DNA target sequences, and the LBD plays a crucial role in ligand-mediated transcription activity; both are highly conserved among vertebrates. The hinge region is not simply a linker but is also important for transcriptional activity of nr5a1 (45–47). However, it displays high diversity in length and amino acid composition between species (Fig. 2D).

The reference sequence (RefSeq) model of the Pogona nr5a1 gene (XM_020793311.1) spans 32,976 bp and is predicted to contain nine exons (Fig. 2A). The region containing the DBD (encoded by exons 2 to 4) is well resolved and shares 96% protein identity with the human DBD sequence. The LBD (encoded by exons 6 to 9) shows 72% protein identity to human SF1. The protein sequence of the hinge region cannot be determined confidently from the genome assembly because 376 Ns were absent in other species, so they appear to be Pogona specific (SI Appendix, Fig. S2 and Supplementary Text).

In the ZZ genomic assembly Pvi1.1, a duplication of identical sequences, each 200 bp in length, immediately follows the previously unresolved stretch of Ns (Fig. 2B and C). All 422 amplicons from 33 individuals that we sequenced contained only one 200-bp repeat, not two, suggesting that the second 200-bp repeat is absent, presumably an assembly error in Pvi1.1.

The genomic sequence that we resolved perfectly connects the N-terminal part of Pogona SF1 with its C-terminal part in frame, as a continuous larger exon (hereafter termed exon 4b), as seen for most model organisms. The resulting hinge region of Pogona SF1 is larger than that of mammals, chicken, alligator, and Anolis carolinensis (Fig. 2D and SI Appendix, Supplementary Text; see Fig. 5D, ii).

**Genetic Instability of the STR Region of W-nr5a1.** Sequencing nr5a1 in 33 individuals of P. vitticeps revealed multiple sequences in the nr5a1 STR region that varied at single nucleotides and in repeat copy number. Not only was there polymorphism between individuals, but there were more than two genome sequences within each individual, indicating somatic mutations in this single-copy gene.

We could assign alleles to the Z or W chromosome by pedigree analysis (SI Appendix, Fig. S3) and distinguish variability in Z-nr5a1 and W-nr5a1 alleles based on the principle that ZW offspring inherit their W from the ZW mother and that from the ZZ father. We sequenced to achieve a target of 10 clones per individual (range 6 to 39) of the nr5a1 STR region (Dataset S1). Based on allele-specific SNPs and indels in the STR region, genetic clones were categorized as Z borne (Z-nr5a1) or W borne (W-nr5a1).

The spectrum of acquired somatic mutations for the nr5a1 STR region can be categorized into 1) substitutions and indels of a single nucleotide, 2) copy number change in the tandem repeats, 3) recombination between two alleles, and 4) large fragment deletions. Category 1 and 2 mutations occurred with slightly higher frequency in W-nr5a1 than Z-nr5a1, but the difference was not statistically significant (Fig. 3A).

Differentiated sex chromosomes are typically characterized by a loss of homology so that the differentiated regions no longer pair and recombine at meiosis. In contrast, we observed that mitotic recombination still occurs between the Z and W chromosomes in P. vitticeps, despite their cytogenetic differences. However, the frequency of somatic recombination between the Z and W chromosomes in ZZ females was lower than that between the two Z chromosomes in ZZ animals (Fig. 3B), suggesting repression in the heterogametic sex.

**Gonad-Specific Expression of nr5a1.** We used RNA sequencing (RNA-seq) data from gonadal and various somatic tissues of adult individuals (48) to examine the nr5a1 expression profile in P. vitticeps. Of the tissues we examined (Fig. 4A), higher levels of nr5a1 messenger RNA (mRNA) were found only in gonads, consistent with its expression profile in other vertebrates (GTEXPortal; https://gtexportal.org/home/) (Fig. 4A) and a significant role in gonadal development and maintenance.

We observed no significant difference in nr5a1 expression levels between adult ZZ testis and ZW ovaries (Fig. 4A). This was surprising for a putative sex-determining gene, although this does not exclude differential expression during embryonic development. The presence of numerous heterozygous SNPs confirmed that nr5a1 is transcribed from both alleles in ZZ and ZW individuals, as was the case for autosomal genes (Fig. 4B). This is not consistent with a simple SF1 dosage mechanism of sex determination, in which the W allele is inactive or less active.

RNA-seq read depth indicated that gonadal nr5a1 transcripts had two different 3’ untranslated regions (UTRs), producing mRNAs of different lengths; a more abundant 3’ UTR had a length of 2,210 nt corresponding to the NCBI annotation, and the other was estimated at 3,895 nt, which was verified by 3’ rapid amplification of cDNA ends (3’ RACE) (Fig. 4C). The longer 3’ UTR is considerably longer than 3’ UTRs of nr5a1
Fig. 2. The hinge region of *P. vitticeps* nr5a1 is especially long and GC rich. (A) Row 1 shows the human nr5a1 gene. Row 4 shows the *Pogona* nr5a1 gene. Both are as annotated in the NCBI RefSeq database (accession no. human NM_004959.5, *Pogona* XM_020793111.1). Black and gray boxes indicate exons, with coding regions in black and noncoding regions in gray. N indicates a sequence between *Pogona* exons 4 and 5, which was not resolved in the genome assembly Pvi1.1. Rows 2 and 3 illustrate the broad structure of the canonical SF1 encoded by nr5a1 for humans (accession no. NP_004950.2) and *Pogona* (accession no. XP_020648970.1). Orange boxes indicate the DBDs, blue boxes indicate the LBDs, and green boxes indicate the hinge regions. The percentage of identity of each region to the human protein sequence is indicated. The length of SF1 as the number of amino acids is given at the right for each species; however, owing to the unresolved sequence in *Pogona*, the hinge region is unlikely to have been annotated correctly by the automated annotation pipeline (as displayed by the red question marks). The positions of two zinc fingers (ZnI, ZnII), the activation function domains 1 and 2 (AF-1 and AF-2, respectively), the Ftz-F1, and the proline-rich region (Prorich) are also denoted (4, 36). (B) Expanded region between exons 4 and 6. N indicates the unresolved region in the Pvi1.1 assembly. Two large blue arrows indicate two copies of a 200-bp identical repeat in the Pvi1.1 assembly. Small black arrows indicate primers used to resolve the sequence in the region between exons 4 and 6. (C) The region between exons 4 and 6 as resolved by cloning and sequencing revealing GC-rich microsatellites. From left to right, dispersed (CCCCC)n pentamers (light yellow arrows), a (CAC)n repeat (light yellow arrows), dispersed (GGGGG)n pentamers (light blue arrows), and a (GGT/A)n repeat (light blue arrows) are shown. The exact gDNA sequence of the regions with the dispersed pentamers is given in boxes for one haplotype of individual Pit_001003387339. The exons 4 to 6 region is likely to represent one continuous exon, indicated by the dashed lines and gray shading and herein termed exon 46. (D) Multiple protein sequence alignment of SF1 from chicken (G.g.; NP_990408.1), humans (H.s.; NP_004950.2), mouse (M.m.; NP_001303616.1), zebrafish (D.r.; NP_571869.1), American alligator (A.m.; XP_006271373.2), *Anolis* (A.c.; XP_008122440.1), and *Pogona* (P.v.) obtained with MUSCLE 3.8 (69). For *Pogona*, the genomic DNA consensus sequence (obtained with EMBOSS Cons) of individual Pit_001003387339 of the unresolved sequence in Pvi1.1 was translated into the protein sequence as one continuous exon (exon 46). The regions are indicated with colored boxes at the top of the sequence blocks (DBD, orange; hinge region, green; LBD, blue).
reported in humans, mice, chicken, alligator, turtle, garter snake, and Anolis. A short interspersed nuclear element (SINE) repeat resides in the longer 3′ UTR variant, which may function in regulation of RNA degradation and gene expression (45).

Sex Differential Alternative Splicing in *P. vitticeps* nr5a1 STR Region. RT-PCR verified that all RefSeq exons, except exon 5 (including the absent 200-bp repeat), are expressed in *P. vitticeps* nr5a1 transcripts. Unexpectedly, RT-PCR amplified multiple bands using a primer pair spanning exon 46. Cloning and sequencing of RT-PCR products identified a variety of complementary DNA (cDNA) isoforms produced by excising different parts within the STR region of exon 46.

In order to distinguish splicing of the Z-nr5a1– and W-nr5a1–borne transcripts, we compared cDNA isoforms from ZZ and ZW animals. In each of five ZZ males, we obtained two cDNA isoforms (nr5a1ΔAS59, nr5a1ΔSS50) (Fig. S4 and Table 1), and one ZZ male had an additional isoform (nr5a1ΔS70). These isoforms must be transcribed from Z-nr5a1 alleles.

RT-PCR on four ZW individuals produced one or all three of these Z-nr5a1 cDNA isoforms. In addition, we found 13 distinct isoforms from ZZ individuals. Each ZZ individual had up to five of these additional isoforms (Fig. S5 and Table 1). These 13 isoforms were never observed in ZZ animals and thus, must be produced by the W-nr5a1 allele.

In addition to the 16 isoforms, 1 isoform present in every male and female individual lacked the entire exon 46 (nr5a1ΔAS60) through a canonical splicing event. We did not retain an isoform containing the full exon 46, although RNA-seq suggested this to be the most abundant isoform. Both methods, RT-PCR and RNA-seq, are challenged by the high repetitive content in this region.

Thus, the Z-nr5a1 and W-nr5a1 alleles, despite showing no differences in expression level, produce an array of nr5a1 isoforms in a sex-specific fashion.Unfortunately, however, we are unable to assess the relative abundance of nr5a1 isoforms from RNA-seq experiments because of problematic read mapping to the splice sites, which are in highly repetitive regions.

Putative Function of Polypeptides Encoded by Alternative Transcripts. Which of these isoforms encodes a functional polypeptide? Two of the three isoforms common to ZZ males and ZW females (nr5a1ΔAS59, nr5a1ΔS50) and one isoform recovered only from ZW females (nr5a1ΔAS60) can be translated from the canonical start of nr5a1 through to its canonical stop codon, producing polypeptides containing the whole DBD and the whole LBD but lacking almost the entire hinge region (Fig. 5C and SI Appendix, Fig. S7).

Nr5a1ΔAS33 from both ZZ and ZW and the other 12 female-specific isoforms all contain a translational frameshift that introduces a premature stop codon. Unless they are degraded by the nonsense-mediated mRNA decay mechanism (50), they would translate into polypeptides containing the DBD but not the LBD (Fig. 5C and SI Appendix, Fig. S7). The three-dimensional structure of the DBD of all truncated isoforms was predicted to fold correctly (Fig. 5D and SI Appendix, Fig. S8), suggesting that it is functional. Therefore, these truncated peptides might act as SF1 repressors (51) by competing with full-length SF1 for binding sites of downstream genes (for example, *dmrt1* or *sox9*).

The remaining isoform (nr5a1ΔAS60) splices out the last third of the DBD, causing a +2 frameshift and premature stop codon. It seems unlikely that this polypeptide is functional.

**Discussion**

The ability of *P. vitticeps* ZZ embryos to develop as females in the absence of a W chromosome led to the hypothesis that sex determination in this species is controlled by a dosage-sensitive gene on the Z chromosome (52). We, therefore, expected to detect qualitative or quantitative differences in sequences between the ZZ and ZW genotypes that could identify the gene whose dose is responsible for sex determination. However, our recent work implying that temperature–sex reversal uses a different regulatory pathway (53) means that we can no longer discount the role of the W chromosome in GSD at normal temperatures.

We, therefore, searched for W-specific sequences. The Z and W chromosomes are cytologically distinct, the longer W having a female-specific repetitive sequence cytologically visible as c banding and able to be used as markers (54). However, we found no unique differences in gene sequences between the Z and W chromosomes using both in vitro and in silico subtraction approaches. Thus, no genetic differences were found in *P. vitticeps* that could initiate male or female development in this species.

**Putative Function of Polypeptides Encoded by Alternative Transcripts.**

Which of these isoforms encodes a functional polypeptide? Two of the three isoforms common to ZZ males and ZW females (nr5a1ΔAS59, nr5a1ΔS50) and one isoform recovered only from ZW females (nr5a1ΔAS60) can be translated from the canonical start of nr5a1 through to its canonical stop codon, producing polypeptides containing the whole DBD and the whole LBD but lacking almost the entire hinge region (Fig. 5C and SI Appendix, Fig. S7).

Nr5a1ΔAS33 from both ZZ and ZW and the other 12 female-specific isoforms all contain a translational frameshift that introduces a premature stop codon. Unless they are degraded by the nonsense-mediated mRNA decay mechanism (50), they would translate into polypeptides containing the DBD but not the LBD (Fig. 5C and SI Appendix, Fig. S7). The three-dimensional structure of the DBD of all truncated isoforms was predicted to fold correctly (Fig. 5D and SI Appendix, Fig. S8), suggesting that it is functional. Therefore, these truncated peptides might act as SF1 repressors (51) by competing with full-length SF1 for binding sites of downstream genes (for example, *dmrt1* or *sox9*).

The remaining isoform (nr5a1ΔAS60) splices out the last third of the DBD, causing a +2 frameshift and premature stop codon. It seems unlikely that this polypeptide is functional.

**Discussion**

The ability of *P. vitticeps* ZZ embryos to develop as females in the absence of a W chromosome led to the hypothesis that sex determination in this species is controlled by a dosage-sensitive gene on the Z chromosome (52). We, therefore, expected to detect qualitative or quantitative differences in sequences between the ZZ and ZW genotypes that could identify the gene whose dose is responsible for sex determination. However, our recent work implying that temperature–sex reversal uses a different regulatory pathway (53) means that we can no longer discount the role of the W chromosome in GSD at normal temperatures.

We, therefore, searched for W-specific sequences. The Z and W chromosomes are cytologically distinct, the longer W having a female-specific repetitive sequence cytologically visible as c banding and able to be used as markers (54). However, we found no unique differences in gene sequences between the Z and W chromosomes using both in vitro and in silico subtraction approaches. Thus, no genetic differences were found in *P. vitticeps* that could initiate male or female development in this species.

**Putative Function of Polypeptides Encoded by Alternative Transcripts.**

Which of these isoforms encodes a functional polypeptide? Two of the three isoforms common to ZZ males and ZW females (nr5a1ΔAS59, nr5a1ΔS50) and one isoform recovered only from ZW females (nr5a1ΔAS60) can be translated from the canonical start of nr5a1 through to its canonical stop codon, producing polypeptides containing the whole DBD and the whole LBD but lacking almost the entire hinge region (Fig. 5C and SI Appendix, Fig. S7).

Nr5a1ΔAS33 from both ZZ and ZW and the other 12 female-specific isoforms all contain a translational frameshift that introduces a premature stop codon. Unless they are degraded by the nonsense-mediated mRNA decay mechanism (50), they would translate into polypeptides containing the DBD but not the LBD (Fig. 5C and SI Appendix, Fig. S7). The three-dimensional structure of the DBD of all truncated isoforms was predicted to fold correctly (Fig. 5D and SI Appendix, Fig. S8), suggesting that it is functional. Therefore, these truncated peptides might act as SF1 repressors (51) by competing with full-length SF1 for binding sites of downstream genes (for example, *dmrt1* or *sox9*).

The remaining isoform (nr5a1ΔAS60) splices out the last third of the DBD, causing a +2 frameshift and premature stop codon. It seems unlikely that this polypeptide is functional.

**Discussion**

The ability of *P. vitticeps* ZZ embryos to develop as females in the absence of a W chromosome led to the hypothesis that sex determination in this species is controlled by a dosage-sensitive gene on the Z chromosome (52). We, therefore, expected to detect qualitative or quantitative differences in sequences between the ZZ and ZW genotypes that could identify the gene whose dose is responsible for sex determination. However, our recent work implying that temperature–sex reversal uses a different regulatory pathway (53) means that we can no longer discount the role of the W chromosome in GSD at normal temperatures.

We, therefore, searched for W-specific sequences. The Z and W chromosomes are cytologically distinct, the longer W having a female-specific repetitive sequence cytologically visible as c banding and able to be used as markers (54). However, we found no unique differences in gene sequences between the Z and W chromosomes using both in vitro and in silico subtraction approaches. Thus, no genetic differences were found in *P. vitticeps* that could initiate male or female development in this species.
Sex differential cDNA isoforms of \(nr5a1\) that differ in the STR region. (A) The gDNA sequence of exon 46 of \(nr5a1\) and the three cDNA isoforms obtained from ZZ individuals. For orientation, NCBI RefSeq exons 4 and 6 are indicated with gray boxes. The nucleotide highlighted in red is the last nucleotide in the transcript before the excised part. The nucleotide highlighted in green is the first nucleotide present in the transcript after the excised part. (B) The same as for A but showing the additional cDNA isoforms in a representative ZW female, Pit_001003342982. (C and D) Putative polypeptides translated from the \(nr5a1\) cDNA isoforms identified in this study. The DBD, hinge region, and LBD are shown in orange, green, and blue, respectively. N and C termini are shown in black. Gray indicates an out-of-frame C-terminal peptide sequence occurring after a frameshift. The full-length protein, SF1 FL, was translated from the herein-resolved gDNA sequence (consensus of individual Pit_001003387339 obtained with EMBOSS Cons). The isoforms are grouped by the frameshift they are causing. (C) Multiple protein sequence alignments (by MUSCLE 3.8). Full-length SF1 is shown on top (FL). All three isoforms without a frameshift are shown next (pink), followed by a representative of isoforms producing +1 (yellow) and +2 frameshifts (purple); 171 amino acids of the hinge region of the full-length protein are not shown, as they are missing in all isoforms. A simplified schematic of the resulting protein is given at the bottom. The full alignment of all isoforms is shown in SI Appendix, Fig. S7. (D) Predicted three-dimensional structures of putative polypeptides translated from \(nr5a1\) cDNA isoforms. The DBDs of all structures are at a similar angle, while the LBDs are at different angles. (D, i) DBD of the full-length protein of \(Pogona\) superposed on that of humans (NP_004950.2) shows nearly identical conformation. (D, ii) \(Pogona\) full-length SF1. The unstructured hinge region containing several homopolymeric tracts is exceptionally long in \(Pogona\). (D, iii) \(nr5a1\Delta549\) is a representative of isoforms with no frameshift. DBD and LBD fold correctly. (D, iv) \(nr5a1\Delta523\) is a representative of isoforms producing a +1 frameshift. The DBD of the truncated polypeptide folds correctly. (D, v) \(nr5a1\Delta575\) is a representative of isoforms producing a +2 frameshift. The DBD of the truncated polypeptide folds correctly, despite the longer unstructured C-terminal peptide that is out of frame. The structures were predicted using AlphaFold (66). The three-dimensional structures of all isoforms are shown in SI Appendix, Fig. S8.
Our subsequent approach was, therefore, to identify potential *P. vitticeps* sex-determining genes by scanning the sequence on the Z and W chromosomes for candidate genes. We demonstrated that the sex microchromosomes contained *nr5a1*, which encodes SF1, a protein with important male-determining roles in all vertebrates. We showed that *nr5a1* maps to both the Z and W chromosomes, and its expression is confined to the gonads in both sexes. We characterized *P. vitticeps nr5a1*, finding that its structure is similar to orthologs in other vertebrates, encoding DBDs and LBDs, separated by a hinge region.

However, we found that among the vertebrates we surveyed, the *P. vitticeps nr5a1* sequence uniquely contains a region (within exon 46, which encodes the hinge region) that is CG rich and includes four STRs. Our most puzzling and most intriguing finding is that this STR region of the W-borne *nr5a1* is transcriptionally unstable, consistent with reports that GC richness and microsatellite tandem repeats can elevate genetic instability and influence transcription (55–57).

We also observed striking differences in *nr5a1* transcripts in ZZ and ZW dragons that relate to differences in splicing in the STR region. Four isoforms that translate into intact SF1 protein were observed across both ZZ and ZW individuals and are presumably attributable to the Z-*nr5a1* allele on the Z chromosome. However, another 13 isoforms are observed specifically in ZW animals and are presumably transcripts of the W-borne W-*nr5a1* allele. We cannot assess the relative abundance of the individual isoforms; however, most of these W-specific transcripts encode a truncated SF1 protein containing the DBD but lacking the activator domain (LBD). These truncated polypeptides could interfere with SF1 action by competing for DNA-binding sites, and the diminished SF1 activity would result in female-determining activity and sex determination (59) (Fig. 6).

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 entire sex chromosomes as well as detailed studies of transcripts early in embryonic development and the effects of temperature on differential splicing.

Materials and Methods

Gene nomenclature for *P. vitticeps* follows that of Kusumi et al. (60).

Animal Husbandry and Tissue Collection. The specimens and tissues used for the RNA isoform analysis (5 males, 4 females) and genomic DNA (gDNA) sequence analysis (14 males, 19 females) are provided in Dataset S1. Genotypic sex was verified using a PCR sex test (54).

*Nr5a1* Copy Number Verification. To test the copy number of *nr5a1*, we used already available ZZ DNA sequencing reads for assembly Pvi1.1: GCF_900067755.1 (28). The ZW reads were obtained from an Illumina run (36x) of a second individual verified as ZW by cytology. BWA (Burrows-Wheeler Alignment Tool Version 0.7.17-r1188) (61) was used with default parameters to map

![Fig. 6.](Image)

Epigenetic control of sex determination by posttranscriptional control of *nr5a1* in *P. vitticeps*. The Z and W chromosomes have the same gene content, but the repeat-rich W assumes an altered conformation. This renders the STR region more vulnerable, and presents an altered template for transcription and folding that controls splicing and ultimately, SF1 activity and sex determination (59) (Fig. 6).

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 entire sex chromosomes as well as detailed studies of transcripts early in embryonic development and the effects of temperature on differential splicing.
the reads to the draft Pogona assembly Pvl1.1. Bedtools [v2.27.1] (62) was used to generate bed graphs.

Molecular Cloning and Sequencing of the GR-Chr Microsatellite Region of *nr5a1*. Genomic DNA was isolated from the DNeasy Blood & Tissue Kits (QIA-GEN). For blood samples on Whatman FTA elute cards (GE Healthcare Life Sciences, Merck, Sigma-Aldrich), a 3-mm disk was removed from the center of each blood spot with a hole puncher. To elute the gDNA, each disk was soaked in 500 μL water at 95 °C for 30 min with vortexing, according to the manufacturer’s instructions.

To sequence the genomic DNA between exons 4 and 6 of *nr5a1* (which includes the unresolved sequence between exons 4 and 5), we designed oligo primers (Fig. 2B and Dataset S2) with high annealing temperature to enhance amplification of troublesome sequences, such as GC-rich regions (63). Primer sequences were manipulated to amplify specifically *nr5a1* by blast against the *P. vitticeps* genome. PCR reactions were carried out using MyFi DNA polymerase (Bio-21117; Bioline; Australia Pty Ltd.) under the following conditions: 250 to 200 ng of genomic DNA, 10 μM each primer, 2 μL of MyFi DNA polymerase, and 5 μL of 5x buffer supplemented in a final volume of 25 μL. An initial denaturation step at 95 °C for 1 min was followed by 35 cycles of denaturation at 95 °C for 15 s and annealing/extension at 72 °C for 1.25 min, followed by a final extension step at 72 °C for 10 min.

PCR products were purified using the QIAquick PCR Purification Kit (QIA-GEN), directly sequenced (Macrogen) or cloned into pc4RAbrTM vector or the TOPO XL-2 Complete PCR Cloning Kit (Invitrogen Australia Pty Ltd.), and transformed into TOP10 chemically competent Escherichia coli cells (Invitrogen Australia Pty Ltd.). The DNA was then purified and sequenced using the Sanger ABI platform (Macrogen).

**DART Analysis to Verify Parentage of Pogona Families.** To verify parentage, prospective parents and offspring were each genotyped using DArT-seq (Diversity Arrays Technology Pty Ltd.) (29, 30), and sequence data were manipulated in software package dartR (64). Briefly, the SNPs were filtered on reproducibility (percentage of technical replicates that agreed, threshold 100%), read depth (threshold 16x), and call rate (threshold 0.95). A relatedness matrix (G) was calculated with gl.grm in dartR, which essentially draws upon the A.mat function of R package rrBLUP (65). The G matrix was visually represented as a network using gl.grm.network in dartR, and clusters were used to verify parent–offspring relationships qualitatively (SI Appendix, Fig. S3).

**Identification of Z- or W-Linked *nr5a1* and Evaluation of Mutation Rates.** The genomic DNA sequence of *nr5a1* exons 4 to 6 including the STR region was used to identify Z- or W-linked *nr5a1* alleles. Families with verified parentage were analyzed for this purpose (Dataset S1 and SI Appendix, Fig. S3). Sequences were aligned using ClustalW in Geneious analysis software (Geneious version 8.1.1 by Biomatters). First, gDNA clones from one individual were categorized into two allele groups based on two sets of SNPs and indels that occurred exclusively in one group of gDNA clones. SNPs or indels that occurred in only one or a few clones were likely acquired somatically during the life of individuals and not regarded for identifying Z- and W-linked alleles. The two alleles (i.e., the two sets of SNP/indels of *nr5a1* clones of the daughters) were compared with those of the mother, and the allele group with the matching set of SNP/indels to one of the mother allele groups was identified as W linked, while the other allele group with SNP/indels matching to one of the father allele groups was identified as Z linked. Subsequently, we assigned the two mother allele groups as Z or W linked.

Evaluating somatic mutation rates was carried out in Geneious using the pairwise Distance function. First, for each individual we obtained the pairwise identity percentages of gDNA clones within an allele group. The clone with the highest identity with others in the allele group was defined as standard. All other clones were compared with the standard for the percentage of identity in the Distance function. The mutation rate was calculated as 100 − percentage of identity. Then, the average and SE of the mutation rates for all clones of the W allele (n = 55) and the Z allele (n = 59) for ZW individuals and of both Z alleles (n = 74) of ZZ individuals were calculated.

Occasionally, one clone had an SNP/indel pattern of one allele group in the first part and an SNP/indel pattern of the other allele group in the last part. When either part was long enough so that at least three SNP/indels of each allele group were present, the clone was not defined as Z or W linked but as a recombination of both alleles (Dataset S1).

**Molecular Cloning and Sequencing of *nr5a1* DNA Variants.** Total RNA extracts of *P. vitticeps* gonad tissues were obtained using the PureLink RNA Micro Kit (ThermoFisher Scientific) following the manufacturer’s guidelines. Contamination with genomic DNA was removed using PureLink DNase (ThermoFisher Scientific). The quality of the RNA was assessed using a Nanodrop 1000 spectrophotometer (ThermoFisher Scientific).

cDNA was synthesized from 1 μg RNA in a total volume of 20 μL using SuperScript IV reverse transcriptase (Invitrogen; ThermoFisher Scientific) and random hexamer oligo primers (Invitrogen; ThermoFisher Scientific) following the manufacturer’s guidelines. Subsequently, 1 μL of the cDNA reaction was used to perform PCR in a 25-μL reaction mix with 2.0 units of Taq Red DNA polymerase (Bioline Australia Pty Ltd.), 10 μM each primer, and 5 μL of 5x buffer (Bioline; Australia Pty Ltd.). The thermal cycling was held at 95 °C for 1 min followed by 35 cycles at 95 °C for 15 s, an annealing step at a temperature dependent on the primer annealing temperature, and an elongation step at 72 °C whose time depended on the expected length of the templates. Primers for the PCR were designed over exon–exon boundaries whenever possible to ensure amplification of cDNA derived from mature mRNA (Dataset S2). Primer sequences were confirmed to amplify specifically *nr5a1* by blast against the *P. vitticeps* genome and alignment to a transcript of *nr5a2*.

**Three-Dimensional Protein Structure Prediction of *nr5a1* DNA Variants.** The three-dimensional structure of the putative translated protein of *nr5a1* DNA variants was predicted with AlphaFold (66).

**Nr5a1 Expression Profiling.** We used the recently published global gene expression analysis data from adult male (ZZ) and female (ZW) dragon tissues, including gonad (n = 2), brain (n = 2), and additional somatic tissues (liver, kidney, lung, heart, and muscle; n = 1) (48). In short, *P. vitticeps* genome Pvl1.1 (28) and accompanying gene annotation (http://igagdb.org/dataset/100166) were used to derive gene expression values with Kallisto (67) using a fixed k-mer length of 30 nt. To assess the significance of the difference in female and male gonadal gene expression, a differential expression analysis was performed with DESeq2 (version 1.26.0) on read counts per gene.

The methods for RDA and VisNetwork and empirical validation, chromosome preparation, microdissection and putative sex genes, and DART-seq analysis are described in SI Appendix, Supplementary Text.

**Data Availability.** The annotated DNA sequence of *nr5a1*, the sequence canonical RNA isoform and other relevant data have been deposited in Dryad publicly accessible database (https://doi.org/10.5061/dryad.cm5shqs/5d) (70).

**Acknowledgments.** We thank the members of the wildlife genetics lab of the Institute for Applied Ecology for the many lively discussions on elements presented in this paper. The dragon lizards used in this study were maintained and bred in the capable hands of Wendy Ruscoe and Jacqui Richard. We thank Sarah A. Whiteley and Meghan Castelli for providing tissue samples and genomic DNA for some individuals and insightful advice on protocols. We also thank Shayer Alam for the fluorescent in-situ hybridization (FISH) experiment and Foyez Shams for image analysis. This work was supported by the Australian Research Council Discovery Grant DP170101147 (to C.E.H., J.E.D., T.E., J.A.M.G., A.G. [lead], Stephen Sarre, Lisa Schwanz, and Paul Waters).


