

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

Xiuwen Zhang^{a,1}, Susan Wagner^{a,1}, Clare E. Holleley^{a,b}, Janine E. Deakin^a, Kazumi Matsubara^{a,2}, Ira W. Deveson^{c,d}, Denis O'Meally^{a,3}, Hardip R. Patel^e, Tariq Ezaz^a, Zhao Li^a, Chexu Wang^a, Melanie Edwards^{a,4}, Jennifer A. Marshall Graves^{a,f,5,6}, and Arthur Georges^{a,5,6}

^aInstitute for Applied Ecology, University of Canberra, Bruce, ACT 2617, Australia; ^bAustralian National Wildlife Collection, Commonwealth Scientific and Industrial Research Organisation, Crace, ACT 2911, Australia; ^cKinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia; ^dSchool of Biotechnology and Biomolecular Sciences, Faculty of Science, University of New South Wales, Sydney, NSW 2052, Australia; ^eGenome Sciences Department, John Curtin School of Medical Research, Australian National University, Canberra, ACT 2601, Australia; and ^fSchool of Life Sciences, La Trobe University, Bundoora, VIC 3186, Australia

Contributed by Jennifer A. Marshall Graves; received September 9, 2021; accepted December 3, 2021; reviewed by Manfred Scharl, Changwei Shao, and Andrew Sinclair

Pogona vitticeps has female heterogamety (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.

reptile sex determination | *nr5a1* | chromosome conformation | sex-specific splicing

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/RSPO1/B-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 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.

Author contributions: X.Z., S.W., C.E.H., J.E.D., I.W.D., T.E., J.A.M.G., and A.G. designed research; X.Z., S.W., C.E.H., K.M., T.E., Z.L., and C.W. performed research; X.Z., S.W., C.E.H., J.E.D., I.W.D., D.O., H.R.P., T.E., C.W., M.E., and A.G. analyzed data; X.Z. and S.W. provided supervision and administration; and X.Z., S.W., C.E.H., J.E.D., K.M., I.W.D., J.A.M.G., and A.G. wrote the paper.

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 *dmrt1* and *sox3* 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 SF1 whose action is required for male development in all mammals. We found copies of *nr5a1* on the Z and W that encoded functional SF1. 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 SF1, but ZW females also produced many transcript variants with premature stop codons whose truncated polypeptide products could act as competitive inhibitors of SF1 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 ().

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-seqTM (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: TGCAGCTGTTTTTACGAATCAAAAAAGCTTTTTTGTTC TGCCTATCTTAGCAG[T^W/G^Z]GTGTTTGCCAAA (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 are present on scaffolds 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 I box (*Ftz-FI*), expressed early in *Drosophila* embryogenesis (33), and have

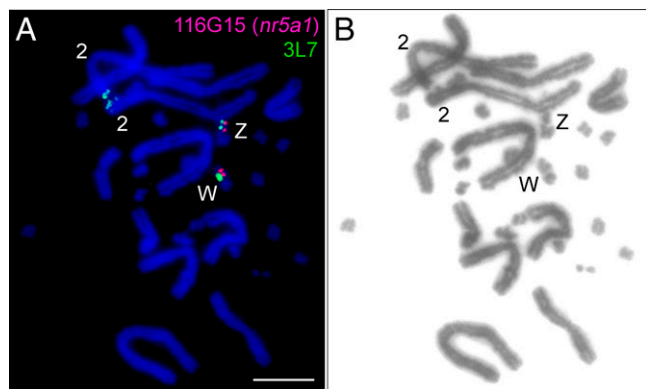


Fig. 1. An *nr5a1*-containing BAC hybridizes to the Z and W chromosomes in a female *P. vitticeps*. (A) BAC 116G15 containing the *nr5a1* gene (red) hybridizes to the opposite end of the W chromosome to the sex chromosome marker BAC 3L7 (green). Note that BAC 3L7 also contains a repetitive sequence that hybridizes to the end of *Pogona* chromosome 2 (68). (B) Inverted DAPI image (stained with 4',6-diamidino-2-phenylindole) of the same metaphase highlighting Z, W, and chromosomes 2. (Scale bar: 10 μm.)

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 fill a gap between exons 4 and 5. For most species, the hinge region is encoded within a larger exon in the middle of the gene (human information is in Fig. 2A), but the model RefSeq transcript generated by National Center for Biotechnology Information (NCBI) for *Pogona* splits it into three exons and defines an intron that spans the assembly gap.

We resolved the genomic sequence of this region by Sanger sequencing. The region has a high GC content (SI Appendix, Fig. S2) harboring four microsatellites (short tandem repeats [STRs]) (Fig. 2C). The four STRs include monomeric (CCCC)_{6–7} repeats, trinucleotide (CAC)_{13–18} repeats, monomeric (GGGG)_{2–6} repeats, and trinucleotide (GGT/A)_{4–7} repeats. The monomeric C and G microsatellite repeats 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 46), 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 their Z 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*).

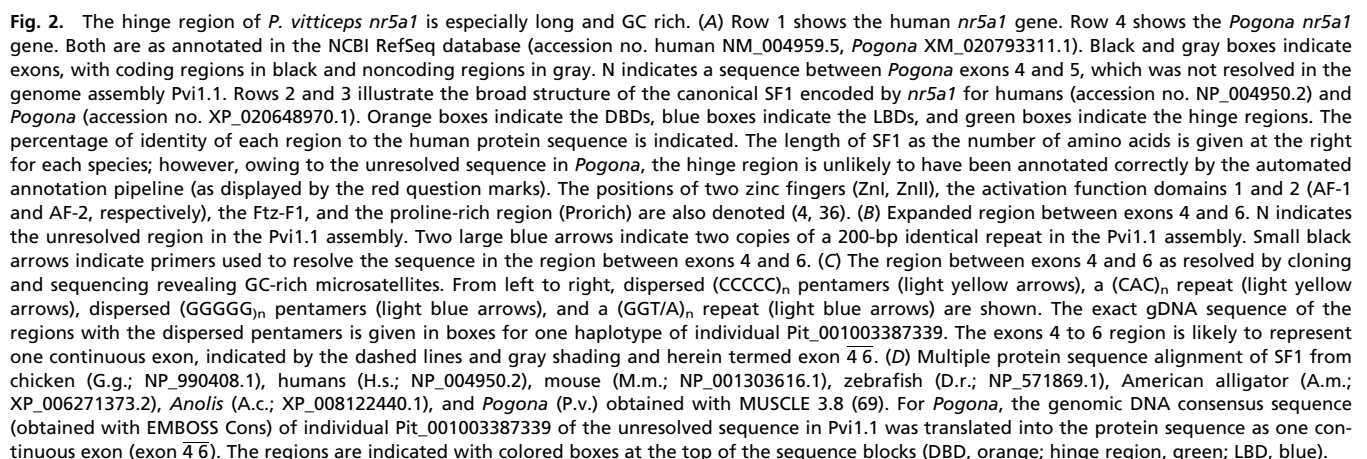
The spectrum of acquired somatic mutations for the *nr5a1* STR region can be categorized into 1) substitutions and indels of a single nucleotide, 2) copy number change in the tandem repeats, 3) recombination between two alleles, and 4) large fragment deletions. Category 1 and 2 mutations occurred with 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 ZW 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*



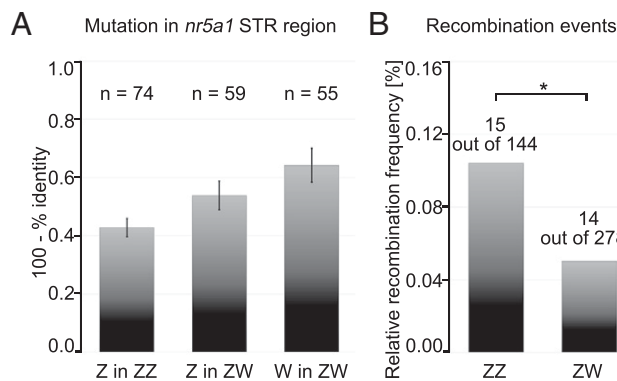


Fig. 3. Genetic instability of the STR region in *Pogona nr5a1*. (A) Average intraindividual rate of nucleotide substitutions, indels, and repeat copy number changes in the STR region of exon 46 of *nr5a1* gDNA for the Z allele of either ZZ individuals or ZW individuals and the W allele of ZW individuals. Average \pm SEM is shown. (B) Fraction of clones with evidence of recombination events observed in the STR region of *nr5a1* between Z and Z in ZZ individuals and Z and W in ZW individuals. *Significance according to the Fisher exact test statistic value of 0.0438.

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 (49).

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 (*nr5a1*_{Δ549}, *nr5a1*_{Δ533}) (Fig. 5A and Table 1), and one ZZ male had an additional isoform (*nr5a1*_{Δ570}). 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 ZW individuals. Each ZW individual had up to five of these additional isoforms (Fig. 5B 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 (*nr5a1*_{ΔE4-6}) 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 (*nr5a1*_{Δ549}, *nr5a1*_{Δ570}) and one isoform recovered only from ZW females (*nr5a1*_{Δ608}) 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*).

*Nr5a1*_{Δ533} 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 (*nr5a1*_{ΔE4-6}) 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.

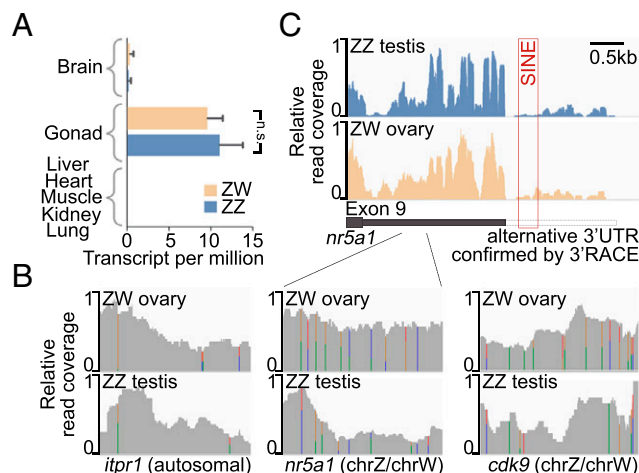


Fig. 4. No difference in female and male gonadal *nr5a1* expression levels shown by RNA-seq expression analysis. (A) *nr5a1* expression in several tissues in ZZ males and ZW females (gonad and brain: $n = 2$; liver, kidney, lung, heart, and muscle: $n = 1$). Adjusted P value = 0.4979. n.s., not significant. (B) Relative read coverage is shown over a 500-bp window of three genes in gonads: *nr5a1* and *cdk9* on the sex chromosomes and the autosomal *itpr1*. Mismatches to the reference, indicating SNPs, are colored according to the substituted base. Bicolor signifies heterozygosity. (C) Read coverage tracks over the 3' end of *nr5a1* in gonads showing two alternative 3' UTR ends. The red box in the 3' UTR indicates an SINE.

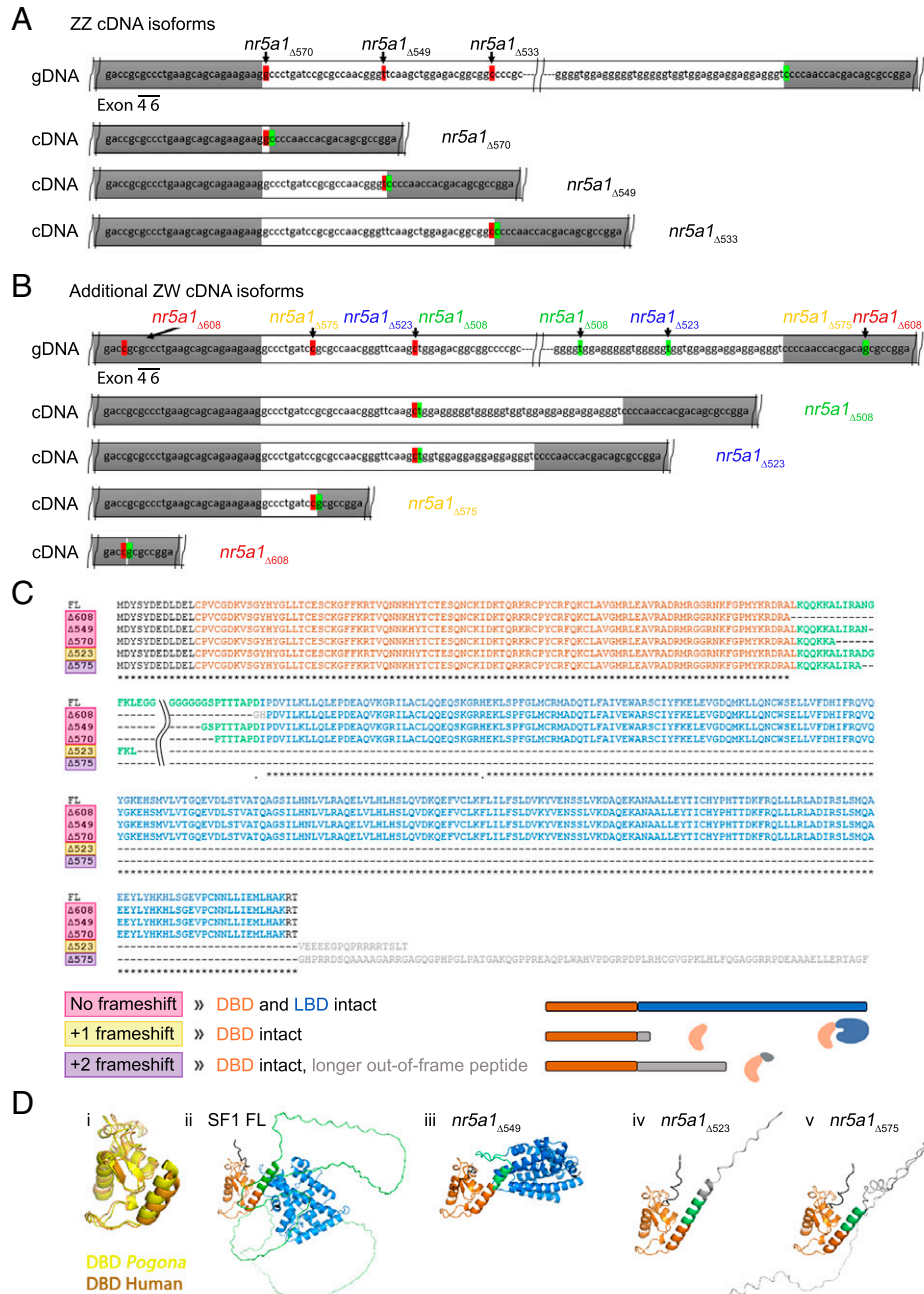


Fig. 5. Sex differential cDNA isoforms of *nr5a1* that differ in the STR region. (A) The gDNA sequence of exon 46 of *nr5a1* and the three cDNA isoforms obtained from ZZ individuals. For orientation, NCBI RefSeq exons 4 and 6 are indicated with gray boxes. The nucleotide highlighted in red is the last 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 EMBOSSE 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). The 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*_{Δ549} is a representative of isoforms with no frameshift. DBD and LBD fold correctly. (D, iv) *nr5a1*_{Δ523} is a representative of isoforms producing a +1 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*.

Table 1. Different *nr5a1* cDNA variants identified in male and female *Pogona* individuals

Gender and individual identification	cDNA variants
ZZ	
Pit_001003379128	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ533} *
Pit_001003344020	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ533} * <i>nr5a1</i> _{Δ570} *
Pit_001003351790	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ533} *
Pit_001003344962	<i>nr5a1</i> _{Δ549v} * <i>nr5a1</i> _{Δ533} *
Pit_005005003628	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ533} *
ZW	
Pit_005005003514	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ520} <i>nr5a1</i> _{Δ526} <i>nr5a1</i> _{Δ568} <i>nr5a1</i> _{Δ703} <i>nr5a1</i> _{Δ881}
Pit_005005003443	<i>nr5a1</i> _{Δ533} * <i>nr5a1</i> _{Δ537} <i>nr5a1</i> _{Δ604} <i>nr5a1</i> _{Δ640} <i>nr5a1</i> _{Δ679}
Pit_001003182571	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ533} * <i>nr5a1</i> _{Δ570} * <i>nr5a1</i> _{Δ523} <i>nr5a1</i> _{Δ608}
Pit_001003342982	<i>nr5a1</i> _{Δ549} * <i>nr5a1</i> _{Δ533} * <i>nr5a1</i> _{Δ570} * <i>nr5a1</i> _{Δ508} <i>nr5a1</i> _{Δ523} <i>nr5a1</i> _{Δ575} <i>nr5a1</i> _{Δ608}

Variants result from alternative splicing in the STR region (within exon 4–6). Isoforms are named according to the number of missing nucleotides in relation to the genomic DNA consensus sequence of individual Pit_001003387339 (exon 4–6 has an assembly gap in the genome assembly Pvi1.1, which we have resolved in this study).

*Isoforms found in both male and female individuals.

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 4–6, 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. Two 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 development of ZW embryos. A similar competitive inhibition model has been advanced to explain the female-determining action of a truncated *dmrt1* gene on the W chromosome in *X. laevis* (9). Thus, *nr5a1* may control sexual fate in *P. vitticeps* through dosage operating at a posttranscription level.

Formally, therefore, we propose that W-*nr5a1* acts as female dominant in a GSD system. However, in the absence of genomic difference between Z-*nr5a1* and W-*nr5a1*, the mechanism of action of this gene in determining sex is epigenetic rather than genetic. Previously, DNA methylation differences have been invoked to explain the silencing of *Dmrt1* in ZW female half-smooth tongue sole (58). In the absence of genomic differences between Z-borne and W-borne *nr5a1* alleles, we must look for epigenetic factors that might differentiate them in the landscape of the Z and W chromosomes. The most obvious difference is the greater repeat content of the W chromosome. We hypothesize that the W chromosome is held in an altered conformation, which

lowers homologous pairing and recombination, makes 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.

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 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

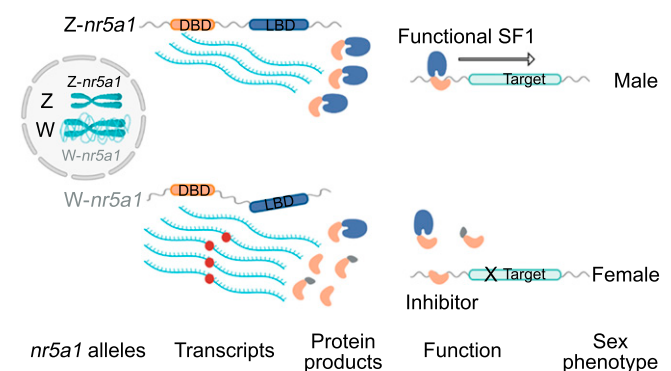


Fig. 6. Epigenetic control of sex determination by posttranscriptional control of *nr5a1* in *P. vitticeps*. The Z and W chromosomes have the same gene content, but the repeat-rich W assumes an altered conformation. This renders the STR region in the middle of W-*nr5a1* genetically unstable and affects transcription and secondary structure of the primary transcript. This affects splicing, leading to many isoforms with premature stop codons (red circles) that are translated into truncated SF1 polypeptides that have the DBD (orange) but lack the LBD (blue) or that bear a small part at the DBD terminus that is out of frame (gray). These truncated polypeptides act as competitive inhibitors of SF1 function, blocking the binding of full-length SF1 and the activation of target genes via the LBD. Suppression of SF1 leads to female development of ZW animals. The figure was created with Biorender.

the reads to the draft *Pogona* assembly Pvi1.1. Bedtools [v2.27.1 (62)] was used to generate bed graphs.

Molecular Cloning and Sequencing of the GC-Rich Microsatellite Region of *nr5a1*. Genomic DNA was isolated with the DNeasy Blood & Tissue Kits (QIAGEN). 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 μ L 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 confirmed to amplify specifically *nr5a1* by blast against the *P. vitticeps* genome. PCRs were carried out using MyFi DNA polymerase (Bio-21117; Bioline; Australia Pty Ltd.) under the following conditions: 50 to 200 ng of genomic DNA, 10 μ M each primer, 2.0 units of MyFi DNA polymerase, and 5 μ L of 5 \times buffer supplemented in a final volume of 25 μ L. 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 (QIAGEN), directly sequenced (Macrogen) or cloned into pCR4-TOPO™ 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 16 \times), and call rate (threshold 0.95). A relatedness matrix (G) was calculated with gl.glm 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.glm.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 ClustaW in Geneious sequence analysis software (Geneious version 8.1 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 only 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 SNPs/indels of gDNA clones of the daughters) were compared with those of the mother, and the allele group with the matching set of SNPs/indels to one of the mother allele groups was identified as W linked, while the other allele group with SNPs/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 SNPs/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* cDNA 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 μ g RNA in a total volume of 20 μ L using SuperScript IV reverse transcriptase (Invitrogen; ThermoFisher Scientific) and random hexamer oligo primers (Invitrogen; ThermoFisher Scientific) following the manufacturer's guidelines. Subsequently, 1 μ L of the cDNA reaction was used to perform PCR in a 25- μ L reaction mix with 2.0 units of Taq Red DNA polymerase (Bioline Australia Pty Ltd.), 10 μ M each primer, and 5 μ L of 5 \times 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* cDNA Variants. The three-dimensional structure of the putative translated protein of *nr5a1* cDNA variants was predicted with AlphaFold (66).

3' RACE. The 3' end of *nr5a1* cDNA was determined by the 3' RACE PCR method as detailed in the manufacturer's protocols (Gene Racer kit; Invitrogen; ThermoFisher Scientific). cDNA was synthesized by incubating 2 μ g of total RNA with reverse transcriptase SuperScript IV (Invitrogen; ThermoFisher Scientific) for 40 min at 50°C using the Gene Racer oligo dT primer containing a 3' adaptor. The subsequent two RACE PCR reactions were carried out using gene-specific primers (Sf3RC1 and Sf3RC2), GeneRacer 3' Primer, and GeneRacer 3' Nested Primer at an annealing temperature of 60°C. The RT-PCR and 3' RACE PCR products were cloned and sequenced using the method described above. Sequences obtained from RT-PCR were aligned to the gDNA sequence using Geneious.

***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 Pvi1.1 (28) and accompanying gene annotation (<http://gigadb.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-DSC, subtraction 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.cnp5hqc5d>) (70).

ACKNOWLEDGMENTS. We thank the members of the wildlife genetics laboratory 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 Richardson. 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 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).

1. P. Koopman, A. Münsterberg, B. Capel, N. Vivian, R. Lovell-Badge, Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348**, 450–452 (1990).
2. A. H. Sinclair et al., A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240–244 (1990).

3. Y. Kim, B. Capel, Balancing the bipotential gonad between alternative organ fates: A new perspective on an old problem. *Dev. Dyn.* **235**, 2292–2300 (2006).
4. M. P. Mello, E. S. França, H. C. Fabbri, A. T. Maciel-Guerra, G. Guerra-Júnior, Multifunctional role of steroidogenic factor 1 and disorders of sex development. *Arq. Bras. Endocrinol. Metabol.* **55**, 607–612 (2011).

5. R. Sekido, R. Lovell-Badge, Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930–934 (2008).
6. L. Lin, J. C. Achermann, Steroidogenic factor-1 (SF-1, Ad4BP, NR5A1) and disorders of testis development. *Sex Dev.* **2**, 200–209 (2008).
7. C. A. Smith *et al.*, The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).
8. L. S. Lambeth *et al.*, Over-expression of DMRT1 induces the male pathway in embryonic chicken gonads. *Dev. Biol.* **389**, 160–172 (2014).
9. S. Yoshimoto *et al.*, A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2469–2474 (2008).
10. Y. Nagahama, T. Chakraborty, B. Paul-Prasanth, K. Ohta, M. Nakamura, Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiol. Rev.* **101**, 1237–1308 (2021).
11. E. E. Montiel, D. Badenhorst, J. Tamplin, R. L. Burke, N. Valenzuela, Discovery of the youngest sex chromosomes reveals first case of convergent co-option of ancestral autosomes in turtles. *Chromosome Res.* **126**, 105–113 (2017).
12. L. Lee, E. E. Montiel, B. M. Navarro-Domínguez, N. Valenzuela, Chromosomal rearrangements during turtle evolution altered the syntax of genes involved in vertebrate sex determination. *Cytogenet. Genome Res.* **157**, 77–88 (2019).
13. D. O'Meally *et al.*, Non-homologous sex chromosomes of birds and snakes share repetitive sequences. *Chromosome Res.* **18**, 787–800 (2010).
14. J. F. Hughes, H. Skaletsky, D. C. Page, Sequencing of rhesus macaque Y chromosome clarifies origins and evolution of the DAZ (deleted in AZoospermia) genes. *BioEssays* **34**, 1035–1044 (2012).
15. H. Skaletsky *et al.*, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
16. Y. Q. S. Soh *et al.*, Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**, 800–813 (2014).
17. B. M. Skinner *et al.*, The pig X and Y chromosomes: Structure, sequence, and evolution. *Genome Res.* **26**, 130–139 (2016).
18. J. F. Hughes *et al.*, Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
19. M. Tomaszewicz, P. Medvedev, K. D. Makova, Y and W chromosome assemblies: Approaches and discoveries. *Trends Genet.* **33**, 266–282 (2017).
20. T. Gamble, D. Zarkower, Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol. Ecol. Resour.* **14**, 902–913 (2014).
21. T. Gamble, Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Mol. Ecol.* **25**, 2114–2116 (2016).
22. D. S. B. Dissanayake *et al.*, Identification of Y chromosome markers in the eastern three-lined skink (*Bassiana duperreyi*) using *in silico* whole genome subtraction. *BMC Genomics* **21**, 667 (2020).
23. S. Mahajan, K. H. C. Wei, M. J. Nalley, L. Gibilisco, D. Bachtrog, De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol.* **16**, e2006348 (2018).
24. Y. Yano, T. Chiba, H. Asahara, Analysis of the mouse Y chromosome by single-molecule sequencing with Y chromosome enrichment. *Front. Genet.* **11**, 406 (2020).
25. L. F. K. Kuderna *et al.*, Selective single molecule sequencing and assembly of a human Y chromosome of African origin. *Nat. Commun.* **10**, 4 (2019).
26. T. Ezaz *et al.*, The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. *Chromosome Res.* **13**, 763–776 (2005).
27. J. H. Luo *et al.*, Differential subtraction chain, a method for identifying differences in genomic DNA and mRNA. *Nucleic Acids Res.* **27**, e24 (1999).
28. A. Georges *et al.*, High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *Gigascience* **4**, 45 (2015).
29. A. Kilian *et al.*, “Diversity arrays technology: A generic genome profiling technology on open platforms” in *Data Production and Analysis in Population Genomics: Methods and Protocols*, F. Pompanon, A. Bonin, Eds. (Humana Press, 2012), pp. 67–89.
30. C. Sansaloni *et al.*, Diversity arrays technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proc.* **5**, 54 (2011).
31. J. E. Deakin *et al.*, Anchoring genome sequence to chromosomes of the central bearded dragon (*Pogona vitticeps*) enables reconstruction of ancestral squamate macrochromosomes and identifies sequence content of the Z chromosome. *BMC Genomics* **17**, 447 (2016).
32. T. Ezaz *et al.*, Sequence and gene content of a large fragment of a lizard sex chromosome and evaluation of candidate sex differentiating gene R-spondin 1. *BMC Genomics* **14**, 899 (2013).
33. G. Lavorgna, H. Ueda, J. Clos, C. Wu, FTZ-F1, a steroid hormone receptor-like protein implicated in the activation of fushi tarazu. *Science* **252**, 848–851 (1991).
34. R. J. Agate, W. R. Perlman, A. P. Arnold, Cloning and expression of zebra finch (*Taeniopygia guttata*) steroidogenic factor 1: Overlap with hypothalamic but not with telencephalic aromatase. *Biol. Reprod.* **66**, 1127–1133 (2002).
35. K. Kawano, I. Miura, K. Morohashi, M. Takase, M. Nakamura, Molecular cloning and expression of the SF-1/Ad4BP gene in the frog, *Rana rugosa*. *Gene* **222**, 169–176 (1998).
36. M. Wong, M. S. Ramayya, G. P. Chrousos, P. H. Driggers, K. L. Parker, Cloning and sequence analysis of the human gene encoding steroidogenic factor 1. *J. Mol. Endocrinol.* **17**, 139–147 (1996).
37. T. Kudo, S. Sutou, Molecular cloning of chicken FTZ-F1-related orphan receptors. *Gene* **197**, 261–268 (1997).
38. M. W. Kuo *et al.*, Gene duplication, gene loss and evolution of expression domains in the vertebrate nuclear receptor NR5A (Ftz-F1) family. *Biochem. J.* **389**, 19–26 (2005).
39. D. S. Lala, D. A. Rice, K. L. Parker, Steroidogenic factor I, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu-factor I. *Mol. Endocrinol.* **6**, 1249–1258 (1992).
40. T. Nakajima, M. Takase, I. Miura, M. Nakamura, Two isoforms of FTZ-F1 messenger RNA: Molecular cloning and their expression in the frog testis. *Gene* **248**, 203–212 (2000).
41. K. Oba *et al.*, Structural characterization of human Ad4bp (SF-1) gene. *Biochem. Biophys. Res. Commun.* **226**, 261–267 (1996).
42. K. L. Parker, B. P. Schimmer, Steroidogenic factor 1: A key determinant of endocrine development and function. *Endocr. Rev.* **18**, 361–377 (1997).
43. J. von Hofsten, P. E. Olsson, Zebrafish sex determination and differentiation: Involvement of FTZ-F1 genes. *Reprod. Biol. Endocrinol.* **3**, 63 (2005).
44. P. S. Western, J. L. Harry, J. A. Marshall Graves, A. H. Sinclair, Temperature-dependent sex determination in the American alligator: Expression of SF1, WT1 and DAX1 during gonadogenesis. *Gene* **241**, 223–232 (2000).
45. S. Tantawy *et al.*, Analysis of the gene coding for steroidogenic factor 1 (SF1, NR5A1) in a cohort of 50 Egyptian patients with 46,XY disorders of sex development. *Eur. J. Endocrinol.* **170**, 759–767 (2014).
46. Q. Ou *et al.*, The DEAD box protein DP103 is a regulator of steroidogenic factor-1. *Mol. Endocrinol.* **15**, 69–79 (2001).
47. T. H. Little *et al.*, Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: A helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol. Endocrinol.* **20**, 831–843 (2006).
48. I. W. Deveson *et al.*, Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* **3**, e1700731 (2017).
49. R. A. Elbarbary, B. A. Lucas, L. E. Maquat, Retrotransposons as regulators of gene expression. *Science* **351**, aac7247 (2016).
50. L. E. Maquat, Nonsense-mediated mRNA decay. *Curr. Biol.* **12**, R196–R197 (2002).
51. A. Damianov, D. L. Black, Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA* **16**, 405–416 (2010).
52. A. E. Quinn *et al.*, Temperature sex reversal implies sex gene dosage in a reptile. *Science* **316**, 411 (2007).
53. S. L. Whiteley *et al.*, Two transcriptionally distinct pathways drive female development in a reptile with both genetic and temperature dependent sex determination. *PLoS Genet.* **17**, e1009465 (2021).
54. C. E. Holleley *et al.*, Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. *Nature* **523**, 79–82 (2015).
55. E. Buschiazio, N. J. Gemmell, The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* **28**, 1040–1050 (2006).
56. D. A. Kiktev, Z. Sheng, K. S. Lobachev, T. D. Petes, GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7109–E7118 (2018).
57. A. R. Iglesias, E. Kindlund, M. Tammi, C. Wadelius, Some microsatellites may act as novel polymorphic cis-regulatory elements through transcription factor binding. *Gene* **341**, 149–165 (2004).
58. C. Shao *et al.*, Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res.* **24**, 604–615 (2014).
59. E. Agirre, A. J. Oldfield, N. Bellora, A. Segelle, R. F. Luco, Splicing-associated chromatin signatures: A combinatorial and position-dependent role for histone marks in splicing definition. *Nat. Commun.* **12**, 682 (2021).
60. K. Kusumi *et al.*, Developing a community-based genetic nomenclature for anole lizards. *BMC Genomics* **12**, 554 (2011).
61. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
62. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
63. L. Y. Li *et al.*, A primer design strategy for PCR amplification of GC-rich DNA sequences. *Clin. Biochem.* **44**, 692–698 (2011).
64. B. Gruber, P. J. Unmack, O. F. Berry, A. Georges, dartr: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* **18**, 691–699 (2018).
65. J. B. Endelman, Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255 (2011).
66. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
67. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
68. M. J. Young, D. O'Meally, S. D. Sarre, A. Georges, T. Ezaz, Molecular cytogenetic map of the central bearded dragon, *Pogona vitticeps* (Squamata: Agamidae). *Chromosome Res.* **21**, 361–374 (2013).
69. F. Madeira *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47** (W1), W636–W641 (2019).
70. X. Zhang *et al.*, Sex-specific splicing of Z- and W-borne nr5a1 alleles suggests sex determination is controlled by chromosome conformation. *Dryad*. 10.5061/dryad.cnp5hqc5d. Deposited 7 September 2021.