

# Characterization, chromosomal location, and genomic neighborhood of a ratite ortholog of a gene with gonadal expression in mammals

Daniel E. Janes,<sup>1,\*</sup> Tariq Ezaz,<sup>†</sup> Jennifer A. Marshall Graves,<sup>†</sup> and Scott V. Edwards\*

\*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA;

<sup>†</sup>Comparative Genomics Group, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra ACT 2601, Australia

**Synopsis** A locus that we name SubA was discovered during large-scale sequencing and characterization of a bacterial artificial chromosome library from an emu, *Dromaius novaehollandiae*. This locus yields a significantly negative Tajima's D in emus and is conserved across emu, chicken, mouse, and human. Expression of SubA orthologs has been reported in human ovaries and in mouse testes, but remains unknown in emus. The locus was physically mapped onto a pair of microchromosomes in emus by fluorescent *in situ* hybridization and also in chicken as previously reported. By characterizing emu SubA in this article, we aim to improve current descriptions of the cascade of genes associated with avian sex differentiation. Future experimentation will report the expression of SubA in ratites, other birds, and nonavian reptiles.

## Introduction

Genes are frequently discovered and characterized with the aid of novel genomic technology and resources. As a result, developmental biology, among other fields, has taken great strides recently (Canestro et al. 2007). For example, bacterial artificial chromosome (BAC) libraries have been generated for several vertebrates that allow characterization of the evolutionary history of typically mammalian, avian, and nonavian reptilian traits. In this article, we describe SubA, a gene discovered in a genomic library for the emu (*Dromaius novaehollandiae*) that is highly conserved across birds and mammals, and appears to be involved in gonadal development (Wheeler et al. 2003).

As part of an ongoing study (Janes et al., manuscript in preparation), several BAC clones were randomly selected from a genomic library for the emu to map to chromosomes and to characterize the population genetics of sex chromosomes and autosomes. The emu is a valuable species with which to study these genomic compartments, because emus, like all paleognaths, bear similarity to many reptiles in that their sex chromosomes are virtually homomorphic. Homomorphic sex chromosomes have been characterized in genotypically sex-determined

turtles but birds universally exhibit cytogenetically distinguishable sex chromosomes (Ewert et al. 2004). As part of a study of the evolution of heteromorphism of sex chromosomes, we discovered an emu gene that appears to be expressed in a sex-biased pattern in gonadal and other tissues of other vertebrates (Wheeler et al. 2003). The genetics of sex differentiation can be studied via sex-biased expression of large series of genes as measured by microarray and suppressive subtractive hybridization techniques (Diatchenko et al. 1996; Ellegren and Parsch 2007). Generation and annotation of novel genomic sequences offer an additional option for discovery of genes associated with sex determination. As the cost of sequencing continues to decrease, more genes will likely be reported by an approach similar to that taken by this study.

## Materials and methods

### Fluorescent *in situ* hybridization (FISH) mapping of BAC clone

We used a publicly available BAC library (www.sym-bio.com) as a starting point for our study. As part of a larger study, several BAC clones were randomly selected from an emu genomic library. To determine the chromosomal location of these

From the symposium "Reptile Genomics and Evolutionary Genetics" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 3–6, 2008, at San Antonio, Texas.

<sup>1</sup>E-mail: djanes@oeb.harvard.edu

*Integrative and Comparative Biology*, pp. 1–7

doi:10.1093/icb/icn024

clones, BAC insert #Drn194e24 was isolated and mapped to emu chromosome spreads by FISH. Metaphase chromosome spreads were prepared from emu fibroblasts and slides were prepared as described by Ezaz et al. (2005). BAC clone #Drn194e24 was grown overnight in 15 ml of Difco LB broth at 37°C with shaking and purified as previously described (Shedlock et al. 2008). The insert was labeled with Biotin-16-dUTP by nick translation following the protocol as described by Ezaz et al. (2005). BAC DNA was precipitated overnight at -20°C with 2.5 volumes of 100% ethanol and 20 µg glycogen (as carrier), pelleted and air-dried, then resuspended in 20 µl hybridization buffer (50% formamide, 10% dextran sulphate, 2× SSC, 40 mM sodium phosphate pH 7.0, and 1× Denhardt's solution) and incubated at 37°C for 30 min. Drn194e24 probe was denatured at 70°C for 10 min, quenched on ice for 2 min, and preannealed at 37°C for 15 min (Shetty et al. 2002). Metaphase chromosome slides from emus were denatured for 2 min at 70°C, dehydrated by ethanol series, and air-dried for 30 min. Eighteen microliters of each probe were added onto each denatured metaphase chromosome slide and hybridized overnight at 37°C and washed twice each (50% 2× SSC, 50% formamide at 45°C, followed by 0.2× SSC at 45°C). Probes were detected with Avidin, conjugated with FITC for 45 min, washed at room temperature, air-dried, and mounted with anti-fade medium Vectashield (Vector Laboratories, Burlingame, CA) containing 1.5 µg/ml DAPI. Images were captured using a Zeiss Axioplan epifluorescence microscope equipped with a CCD (charge-coupled device) camera (RT-Spot, Diagnostic Instrument, Sterling Heights, MI) using either filters 02, 10, or 15 from the Zeiss fluorescence filter set or the Pinkel filter set (filter set 8300, Chroma technologies, Rockingham, VT). IPLab scientific imaging software (V.3.9, Scanalytics Inc., Rockville, MD) was used to capture and super-impose gray-scale images to produce colored images.

### Sequencing BAC DNA

DNA from Drn194e24 was isolated, hydrosheared, subcloned, and sequenced as described by Shedlock et al. (2008). Subclone sequences were inspected for quality by Phred software, assembled into contigs by Phrap, and visualized with Consed (Gordon et al. 1998). Contigs produced from the BAC insert were queried for Refseq genes using the BLAT function within the University of California at Santa Cruz (UCSC) genome browser (Kent 2002; Pruitt et al. 2005). With the UCSC genome browser, emu BAC sequence was queried individually against the most

recent chicken (*Gallus gallus*; Build 2.1), mouse (*Mus musculus*; Build 37.1), and human (*Homo sapiens*; Build 36.1) assemblies. Identified Refseq orthologs were aligned to Drn194e24 and visualized with the UCSC genome browser.

### Polymorphism and estimation of Tajima's *D* and Fu and Li's *D*

DNA samples were collected from 17 wild-caught emus from Western Australia ( $n=8$ ), Queensland ( $n=2$ ), South Australia ( $n=4$ ), and New South Wales ( $n=3$ ). Primers for SubA, a Drn194e24-linked locus, were designed with Oligo Primer Analysis Software (Molecular Biology Insights, Cascade, CO). Amplifications of SubA were performed with primers F: TTCTTTAGGGCATAGCATAGG and R: AGCACTTTGCCGGTAA using an initial denaturation step at 94°C for 2 min, followed by 35 cycles of 94°C for 40 s, 58.5°C for 1 min, and 72°C for 1 min, followed by a final step of 72°C for 5 min. Amplified fragments were purified using Multiscreen PCR<sub>µ96</sub> Filter Plates (Millipore, Billerica, MA) and sequenced directly using BigDye<sup>TM</sup> Terminator Cycle Sequencing chemistry with original primers (Applied Biosystems, Foster City, CA). Sequences were recorded with an ABI3100 automated sequencing instrument (Applied Biosystems). Forward and reverse sequences from the locus were reconciled with each other using Sequencher v. 4.5 (Gene Codes, Ann Arbor, MI). Sequences were aligned among the 17 individuals and haplotypes were estimated by PHASE (Stephens et al. 2001). PHASE outfiles were converted to Nexus files with python script (C. Chapus, personal communication) and analyzed by DNAsp (Rozas et al. 2003) for calculation of Tajima's *D* and Fu and Li's *D* (Fu and Li 1993; Tajima 1989). Both Tajima's *D* and Fu and Li's *D* can be used to detect deviations of the site frequency distribution from those expected from a neutral panmictic population. Such deviations can suggest positive selection, but can also have a demographic interpretation (Loewe and Charlesworth 2007).

### Estimation of repeat densities

In order to compare the composition of our emu BAC with that for chicken, repeat (retroelement) densities were estimated for chicken macrochromosomes and microchromosomes, and all emu BAC sequences currently available from Genbank. Chicken macrochromosomal sequence data were collected from chromosomes 1–5 (2.5 Mb/chromosome) and microchromosomal sequences were collected from chromosomes 15, 17–20 (2.5 Mb/chromosome). Chicken chromosome 16 was not analyzed because

2.5 Mb contiguous sequence are not, at present, available from that chromosome. All available emu BAC sequences (2.47 Mb) were also collected from Genbank. We used repeatmasker software (Smit et al. 2004) to calculate average GC content, simple repeats, low complexity repeats, and CR1 repeats for chicken macrochromosomal and microchromosomal sequences, all available emu BAC sequence, and emu BAC insert #Drn194e24, from which we were able to shotgun sequence a total of 70 kb.

### Comparison of expression data

To estimate SubA function, orthologs were referenced with the UCSC genome browser and the National Center for Biotechnology Information (NCBI) UniGene Expression Profile Viewer. Genomic locations and expression profiles of orthologs were reported for chicken (*G. gallus*), mouse (*M. musculus*), and human (*H. sapiens*).

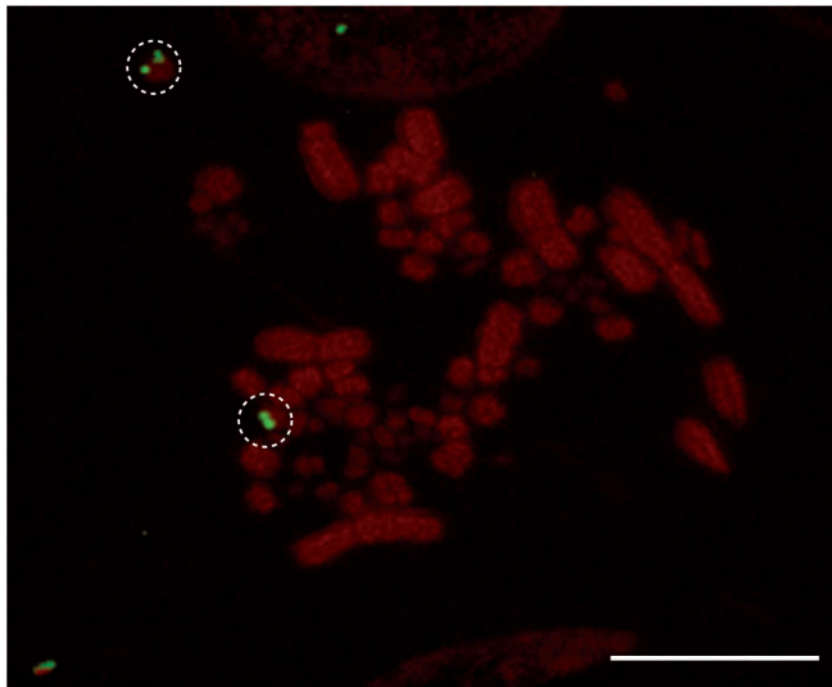
## Results

### Physical mapping

Emu BAC insert #Drn194e24 hybridized onto a pair of microchromosomes in metaphase chromosome spreads of emus (Fig. 1). The microchromosomal location of the sequence is noteworthy because, among other reasons, its ortholog is microchromosomal in chicken as well (Table 1).

### Sequence analysis

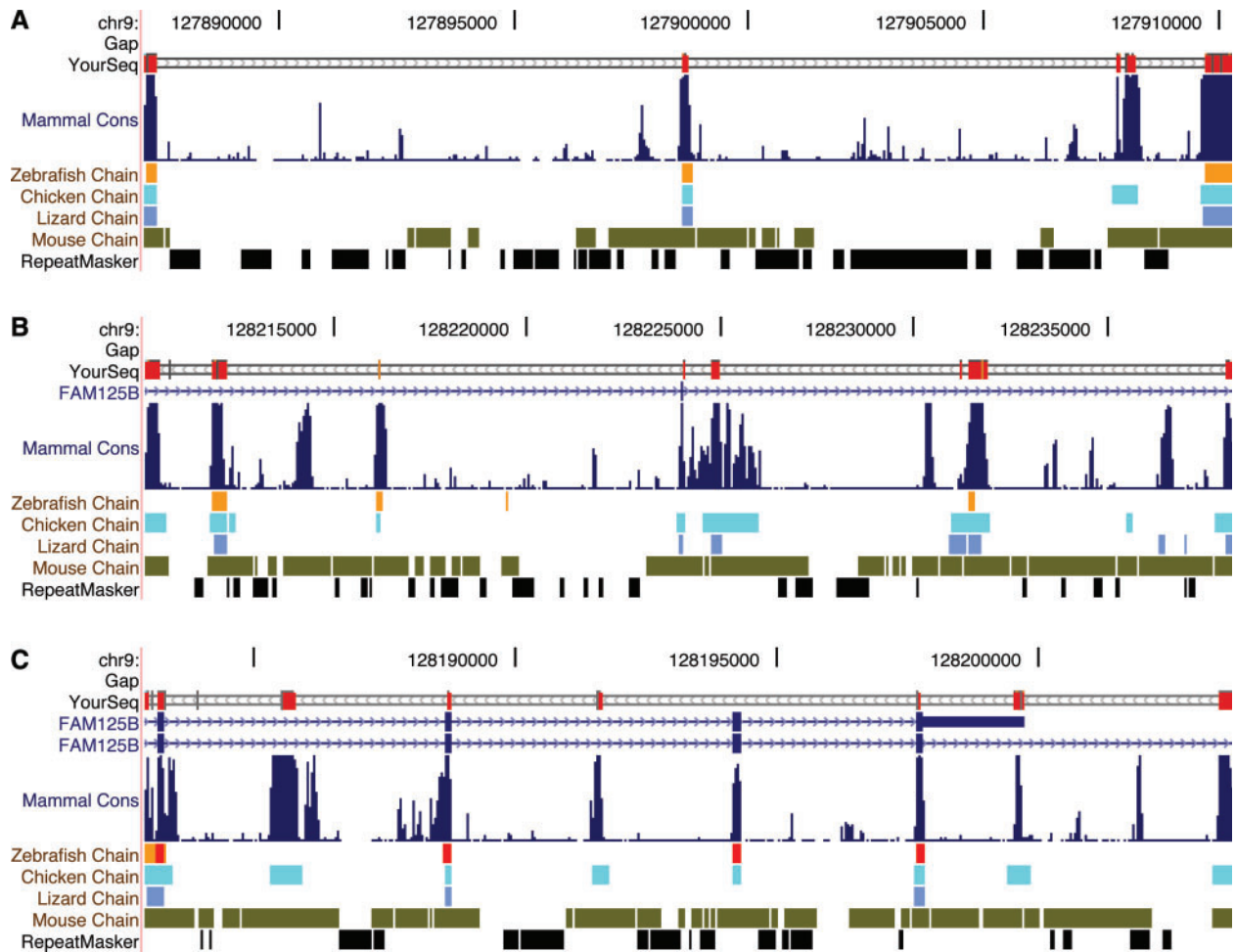
The shotgun subclones of BAC insert #Drn194e24 yielded 4200 sequences and were assembled by Phrap software into several contigs. The two largest contigs from Drn194e24 yielded ~70 kb of high-quality sequence (Phred scores of each base  $\geq 20$ ) and have been submitted to Genbank as two unordered pieces with Accession #EU200931. Alignments from the UCSC genome browser show conservation of



**Fig. 1** Bacterial artificial chromosome insert #Drn194e24 from the emu library maps to a pair of microchromosomes on emu metaphase chromosome spreads. White dotted circles indicate the hybridized Drn194e24 probes and scale bar represents 10  $\mu$ m.

**Table 1** Sequence generated from emu bacterial artificial chromosome insert #Drn194e24 is orthologous to a gene found in human, mouse, and chicken. Expression data are not yet available for chickens and gonadal expression was inconsistent between humans and mice

Library	Gene	Sequence	Description	Strand	Genomic Size	Ovary	Testis
Human	FAM125B	chr9:128,128,949-128,309,140	Hypothetical protein LOC89853	+	180192	+	-
Mouse	2610528K11Rik	chr2:33,585,476-33,743,466	Hypothetical protein LOC72543	-	157991	-	+
Chicken	LOC417092	chr17:10,690,052-10,738,802	Hypothetical protein LOC417092	+	48751	NA	NA



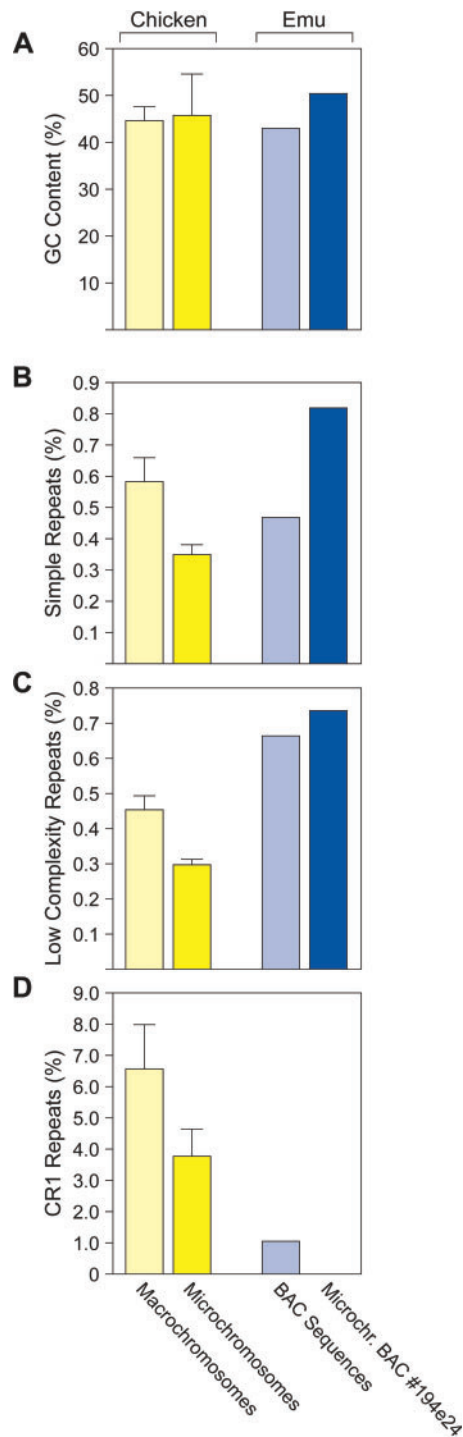
**Fig. 2** UCSC genome browser output displays repeat density and homology among sequence from emu bacterial artificial chromosome insert #Drn194e24 and orthologs from five other vertebrate genomes (human, zebrafish, chicken, anole, and mouse). The human genome is represented by the base numbers on chromosome 9 on the top of each figure. A–C represent three subsections of BAC insert #Drn194e24. Note the order between sections (B) and (C) is inverted in the human genome.

Drn194e24 and human Refseq gene FAM125B, a 4815 bp cDNA consisting of multiple open reading frames (Hattori et al. 2000). Elements within Drn194e24 are also conserved among mouse, chicken, anole (*Anolis carolinensis*), and zebrafish (*Danio rario*; Fig. 2). Many of these elements appear to be exons separated by less conserved introns or noncoding regions. PCR successfully amplified a 292 bp fragment of Drn194e24 that we have named SubA. SubA has significantly low Tajima's  $D$  ( $-2.03,143$ ;  $P < 0.05$ ) and Fu and Li's  $D$  ( $-3.23,469$ ;  $P < 0.02$ ). We did not find any additional major gene hits for this BAC clone using the genome browser or nucleotide blast in Genbank—a surprising result given the generally high gene densities in birds. NCBI's Unigene Expression Profile Viewer and the UCSC genome browser indicate expression of SubA orthologs in human ovaries (but not testes) and mouse testes (but not ovaries).

Data on expression are not yet available for SubA's chicken ortholog (Table 1).

### Estimation of repeat densities

In this sample, variation in GC content shows no strong pattern across chicken macro- and microchromosomal sequences, all available emu BAC sequence, and Drn194e24, although it should be noted that previous reports indicate greater CpG island richness in microchromosomes than in macrochromosomes in chickens (Auer et al. 1987; McQueen et al. 1996; ICGSC 2004). In our sample of genomic sequence, simple and low complexity repeats are more prevalent in the emu sequence than in either chicken macrochromosomal or microchromosomal sequences, and CR1 repeats are less prevalent in emu sequence than in chicken sequence. Repeatmasker results suggest a relationship between Drn194e24 and overall emu BAC sequences that is different from the



**Fig. 3** Average GC content and repeat densities (**A**, GC content; **B**, simple repeats; **C**, low complexity repeats; **D**, CR1 repeats) of chicken and emu sequences. The first two bars display average repeat densities from 12.5 Mb of sequence from chicken macrochromosomes 1–5 and 12.5 Mb of sequence from chicken microchromosomes 15, 17–20 (2.5 Mb/chromosome). The third bar shows total values for all BAC sequences from the emu library that are available in Genbank (2.47 Mb). The fourth bar shows total values for the emu BAC insert #Drn194e24 that contains mapped microchromosomal sequence (70 kb). Repeat densities were calculated by the Repeatmasker package.

relationship between reported chicken macrochromosomal and microchromosomal sequences. In our sample, chicken repeat densities are lower in microchromosomal than in macrochromosomal sequence as expected (ICGSC 2004), but Drn194e24 has greater repeat density than does the overall emu BAC sample that very likely represents macrochromosomal as well as microchromosomal sequences (Fig. 3). The significance of these differences is unclear but similarities in genomic landscape may reflect some structural similarities between microchromosomes and macrochromosomes, or an inability to distinguish these compartments, given our small sampling of the emu genome.

## Discussion

SubA is a microchromosomal locus from the emu that was discovered from a subcloned, partially assembled insert from an emu BAC library. This gene joins 18S–28S RNA genes and a painting probe from chicken chromosome 4 on a short list of markers that have been mapped to microchromosomes in the emu (Nishida-Umehara et al. 2007). The gene is tentatively named SubA because it was the first locus identified from a subcloning effort involving emus in our laboratory. Future studies will ascribe function to the gene and it should be renamed with consideration of its function (Janes et al., manuscript in preparation). We predict that the gene will be expressed in gonads of emus, other birds, and nonavian reptiles, although in view of its apparently labile expression profile in mammals, we withhold expectation of male or female biases in expression.

The gene has a suggestion of positive selection in the form of significant Tajima's and Fu and Li's  $D_s$ , is conserved across birds and mammals, and appears to be expressed in a sex-biased pattern; its genomic neighborhood aligned well with orthologs from chicken, mouse, and human (Fig. 2). Aside from the single human refseq ortholog detected in this region (FAM125B), the region in emus otherwise appear devoid of genes. With regard to the genomic landscape of SubA, the GC content results do not match previous reports of microchromosomal GC richness in birds. However, GC in the form of CpG islands is highly localized to break-prone regions of chromosomes and chicken chromosomes 15 and 17–20 have not been sampled by our method previously (Gordon et al. 2007). It is possible that our sample did not include break-prone regions and therefore highly localized GC concentrations were not detected. Future research will investigate

break-prone regions in chromosomes 15 and 17–20 as well as measure expression of SubA.

The suggestion of positive selection at SubA is tentative, since deviations from a neutral site-frequency spectrum can be due to demographic effects, future studies will estimate population growth and structure among the individuals in this study. Also, comparison of SubA with additional loci from emu is required to confirm the selection hypothesis. The emu locus is now known to map to a microchromosome, the chicken locus was known to map to a microchromosome, and the mouse and human orthologs are found on chromosomes 2 and 9, respectively (Table 1). Gonadal expression of the gene in the mouse and human exhibits an intriguing pattern. In the mouse, hypothetical protein LOC72543 is expressed in testes but not in ovaries and the reverse is reported for humans. Expression data for chickens are not yet available. It remains to be seen if LOC417092 is expressed in gonads of chickens or if SubA is expressed in the gonads of emus. Future experimentation will measure expression of SubA in embryos before and after sex differentiation begins and in various juvenile tissues in emus.

The locus SubA is of interest to developmental biology because of its potential role in gonadal development. The recognized cascade of genes involved with avian gonadal development (Smith and Sinclair 2004) is not exhaustive and SubA may be involved in either testicular or ovarian development, or both. SubA appears to be expressed in a sex-biased manner, with male upregulation in mice and female upregulation in humans (Wheeler et al. 2003). Recognition of a previously unrecognized component of avian sex differentiation suggests a series of experiments testing conservation of function and differences in thermal sensitivity between genes involved with genotypic sex determination and environmental sex determination among vertebrates. Testicular development is associated with upregulated gonadal expression of *dmrt1*, *amh*, and *sox9* and ovarian development is associated with upregulated gonadal expression of *fet1*, *foxl2*, and *aromatase* (Smith and Sinclair 2004). Incomplete description of the genes governing gonadal development is a major obstacle for efforts to characterize the evolution of sex-determining mechanisms and sexual differences. In the near future, SubA expression will be tested across genotypically and environmentally sex-determined vertebrates. In light of reports of conserved function of other genes expressed in gonads, we predict SubA will be found to play

a role in gonadal development of avian as well as nonavian reptiles.

## Acknowledgments

We thank the American Museum of Natural History, the Australian Museum, the Commonwealth Scientific and Industrial Research Organization, the National Museum of Natural History, and the South Australian Museum for donating tissue. Samples were shipped to D.E.J. under U.S. Department of Agriculture Animal and Plant Health Inspection Service permit #54119. Christopher Organ and Nicole Valenzuela contributed greatly to the success of the symposium in which this material was presented. Funding was provided by the Division of Developmental and Cell Biology and Division of Evolutionary Developmental Biology within the Society for Integrative and Comparative Biology as well as the National Science Foundation (IOS #005309844). Individual funding was also provided to D.E.J. by the National Institutes of Health (NRSA 1F32GM072494).

## References

- Auer H, Mayr B, Lambrou M, Schlegel W. 1987. An extended chicken karyotype, including the nor chromosome. *Cytogenet Cell Genet* 45:218–21.
- Canestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. *Nat Rev Genet* 8:932–42.
- Diatchenko L, et al. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 93:6025–30.
- Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* 8:689–98.
- Ewert MA, Etchberger CR, Nelson CE. 2004. Turtle sex-determining modes and TSD patterns, and some TSD pattern correlates. In: Valenzuela N, Lance V, editors. *Temperature-dependent sex determination in vertebrates*. Washington, DC: Smithsonian Books. p 21–32.
- Ezaz T, Quinn AE, Miura I, Sarre SD, Georges A, Graves JAM. 2005. The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. *Chromosome Res* 13:763–76.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202.
- Gordon L, et al. 2007. Comparative analysis of chicken chromosome 28 provides new clues to the evolutionary fragility of gene-rich vertebrate regions. *Genome Res* 17:1603–13.

- Hattori A, Okumura K, Nagase T, Kikuno R, Hirose M, Ohara O. 2000. Characterization of long cDNA clones from human adult spleen. *DNA Res* 7:357–66.
- International Chicken Genome Sequencing Consortium (ICGSC). 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
- Kent WJ. 2002. BLAT - the BLAST-like alignment tool. *Genome Res* 12:656–64.
- Loewe L, Charlesworth B. 2007. Background selection in single genes may explain patterns of codon bias. *Genetics* 175:1381–93.
- McQueen HA, Fantes J, Cross SH, Clark VH, Archibald AL, Bird AP. 1996. CpG islands of chicken are concentrated on microchromosomes. *Nat Genet* 12:321–4.
- Nishida-Umehara C, Tsuda Y, Ishijima J, Ando J, Fujiwara A, Matsuda Y, Griffin DK. 2007. The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds. *Chromosome Res* 15:721–34.
- Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33:D501–4.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–7.
- Shedlock AM, Janes DE, Edwards SV. 2008. Amniote phylogenomics: testing evolutionary hypotheses with BAC library scanning and targeted clone analysis of large-scale DNA sequences from reptiles. In: Murphy W, editor. *Phylogenomics*. Totowa, NJ: Humana Press.
- Shetty S, Kirby P, Zarkower D, Graves JAM. 2002. DMRT1 in a ratite bird: evidence for a role in sex determination and discovery of a putative regulatory element. *Cytogene Genome Res* 99:245–51.
- Smit AFA, Hubley R, Green P. 2004. RepeatMasker Open-3.0.5. [Cited 28 March 2008]. Available from: <http://www.repeatmasker.org>.
- Smith CA, Sinclair AH. 2004. Sex determination: insights from the chicken. *Bioessays* 26:120–32.
- Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–89.
- Tajima F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95.
- Wheeler DL, et al. 2003. Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31:28–33.