Microsatellite-encoded domain in rodent Sry functions as a genetic capacitor to enable the rapid evolution of biological novelty

Yen-Shan Chen^a, Joseph D. Racca^a, Paul W. Sequeira^a, Nelson B. Phillips^a, and Michael A. Weiss^{a,b,c,1}

Departments of ^aBiochemistry, ^bBiomedical Engineering, and ^cMedicine, Case Western Reserve University, Cleveland, OH 44106

Edited by Patricia K. Donahoe, Massachusetts General Hospital, Boston, MA, and approved June 7, 2013 (received for review January 16, 2013)

The male program of therian mammals is determined by Sry, a transcription factor encoded by the Y chromosome. Specific DNA binding is mediated by a high mobility group (HMG) box. Expression of Sry in the gonadal ridge activates a Sox9-dependent gene regulatory network leading to testis formation. A subset of Sry alleles in superfamily Muroidea (order Rodentia) is remarkable for insertion of an unstable DNA microsatellite, most commonly encoding (as in mice) a CAG repeat–associated glutamine-rich domain. We provide evidence, based on an embryonic pre-Sertoli cell line, that this domain functions at a threshold length as a genetic capacitor to facilitate accumulation of variation elsewhere in the protein, including the HMG box. The glutamine-rich domain compensates for otherwise deleterious substitutions in the box and absence of nonbox phosphorylation sites to ensure occupancy of DNA target sites. Such compensation enables activation of a male transcriptional program despite perturbations to the box. Whereas human SRY requires nucleocytoplasmic shuttling and coupled phosphorylation, mouse Sry contains a defective nuclear export signal analogous to a variant human SRY associated with inherited sex reversal. We propose that the rodent glutamine-rich domain has (i) fostered accumulation of cryptic intragenic variation and (ii) enabled unmasking of such variation due to DNA replicative slippage. This model highlights genomic contingency as a source of protein novelty at the edge of developmental ambiguity and may underlie emergence of non-Sry-dependent sex determination in the radiation of Muroidea.

nucleocytoplasmic trafficking | protein–DNA recognition | sexual dimorphism | transcriptional activation | triplet expansion

Protein innovation can emerge through gradual accumulation of mutations (1), rearrangement of DNA segments (2), alternative RNA splicing (3), and RNA editing (4). Exon shuffling among eukaryotic genes and pseudogenes, for example, has provided combinatorial opportunities for protein diversity within a given taxonomy of folds (5). The present study focuses on clade-specific divergence of a transcription factor (6) in association with insertion of a CAG triplet repeat (7, 8). Can microsatellite dynamics (9) in itself influence the pace and direction of protein evolution? A model is provided by Sry, an architectural transcription factor in therian mammals encoded by the sex-determining region of the Y chromosome (10). Our results rationalize rapid changes in the mechanism and fate of a developmental switch in the radiation of rodent superfamily Muroidea (SI Appendix, Fig. S1).

Sry is a sequence-specific DNA-binding protein containing a high mobility group (HMG) box, a conserved motif of DNA bending (11). In the differentiating gonadal ridge Sry activates Sox9, an autosomal gene that in turn regulates male gonadogenesis (12). Binding of murine Sry (mSry) to the testis-specific core enhancer of Sox9 (TESCO) (12) thus activates a Sertoli cell–specific gene regulatory network that mediates programs of cell–cell communication, migration, and differentiation leading to formation of the fetal testis (11). The Sry HMG box provides the signature motif of an extensive family of cognate transcription factors (designated Sox; Sry-related HMG box) with broad

functions in metazoan development and tissue-specific gene regulation (13). Sry itself arose by duplication of Sox3, an X-linked member of this family (14). Whereas Sox3 is highly conserved among mammals (SI Appendix, Table S1), Sry has undergone rapid evolution (SI Appendix, Table S2) (15), particularly within Rodentia (16). As a seeming paradox, some members of Muroidea lack Sry (such as spiny rats *Tokudaia osimensis* and *T. tokunoshimensis* and vole *Ellobius lutescens*), leading to new (and uncharacterized) mechanisms of sex determination (17, 18). We thus sought to investigate variation in the biochemical properties of Sry as a model Y-encoded protein undergoing rapid change. Our studies focused on mSry (derived from *Mus musculus domesticus*) and human SRY (hSRY); their respective domain organizations are shown in Fig. 1 in relation to the structure of the HMG box (19). Whereas hSRY (like many nonrodent Sry alleles) contains an HMG box embedded between N- and C-terminal domains (NTD/CTD), murine and rat Sry lack an NTD and contain a CTD extended by a glutamine-rich domain (Fig. 1A) containing 3–20 poly-Gln blocks separated by His-rich spacers (consensus FHDHH). Encoded by a CAG microsatellite unique to the Y chromosomes of Muroidea, the glutamine-rich domain of mSry is required for its function as a transgene in XX mice (20).

Our investigation of mSry builds on studies of inherited mutations in hSRY at a functional threshold of gonadogenesis (6, 21). Whereas glutamine-rich domains in other transcription factors flank conserved DNA-binding motifs without change in mutational clocks (22), the HMG boxes of mSry and its orthologs in Muroidea exhibit greater sequence variation (with respect to both synonymous and nonsynonymous base substitutions) than do Sry boxes in other mammalian orders (23, 24). Our results demonstrate that

Significance

Gene duplication is prominent among evolutionary pathways through which novel transcription factors and gene regulatory networks evolve. A model in mammals is provided by Sry, a Y-encoded Sox factor that initiates male development. We provide evidence that a CAG DNA microsatellite invasion into the Sry gene of a rodent superfamily enabled its rapid evolution. This unstable microsatellite encodes a variable length glutamine-rich repeat domain. Our results suggest that intragenic complementation between the glutamine-rich domain and canonical Sry motifs accelerated their divergence through repeat length-dependent biochemical linkages. Such novelty may underlie emergence of non-Sry-dependent mechanisms of male sex determination.

Author contributions: M.A.W. designed research; Y.-S.C., J.D.R., and N.B.P. performed research; Y.-S.C., J.D.R., P.W.S., N.B.P., and M.A.W. analyzed data; and Y.-S.C. and M.A.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: michael.weiss@case.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1300860110/-DCSupplemental.

such variation is associated with (i) impaired nuclear export by a mechanism analogous to a clinical mutation in hSRY, (ii) biophysical perturbations of the mSry HMG box, and (iii) impaired occupancy of TESCO in the absence of the glutamine-rich domain. Biochemical compensation is provided by the glutamine-rich domain functioning at a threshold number of poly-Gln blocks. We envisage that variation in rodent *Sry*—suppressed or unmasked at the protein level by an unstable CAG-encoded glutamine-rich domain (25)—has been a source of evolutionary innovation: an historical contingency of genomic dynamics leading to divergence of a master switch and even to its anomalous disappearance (18, 26). The *Sry* glutamine-rich domain, thus functioning as a genetic capacitor (27, 28), has fostered the rapid generation of biological novelty in the radiation of a mammalian taxon.

Results

Rat embryonic pre-Sertoli cell line CH34 (29, 30) was used as our primary platform to monitor the gene regulatory activities of N-terminal hemagglutinin-tagged (HA) *Sry* constructs following transient transfection (31). A subset of key findings was then

Fig. 1. Structures of hSRY and mSry. (A) Human protein (204 residues; upper bar) comprises N-terminal domain (violet; NTD; residues 1–55) with Ser phosphorylation sites (gray; residues 31–33); HMG box (black; residues 56–141) containing the basic tail (dark gray; bt; residues 129–141); and C-terminal domain (white; CTD; residues 142–204) containing bridge- (Br) and PDZ-binding motifs (orange and dark purple, respectively). Murine protein (395 residues; lower bar) comprises HMG box (green; residues 3–86) with basic tail (dark gray; 74–86); Br motif (orange); and C-terminal nonconserved domain (light gray) directly linked to glutamine-rich domain (chartreuse; residues 144–367). (B) Ribbon model of human HMG box/DNA complex (19). (Left) Front view of bent DNA site (blue ribbon) overlying box with basic tail (black and gray). Side chains at the protein–DNA interface are shown in red (R7, F12, I13, Y74, and P76; consensus HMG box numbering scheme), brown (R4, K6, Q62, R66, K73, K79, and K81), or auburn (R20, N32, S33, and S36). (Right) A 90° rotation about vertical axis. Coordinates were obtained from PDB entry 1J46. (C) Corresponding space-filling model of hSRY HMG box (front view). Color code of DNA contacts as in B; noncontact surfaces are gray. (D) Homology model of mSry HMG box. Amino acid substitutions are indicated by darker shades of respective colors (DNA-binding surface) or darker gray (non-DNA-binding surface).

Fig. 2. Gln-rich domain of mSry contributes to transcriptional activation of *Sox9* and TESCO occupancy. (A) Histogram showing baseline extent of *Sox9* mRNA accumulation on transfection by WT hSRY or mSry at low dose (0.02 μ g). Inactive hSRY variant I68A served as negative control (Right). (B) Schematic diagram and amino acid sequence of mSry glutamine-rich domain, comprising 20 Gln-repeat tracts (GRTs; chartreuse) separated by spacer with conserved FHDHH element (black). (C) C-terminal deletion constructs of HA-tagged mSry (brown boxes indicate the HA tag at N terminus): WT, 20 GRTs (Upper); Δ 1, 10 GRTs; Δ 2, 8 GRTs; Δ 3, 4 GRTs; Δ 4, 3 GRTs; Δ 5, 2 GRTs; Δ 6, 1 GRT (Lower). (D) Histogram showing qPCR results of *Sox9* expression by the successive C-terminal deletion constructs with low-dose transfection (0.02 μ g plasmid with 50 $\times$ empty-vector dilution as in A). In A and D, horizontal brackets designate statistical comparisons: (* or ns), Wilcoxon $P < 0.05$ or > 0.05 , respectively.

replicated in human male cell lines. Transcriptional activation of endogenous *Sox9* was probed by quantitative PCR (qPCR) and ChIP (31). Despite their structural differences (Fig. 1), mSry and hSRY exhibit similar activities in these assays in accordance with the ability of either protein to induce testicular differentiation in transgenic XX mice (10, 11). Consistent with transcriptional profiling of the differentiating XY gonadal ridge (32, 33), *Sry*-dependent activation of *Sox9* is associated with selective activation of downstream genes *Sox8* and *fibroblast growth factor 9* (*Fgf9*) (SI Appendix, Fig. S2). Transient transfection of *Sry* constructs does not alter abundances of control *Sox* mRNAs uninvolved in testis determination or mRNAs encoding housekeeping genes.

Our transfection protocol to evaluate mSry-dependent *Sox9* expression employs dilution of the expression plasmid by an empty vector to avoid transcription factor overexpression.* The transcriptional activities of mSry and hSRY are nonetheless similar at high plasmid dose without dilution (1 μ g per well) and on 50-fold plasmid dilution (Fig. 2A). Negative controls were provided in this assay by the empty vector and a variant hSRY

*Use of a strong viral promoter (derived from the CMV) to express mSry and hSRY leads to $\sim 10^6$ protein molecules per nucleus, a significantly higher concentration than is typical of lineage- and stage-specific transcription factors in metazoan development (10^2 – 10^4 molecules per nucleus). Dilution of the expression plasmid by its empty parent mitigates such overexpression, leading to nuclear accumulation in the physiological range.

(I68A; consensus position 13 of the HMG box) unable to bind specific DNA sites (34, 35). Equal expression of the mSry/hSRY constructs was verified by anti-HA Western blot; loading controls were provided by housekeeping protein α -tubulin (32).

Deletion Analysis. The mSry glutamine-rich domain in *M. musculus domesticus* contains 20 Gln-rich blocks separated by His-rich spacers (Fig. 2B). Stepwise C-terminal deletion (constructs $\Delta 1$ – $\Delta 6$ in Fig. 2C) unmasked a threshold requirement for at least three blocks to maintain mSry-dependent *Sox9* expression (Fig. 2D); 3–20 Gln-rich blocks conferred similar activities and ChIP-based estimates of TESCO occupancy (Fig. 3A and B; experimental design as defined in *SI Appendix*, Fig. S3). This dependence of TESCO occupancy on threshold glutamine-rich domain length is in accordance with its necessary inclusion in transgenes able to induce testicular differentiation in XX mice (20).

Biophysical Degeneration of Murine HMG Box. The mSry HMG box has diverged relative to nonrodent Sry domains (*SI Appendix*, Table S2) and is associated with less precise DNA bending (36). The murine domain also exhibits an anomalous sensitivity to chemical denaturation by guanidine-HCl (Fig. 4A). Similarly, its thermal stability is reduced by 3–5 °C relative to the hSRY domain (Fig. 4B, *Inset*). In both mSry and hSRY domains, partial unfolding occurs at physiological temperatures as indicated by attenuated α -helical CD features (Fig. 4B, spectra). α -Helical structure was in each case enhanced on specific DNA binding but to a more marked extent in the less stable mSry complex (Fig. 4B–D).

Respective affinities of murine and human boxes (K_d) for a consensus DNA target site (5'-TCGGTGATTGTTTCAG-3'; complement in bold), as determined by equilibrium FRET-based titration (31), are similar at 15 °C [11.2 ± 3 (murine) and 14.5 ± 2 nM (human)] and differ by ~ 1.5 -fold at 37 °C [22 ± 7 (murine) and 14.2 ± 2 nM (human)]. Such similar affinities mask compensating changes in rates of protein–DNA dissociation and (by inference) protein–DNA association (Fig. 4E; for experimental design, see *SI Appendix*, Fig. S4). At 15 °C, the lifetime of the mSry complex (6.6×10^3 ms, corresponding to $k_{\text{off}} 0.15 \pm 0.002$ s $^{-1}$) is foreshortened relative to the hSRY complex (31.3×10^3 ms;

$k_{\text{off}} 0.032 \pm 0.001$ s $^{-1}$). At 37 °C, the lifetime of the murine complex is markedly reduced (Fig. 4F); only an upper limit could be estimated ($<0.7 \times 10^3$ ms) relative to hSRY (12.5×10^3 ms; $k_{\text{off}} = 0.08 \pm 0.003$ s $^{-1}$). The small reduction in affinity of mSry (relative to hSRY) at 37 °C thus reflects insufficient compensation in rate of association. Previous studies of hSRY variants have suggested that the lifetime of the box–DNA complex correlates with transcriptional potency (37).

mSry/hSRY Chimeric Constructs. “Swap” of murine and human boxes within hSRY (chimera 1; Fig. 5A; *SI Appendix*, Table S3) resulted in reduced TESCO occupancy relative to hSRY or mSry (Fig. 3D). Impaired occupancy was not due to inefficient nuclear import as chimera 1 exhibited enhanced nuclear localization relative to hSRY with decreased frequency of cells with pan-cellular distribution (Fig. 5B and C), a pattern similar to mSry and a clinical hSRY variant bearing a defective nuclear export signal (NES) [I90M at consensus position 35 (21)] (Fig. 5B and C). Inspection of the mSry box sequence revealed a nonconservative substitution (bold) in its putative NES (human→mouse: IxxxLxxxxML→IxxxLxxxxSL). Substitution of the rodent-specific Ser by Met increased the percentage of cells exhibiting a pancellular distribution (Fig. 5B and C), rescued *Sox9* transcriptional activation (Fig. 5D), and restored CRM1 coimmunoprecipitation (IP) activity (*SI Appendix*, Fig. S5). Impaired export of mSry and I90M hSRY stands in contrast to the impaired import of a control hSRY variant bearing a nuclear localization signal (NLS) mutation (R62G; consensus position 7) as previously characterized (38). Distinct distributions of mSry and hSRY in CH34 cells are in accordance with (i) the exclusive nuclear localization of mSry in the differentiating murine XY gonadal ridge (8, 12) and (ii) the partial pancellular distribution of hSRY in an aborted human XY specimen (39, 40).

Like I90M hSRY (21), chimera 1 exhibited twofold attenuation of *Sox9* activation (SSS bars in Fig. 5D), which was rescued by acidic substitutions within the NTD protein kinase A (PKA) site [hSRY residues 26–38; PALRRSSSFLCTE (phosphorylation site is underlined)] in accordance with a nucleocytoplasmic shuttling-dependent activating phosphorylation (DDD bars in

Fig. 3. ChIP analysis of *Sox9* TESCO occupancy. (A) ChIP assays of representative mSry C-terminal deletions (Fig. 2C); experimental design is defined in *SI Appendix*, Fig. S3. (B) Histogram showing relative TESCO occupancies by C-terminal deletion series; WT mSry signal is defined as 100%. (C) ChIP assays of chimeric proteins. Chimera 1 variants contain a mouse HMG box within hSRY framework (SSS, native NTD of hSRY; DDD, S→D mutation at the phosphorylation site of NTD; NES, corrected NES in mouse box; Fig. 5A). Chimera 3 contains a human box within mSry framework (Fig. 6B). A negative control is provided by inactive hSRY variant I68A. ChIP primer sets a and c probed for SRY occupancy, whereas primer set b served as a negative control (*SI Appendix*, Table S5). At *Right* in A and C are shown nonspecific IgG (IgG) controls; equal loading was verified by primer set b. (D) Histogram showing relative TESCO occupancies by chimeric proteins; the WT hSRY signal is defined as 100%. Horizontal brackets in B and D designate statistical comparisons as in Fig. 2.

Fig. 4. Biochemical differences between HMG boxes of mSry and hSRY. (A) The murine domain (green circles) exhibits increased sensitivity to chemical denaturation by guanidine-HCl relative to the human (black line) as probed by intrinsic tryptophan fluorescence. (B) Far-UV CD spectra at 37 °C demonstrates greater attenuation of α -helical content of the murine domain (green circles) relative to human domain (black line). (Inset) Thermal unfolding midpoint of murine domain (green circles; monitored by CD at 222 nm) is decreased by ~ 5 °C relative to human (black line). (C) Thermal denaturation of mSry and hSRY box-DNA complexes (green and black circles, respectively) as monitored by CD at 222 nm. Apparent midpoint temperatures (vertical lines) are 53 °C (mSry) and 59 °C (hSRY). The HMG boxes were complexed with 12-bp DNA site 5'-GTGATTGTTTCAG-3' and complement. (D) Far-UV CD spectra at 37 °C of free mSry HMG box (open green circles), DNA complexes of mSry and hSRY (green closed circles and solid black line, respectively) showing the regain of α -helical structure (downward arrow). The spectrum of free DNA is shown as a red line. (E and F) Stopped-flow FRET-based dissociation kinetic assay of HMG-DNA complexes at 15 °C (E) and 37 °C (F); representative data and solid fitted lines showing time-dependent increase in donor fluorescence of FRET-labeled DNA due to dissociation from the SRY complex. Dissociation rate constants (k_{off}) were determined by fitting three to four individual traces to a single exponential equation (see *SI Appendix, Fig. S4* for experimental design) (31). At both temperatures, the dissociation of the murine complex (green) is more rapid than that of the human complex (black).

Fig. 5D) (41); such phosphorylation is documented in *SI Appendix, Fig. S6*. Elimination of this site in hSRY and chimera 1 led to equal residual transcriptional regulatory activities (AAA bars in Fig. 5D). Design of chimeras 2, 3, and 4 is depicted in Fig. 6A–C. The functional dependence of hSRY on NTD phosphorylation state (as probed by AAA and DDD substitutions) was eliminated by swap of CTDs, including the murine glutamine-rich domain (chimera 3; Fig. 6B and E). Native TESCO occupancy and *Sox9* activation by a transcription factor bearing the murine box (mSry) was likewise conferred by the mSry CTD (chimeras 2 and 3; Figs. 6A, D, and E and 3C and D). In the context of

chimera 3, the murine CTD with a glutamine-rich domain also compensates for an inherited mutation in the human box (Y127F in Fig. 6C; consensus position 72) associated with sex reversal and partial reduction of specific DNA binding (42) (threefold in the above FRET assay at 37 °C). *Sox9* activation was impaired by this substitution in the context of hSRY but not chimera 4 (Fig. 6F).

Transgene-Inspired Chimera Probe for Nonbox Sex-Reversal Mechanism. Chimeric transgenes expressing hSRY or goat Sry (also lacking a glutamine-rich domain) (43) under the transcriptional control of *mSry* regulatory DNA sequences are able to direct testicular differentiation in XX mice (*SI Appendix, Fig. S7*) (10, 11).

Fig. 5. Subcellular localization of mSry, hSRY, and chimeric proteins. (A) Design of chimeric mSry/hSRY chimera 1 (Bottom) in relation to WT hSRY (Top) and mSry (Middle). The color code is as in Fig. 1A with the addition of HA tags (brown). NTDs of hSRY and chimera 1 (violet) contain either native PKA site (LRRSSFLC; residues 28–36 with phosphorylation sites underlined); variants contain modified PKA sites LRRAAFLC (phospho-dead), or LRRDDFLC (phospho-mimic). (B) Subcellular localization of epitope-tagged hSRY/mSry constructs as analyzed by immunostaining: DAPI nuclear staining (Upper Row; blue), and SRY immunofluorescence (Lower Row; green). In most cells, WT hSRY localizes in nucleus with a minor fraction exhibiting pancellular distribution (C). Chimera 1 variants (DDD and with corrected NES) exhibited augmented nuclear localization [similar to hSRY variant I90M with defective NES; as predicted in (21)]. Control human mutation R62G [which impairs an NLS (38)] led to consistent pancellular distribution of hSRY. (C) Histogram indicating fractions of transfected CH34 cells exhibiting exclusive nuclear localization of hSRY/mSry (gray bars) vs. pancellular distribution (white bars). Lengths of gray and white bars do not add to 100 due to occasional GFP-positive cells lacking hSRY/mSry expression. The transfected plasmid dose was in each case 1 μ g. (D) Results of qPCR assays of *Sox9* gene expression following low-dose transfection (0.02 μ g with 50 $\times$ empty-vector dilution). Respective right and left sets of data pertain to hSRY NTD variants (SSS, AAA, and DDD as in A) or corresponding variants of chimera 1. Inactive hSRY variant I68A (Far Right) served as a negative control. Horizontal brackets in C and D indicate statistical comparisons as defined in Fig. 2.

Analogous transgene activity was observed on swap of the mSry HMG box by its X-encoded ancestor *Sox3* or homolog *Sox9* (44) (SI Appendix, Fig. S7). Chimeras 5–8 exploited these findings to demonstrate that, with the exception of swap of the murine box with hSRY (chimera 1 above), the homologous boxes function in the context of either hSRY (Fig. 7A) or human NTD-extended mSry (Fig. 7B). Whereas at high or low plasmid dose the *Sox9*-related transcriptional activities of the NTD-extended chimeras were indistinguishable from WT mSry (i.e., irrespective of box sequence; Fig. 7D), the hSRY-based chimeras exhibited inequivalent activities on plasmid dilution (Fig. 7C) in rank order mSry box < hSRY box, *Sox3* box < goat Sry box.

Control Cell Lines. To extend key findings to a human cellular milieu, additional studies were conducted in male cell lines PC-3 (45) and NT2-D1 (46) (derived from prostate and testicular cancers, respectively). Although endogenous *SOX9* in these lines is less amenable to transcriptional activation by hSRY or mSry, the two factors in each case exhibit similar relative activities (SI Appendix, Figs. S8 and S9). To enable comparative studies of variants despite reduced assay sensitivity, relative activities were further evaluated in NT2-D1 cells on cotransfection of an mSry/hSRY-responsive *luciferase* reporter (SI Appendix, Fig. S9C). The results confirm key findings of the above CH34-based studies with respect to deletion analysis of the mSry glutamine-rich domain and the transcriptional regulatory properties of chimera 1-based constructs (SI Appendix, Fig. S9D), in particular effects of NES repair in the murine box (Ser→Met at box position 45) and DDD-based phospho-mimicry of an activated hSRY N-terminal PKA site.

Discussion

Divergence of biochemical mechanisms underlying cognate gene regulatory networks (47) highlights the complementary roles of chance and necessity in the evolution of biological novelty (48). The present study investigated the relationship between a contingent genomic event—insertion of a DNA microsatellite—and its consequences for protein evolution in the adaptive radiation of a clade. A model was provided by a Y-encoded transcription factor under strong selection (Sry). Interplay between microsatellite instability and protein divergence in Muroidea may underlie emergence of three-component populations (XX females, XY females, and XY males) and non-Y-dependent mechanisms of male sex determination (17, 18, 26).

Sry Domain Organization and Drift of HMG Box. Lacking an NTD, the divergent HMG box of mSry is extended by a C-terminal glutamine-rich domain unique to Muroidea (Fig. 1; SI Appendix, Fig. S10) (20). We used chimeric and deletion constructs, corresponding in part to transgenes previously characterized in XX mice (SI Appendix, Fig. S7), to investigate the interrelation of these domains in a pre-Sertoli cell line (29). Our previous study exploited this line as a model of the differentiating gonadal ridge (30). Impaired coupling is associated with an inherited form of Swyer's syndrome [(46), XY pure gonadal dysgenesis (49)] due to variable effects on hSRY-directed *Sox9* expression (21).

Glutamine-rich domains are well known among eukaryotic transcriptional activation domains (TADs) (50). Such low-complexity sequences are found in diverse transcription factors, including Sox proteins (51), Sp1, Krüppel-related factors, and the cyclic AMP-responsive factor CREB family (52). Glutamine-rich domains can form oligomers (53) and/or contact the basal

Fig. 6. Function of mSry glutamine-rich domain in chimeric constructs. (A–C) Respective designs of chimeras 2–4 in relation to parent proteins. The domain color code and definitions of PKA site variants (SSS, AAA, and DDD; chimeras 2 and 3) are as defined in Fig. 5A. (D–F) Results of qPCR assays of *Sox9* gene expression following low-dose transfection (0.02 μ g as in Fig. 2A). A positive control was in each case provided by WT mSry; negative controls were provided by an empty vector, inactive hSRY variant I68A, or homologous mSry variant M13A. (D) *Sox9* activation by WT mSry or chimera 2 (SSS, AAA, or DDD variants). (E) *Sox9* activation by WT mSry, hSRY (SSS, AAA, or DDD variants), or chimera 3 (SSS, AAA, or DDD variants). (F) *Sox9* activation by WT mSry, chimera 3 (with WT PKA site; SSS), Y127F hSRY, or chimera 4 (WT PKA site). Horizontal brackets indicate statistical comparisons as defined in Fig. 2; $^{*}P < 0.01$.

transcriptional machinery (50). The CAG-encoded domain of mSry was first identified as a potential TAD in a yeast model (8). Its deletion within an *mSry* transgene blocks the ability of the construct to induce testicular differentiation in XX mice (20). A survey of mammalian *Sry* alleles indicates that the CAG microsatellite in Muroidea is associated with loss of (i) an NTD bearing potential phosphorylation sites (41) and (ii) a consensus NES within the HMG box as otherwise observed among mammalian *Sry* and *Sox* family members (40). The inactive NES of mSry (IxxxLxxxxSL; Fig. 8C) is selectively found in that subset of Muroidea rodents whose *Sry* alleles also contain a CAG repeat. In mSry this variant NES blocks nucleocytoplasmic shuttling as characterized in *Sox* proteins (40). Competence for CRM1-mediated nuclear export, conserved in deer and goat *Sry*, was regained on reversion to the *Sry* consensus NES (IxxxLxxxxML) (SI Appendix, Fig. S5). The contribution of the murine glutamine-rich domain to testicular differentiation in vivo (20) and to the gene regulatory activity of mSry in cell culture (present results) may resolve an apparent paradox posed by Swyer's mutations in

hSRY that (akin to WT mSry) are proposed to impair its nuclear export (21).

We speculate that the biochemical activity of the mSry glutamine-rich domain has attenuated selective pressure on its *Sry* HMG box, leading to genetic drift. Whereas the HMG box of *Sox3* (the X-encoded ancestor of *Sry*) (14) is broadly conserved among vertebrates, including within Rodentia, the mSry domain differs from the boxes of primates, ungulates, and other mammalian orders at more sites (and at these sites by less conservative substitutions) than do the latter from the *Sox3* box (SI Appendix, Tables S1 and S2). Such variation in mSry was associated with attenuated thermodynamic stability and foreshortened residence time of a specific DNA complex (Fig. 4).[†]

[†]Although the mSry domain has been described as exhibiting more stringent sequence specificity than hSRY (a seeming biochemical improvement) (71), such findings may represent kinetic artifacts of gel mobility-shift assays (i.e., even more rapid dissociation of variant mSry complexes) (72). Changes in protein–DNA dissociation rates may also account for the murine domain's seeming enhancement of discrimination against AT→IC transitions (71).

Fig. 7. Transgenic-inspired design of SRY chimeras. (A) Domain organization of chimeric proteins 1, 5, and 6 in relation to WT hSRY and (B) chimeric proteins 3, 7, and 8 in relation to native mSry bearing human NTD. Transgenes encoding chimeric proteins 5 and 7 (containing the HMG boxes of Sox3; blue) are able to induce XX sex reversal in mice (44). Similarly, the goat Sry HMG box was used in chimeric proteins 6 and 8 (aquamarine) as motivated by the comparable activity of a goat Sry transgene in XX mice (43). (C) Results of qPCR assays of *Sox9* gene expression activated by WT hSRY or hSRY-based chimeric proteins (1, 5, 6) following low-dose transfection (0.02 μ g as in Fig. 2A; white bars). (D) Corresponding qPCR assays of native mSry and mSry-based chimeric proteins (3, 7, 8). In each case, the function of the chimeric proteins was indistinguishable from that of WT mSry. A negative control was provided by the inactive hSRY variant I68A (Right). Horizontal brackets indicate statistical comparisons; n.s., $P > 0.05$.

Although the native-like α -helical structure is largely regained on specific DNA binding (induced fit), the reduced lifetime of the mSry domain–DNA complex may underlie its impaired transcriptional regulatory activity in the absence of the glutamine-rich domain (20). These biophysical findings suggest that the contribution of the glutamine-rich domain to *Sox9* transcriptional activation (36) compensates for biophysical instability, impaired nucleocytoplasmic shuttling, and absence of NTD phosphorylation site.

Block Glutamine-Rich Domain Dissection. A CAG microsatellite occurs in *Sry* in several lineages within Muroidea, most dramatically in Muridae (old world rats, mice, and gerbils). Repeat lengths are variable, ranging from 20 poly-Gln blocks (as in mSry in *M. musculus domesticus*; Fig. 2B) (54) to 3 (*Rattus norvegicus*) (55). Even among laboratory strains of *M. musculus domesticus*-derived Y chromosomes encode Sry proteins of different lengths relative to *molossinus*-derived Y chromosomes (alleles *Sry*^{B6} and *Sry*¹²⁹) (7). Although block numbers vary, the downstream *Sox9*-dependent gene regulatory network is presumably similar as indicated by heterogametic male development. Tolerance to variation in poly-Gln block number is in accordance with our deletion analysis wherein constructs containing 3 or more blocks activated *Sox9* transcription to an extent similar to that of canonical mSry (20 blocks; Fig. 2D; SI Appendix, Figs. S8 and S9). Further, similar activities were observed in CH34 assays of WT *Sry* alleles derived from *Rattus norvegicus* and *Tokudaia muenninki* (Muennink's spiny rat; Okinawa), which each contain three poly-Gln blocks (accession number: *R. norvegicus*, NP_036904; *T. muenninki*, BAJ08420). ChIP studies focused on Sry-binding sites in TESCO (12) indicated CTDs containing less than three blocks are associated with loss of *Sox9* enhancer occupancy (Fig. 3A and B).

Microsatellite-Based Biochemical Complementation. Chimeric mSry/hSRY constructs were prepared to test whether a CAG-associated TAD could relax biochemical constraints on the function of the HMG box. Chimera 1 is a variant of hSRY containing the murine box (Fig. 5A). Its properties are analogous to those of I90M hSRY [an inherited allele (21)] bearing a dysfunctional NES, leading in each case to reduced activity despite increased nuclear accumulation (Fig. 5B and C). Comparison of human NTD variants indicated that phospho-mimicry through acidic substitutions in a putative PKA site rescued the activity of chimera 1. Such rescue also implies that, on NTD phosphorylation and on enhanced nuclear accumulation due to impaired nuclear export, the function of hSRY (at least in a rodent cell line) tolerates the many substitutions that otherwise distinguish between human and murine boxes. To test whether NTD phosphorylation could modulate the function of mSry, chimera 2 was fused to the human NTD (Fig. 6A). Its gene regulatory properties were found to be robust to AAA or SSS substitutions (Fig. 6D), implying that the glutamine-rich domain renders such regulation superfluous. Chimera 3 contains both the human NTD and box fused to the C-terminal nonbox sequences of mSry, including its glutamine-rich domain (Fig. 6B). Occupancy of TESCO sites was similar to its WT parents (Fig. 3D). The function of chimera 3 was likewise robust to PKA-site substitutions (Fig. 6E).

Biochemical complementation by the mSry CTD was further investigated in relation to an inherited human variant near the protein–DNA interface (Y127F in Fig. 6C; consensus position 72 in the HMG box), which partially impairs specific DNA binding (42). The aromatic ring adjoins V60 (consensus position 5), also a site of inherited mutation (31). Whereas in the context of hSRY Y127F impairs *Sox9* expression by approximately twofold

Fig. 8. Rodent Sry alleles with a CAG-encoded glutamine-rich domain contain attenuated NES motif. (A) Representative species in Muroidea superfamily. Color codes depicting variations in the Sry frame: brown, follows mSry frame, such as HMG-bridge-“domain with repeating Gln-tracts encoded by CAG”; magenta, species with Sry containing poly-A repeating tracts encoded by GCA and species with different evolutionary fates of Sry are framed in boxes. (B) Phylogenetic relationships of three *Tokudaia* species. Tree was adapted from Murata et al. (77). Color codes: brown, as in A; red, species that have lost Sry. (C) Alignments of SRY sequences without CAG-encoded repeating domain (upper bracket) or with CAG-encoded glutamine-rich domain (bottom bracket). NES motifs are highlighted in bold. Cylinders (Upper) show secondary-structural environment of NES motif. Residue numbers correspond to consensus HMG box. The second and third α -helices in hSRY HMG box are labeled $\alpha 2$ and $\alpha 3$; conserved serines proposed to attenuate NES efficiency are in red (bottom bracket).

[as observed in studies of V60L and V60A (31)], the mutation has no effect in the context of chimera 3 (Fig. 6F). Such intragenic complementation supports an evolutionary scenario wherein insertion of a CAG microsatellite in a founding lineage of Muroidea enabled drift of HMG-box sequences. To further explore glutamine-rich domain complementation, chimeric constructs 5–8 used the HMG boxes of mSox3 and goat Sry as inspired by studies of chimeric transgenes (43, 44). Whereas in the context of hSRY, respective “box swap” variants exhibited relative activities in the order mouse < human = Sox3 < goat (Fig. 7C), such functional differences were abolished in the presence of the mSry CTD (Fig. 7D).

Evolution of Male Sex Determination. Whereas Sry is generally conserved among therian mammals as the testis-determining locus (56), an enigma is posed in Muroidea (SI Appendix, Fig. S1). One member of family Cricetidae, the vole *Ellobius lutescens*, has no Sry gene or Y chromosome (Fig. 8A); its mechanism of sex determination is unknown (18). Evidence for the rapid evolution of non-Sry-dependent male-determining mechanisms has likewise been obtained within Muridae. Like genus *Apodemus* (which contains the common mouse and rat), related genus *Tokudaia* contains species with glutamine-rich domain-associated Sry alleles as exemplified by *T. muenninki*. Despite the implication of a common ancestor whose Y chromosome contained the original CAG-associated microsatellite, *Tokudaia* also contains species lacking a Y chromosome (Fig. 8B) (57) as exemplified by *T. tokunoshimensis* (Tokunoshima spiny rat) and *T. osimensis* (Armani spiny rat). We propose a scenario wherein (i) microsatellite invasion of Sry within a Muroidea common ancestor enabled drift of HMG-box sequences with biophysical perturbation and loss of nonbox phosphorylation sites and (ii) subsequent glutamine-rich domain repeat-number instability led (below the

threshold of three poly-Gln blocks) to attenuated Sry-directed Sox9 activation in some lineages, leading in turn to reproductive isolation and recruitment of non-Sry-dependent mechanisms of Sox9 transcriptional activation in the bipotential gonadal ridge. Redundant or nonfunctional Sry alleles were a likely precondition for the rare anomalous loss of the Y chromosomes in this clade.[‡]

The plausibility of this evolutionary scenario is strengthened by intermediate cases found within Cricetidae (grass mice *Akodon boliviensis* and *A. azarae*). Although males represent the heterogametic sex, these species exhibit high percentages of XY females (58). Their variant Sry genes encode a foreshortened NTD (lacking potential phosphorylation sites) and a divergent box with nonconsensus NES (SI Appendix, Table S4), followed by a single-block Gln-rich motif (58). We speculate that this remnant glutamine-rich domain is insufficient to rescue the function of the divergent NTD and box, providing only partial activation of Sox9 at the threshold of testis determination: gonadogenesis would thereby be nonrobust with respect to autosomal variation, environmental fluctuations, or stochastic gene expression. We envisage that such grass mice stand at the crossroads of Sry loss and Y-chromosome degeneration.

Concluding Remarks. Poly-Gln repeats encoded by CAG repeats were first observed in neurological disorders (59–61) in which length-dependent alterations of protein structure, function, and toxicity can correlate with clinical severity or age of onset (62). In Huntington’s disease, for example, aberrant gain of function by

[‡]Unlike in *Rodentia*, primate Y chromosomes have been stable (73). It is not known whether in the absence of microsatellite instability selective pressure to maintain SRY has dampened the pace of Y-chromosome degeneration.

the variant huntingtin perturbs neuronal gene expression, in part through competitive binding of the glutamine-rich domain to transcriptional coactivators and basal transcription factors (22). Whereas microsatellite instability within transcription factors is not generally associated with divergence of respective DNA-binding motifs,[§] the evolution of *Sry* in Muroidea is remarkable for both variation in glutamine-rich domain length and divergence of HMG-box sequences (23, 63). We speculate that glutamine-rich domain-associated gain of function in a TESCO-directed *Sox9* transcriptional regulatory complex circumvents biochemical requirements for nucleocytoplasmic shuttling and nucleocytoplasmic shuttling-coupled phosphorylation as defined in Sox proteins (40).

We propose that the CAG triplet repeat of rodent *Sry* alleles has functioned in the radiation of Muroidea as an intragenic “capacitor” to suppress phenotypic consequences of variation elsewhere in the protein, which (in the case of mSry) includes destabilizing substitutions in the HMG box, loss of nucleocytoplasmic shuttling, and deletion of the NTD. Such variation could then have been unmasked by microsatellite instability leading to truncation of the glutamine-rich domain below its critical threshold. This model extends the paradigm of a genetic capacitor as defined by heat shock protein 90 (Hsp90) (27, 28). Because Hsp90 buffers the misfolding of proteins regulating metazoan development (thereby conferring interim stability to gene regulatory networks), discharge of the *Hsp90* capacitor may underlie rapid morphological evolution as documented in the fossil record (64). Similarly, the microsatellite capacitor of *Sry* in Muroidea may have enabled, via replicative DNA slippage (25), sudden shifts in molecular mechanisms of male sex determination. Operating through the biochemical properties of a glutamine-rich domain in a TESCO complex, this *Sry* capacitor may discharge to create reproductive barriers between nascent species.

We thus envisage that microsatellite instability within *Sry* has promoted the emergence of biological novelty in a mammalian taxon. Such innovation reflects a combination of genomic and biochemical mechanisms distinct from general processes leading to Y degeneration (65, 66). Although biochemical properties of mSry and hSRY differ, each has evolved to regulate *Sox9* expression just above the threshold of Sertoli cell specification. Gonadogenesis at the edge of ambiguity is shared by rare human families (21, 31, 49) and is suggested by the frequency of XY sex reversal among grass mice (58). The thin thread of testis determination (67), first glimpsed in studies of murine Y chromosome-autosome incompatibility (68), represents an apparent violation of the Waddington principle of developmental canalization (69). Addressing why sex is different will

require deciphering a seeming paradox: multilevel selection (70) against the robustness of male gonadogenesis. A fundamental problem at the intersection of biochemistry and evolutionary biology is posed by the developmental, neuroendocrine, behavioral, and social origins of such selection.

Materials and Methods

Plasmids. Plasmids expressing hSRY, mSry, and variants (SI Appendix, Table S1) were constructed by PCR and cloned into cytomegalovirus vector (pCMV) (containing the CMV promoter) (31). The cloning site encoded an N-terminal HA tag in triplicate.

Rodent Cell Culture. CH34 cells (kindly provided by T. R. Clarke and P. K. Donahoe, Massachusetts General Hospital, Boston) (30) were cultured in DMEM containing 5% (vol/vol) heat-inactivated FBS at 37 °C under 5% CO₂.

Human Cell Lines. NT2-D1 cells (46) were grown in DMEM in an atmosphere of 5% CO₂; the complete growth medium contained FBS to a final concentration of 10%. PC-3 cells (45) were cultured in the F-12K medium (ATCC) with 10% FBS in 5% CO₂ atmosphere. Transient transfections were carried out by the Eugene HD protocol (Hoffmann LaRoche). PCR primers were in accordance with human genomic sequences.

Transient Transfection. Transfections were performed as described (76). Efficiencies were determined by the ratio of GFP-positive cells to untransfected cells following cotransfection with pCMX-SRY and pCMX-GFP. Cellular localization was probed by immunostaining 24 h after transfection.

Western Blot. Expression of mSry/hSRY and variants was monitored by Western blot using monoclonal anti-HA antiserum (Sigma-Aldrich).

Real-Time qRT-PCR Assay. Accumulation of *Sox9* mRNA in transfected CH34 cells was probed by qPCR as described (31). Cellular total RNA was extracted using the RNeasy kit (Qiagen). Primer sequences are provided in SI Appendix, Table S5. TFIID was used as an internal control; measurements were made in triplicate with blind coded samples.

Immunocytochemistry. Transfected cells were evenly plated on 12-mm coverslips, fixed with 3% para-formaldehyde in PBS, and visualized by fluorescent microscopy in relation to the total number of GFP-positive cells.

ChIP. Transfected cells were probed by ChIP using an anti-HA antiserum. An expanded high-fidelity PCR protocol was provided by the vender (Hoffmann LaRoche).

Biophysical Assays. Circular dichroism, fluorescent spectroscopy, FRET-based *K_d* determinations, and stopped-flow FRET-based analysis of protein–DNA dissociation rates were performed as described (31).

ACKNOWLEDGMENTS. We thank Prof. P. K. Donahoe for cell line CH34 and P. DeHaseth, H.-Y. Kao, and D. Samols for advice. M.A.W. thanks B. Baker, P. K. Donahoe, F. A. Jenkins, Jr., P. Koopman, R. Lovell-Badge, R. Sekido, and D. Wilhelm for discussion. This work was supported in part by National Institutes of Health Grant GM080505 (to M.A.W.). This article is dedicated to the memory of the late Prof. Farish A. Jenkins, Jr. (Harvard University and the Harvard–MIT Program in Health Sciences and Technology) for his encouragement, humanity, and scientific example.

[§]Microsatellite instability within the coding region of a basal transcription factor may cause disease as demonstrated by spinocerebellar ataxia (type 17) (74). Instability within a specific transcription factor may also underlie the rapid divergence of a morphological program in *Carnivora* (75).

- Soskine M, Tawfik DS (2010) Mutational effects and the evolution of new protein functions. *Nat Rev Genet* 11(8):572–582.
- Bornberg-Bauer E, Beausart F, Kummerfeld SK, Teichmann SA, Weiner J, 3rd (2005) The evolution of domain arrangements in proteins and interaction networks. *Cell Mol Life Sci* 62(4):435–445.
- Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: Diversification, exon definition and function. *Nat Rev Genet* 11(5):345–355.
- Reenan RA (2005) Molecular determinants and guided evolution of species-specific RNA editing. *Nature* 434(7031):409–413.
- Bogard LD, Deem MW (1999) A hierarchical approach to protein molecular evolution. *Proc Natl Acad Sci USA* 96(6):2591–2595.
- Koopman P (1995) The molecular biology of SRY and its role in sex determination in mammals. *Reprod Fertil Dev* 7(4):713–722.
- Coward P, et al. (1994) Polymorphism of a CAG trinucleotide repeat within *Sry* correlates with B6.YDom sex reversal. *Nat Genet* 6(3):245–250.
- Dubin RA, Ostrer H (1994) *Sry* is a transcriptional activator. *Mol Endocrinol* 8(9):1182–1192.
- Schlötter C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109(6):365–371.
- Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R (1991) Male development of chromosomally female mice transgenic for *Sry*. *Nature* 351(6322):117–121.
- Lovell-Badge R, Canning C, Sekido R (2002) *Sex-Determining Genes in Mice: Building Pathways* (John Wiley & Sons Ltd., West Sussex, UK), pp 4–22.
- Sekido R, Lovell-Badge R (2008) Sex determination involves synergistic action of SRY and SF1 on a specific *Sox9* enhancer. *Nature* 453(7197):930–934.
- Guth SI, Wegner M (2008) Having it both ways: Sox protein function between conservation and innovation. *Cell Mol Life Sci* 65(19):3000–3018.
- Katoh K, Miyata T (1999) A heuristic approach of maximum likelihood method for inferring phylogenetic tree and an application to the mammalian SOX-3 origin of the testis-determining gene SRY. *FEBS Lett* 463(1–2):129–132.
- Whitfield LS, Lovell-Badge R, Goodfellow PN (1993) Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature* 364(6439):713–715.
- Pamilo P, O'Neill RJ (1997) Evolution of the *Sry* genes. *Mol Biol Evol* 14(1):49–55.

17. Soullier S, Hanni C, Catzeffis F, Berta P, Laudet V (1998) Male sex determination in the spiny rat *Tokudaia osimensis* (Rodentia: Muridae) is not *Sry* dependent. *Mamm Genome* 9(7):590–592.
18. Just W, et al. (2007) *Ellobius lutescens*: Sex determination and sex chromosome. *Sex Dev* 1(4):211–221.
19. Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM (2001) Structural basis for SRY-dependent 46-X,Y sex reversal: Modulation of DNA bending by a naturally occurring point mutation. *J Mol Biol* 312(3):481–499.
20. Bowles J, Cooper L, Berkman J, Koopman P (1999) *Sry* requires a CAG repeat domain for male sex determination in *Mus musculus*. *Nat Genet* 22(4):405–408.
21. Knowler KC, et al. (2011) Failure of SOX9 regulation in 46XY disorders of sex development with SRY, SOX9 and SF1 mutations. *PLoS ONE* 6(3):e17751.
22. Dunah AW, et al. (2002) Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* 296(5576):2238–2243.
23. Tucker PK, Lundrigan BL (1993) Rapid evolution of the sex determining locus in Old World mice and rats. *Nature* 364(6439):715–717.
24. Pontiggia A, Whitfield S, Goodfellow PN, Lovell-Badge R, Bianchi ME (1995) Evolutionary conservation in the DNA-binding and -bending properties of HMG-boxes from SRY proteins of primates. *Gene* 154(2):277–280.
25. Liu Y, Wilson SH (2012) DNA base excision repair: A mechanism of trinucleotide repeat expansion. *Trends Biochem Sci* 37(4):162–172.
26. Graves JA (2002) Evolution of the testis-determining gene: The rise and fall of SRY. *Novartis Found Symp* 244:86–97.
27. Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396(6709):336–342.
28. Lindquist S (2009) Protein folding sculpting evolutionary change. *Cold Spring Harb Symp Quant Biol* 74:103–108.
29. Haqq CM, et al. (1994) Molecular basis of mammalian sexual determination: Activation of Müllerian inhibiting substance gene expression by SRY. *Science* 266(5190):1494–1500.
30. Haqq CM, Donahoe PK (1998) Regulation of sexual dimorphism in mammals. *Physiol Rev* 78(1):1–33.
31. Phillips NB, et al. (2011) Mammalian testis-determining factor SRY and the enigma of inherited human sex reversal. *J Biol Chem* 286:36787–36807.
32. Moniot B, et al. (2009) The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation. *Development* 136(11):1813–1821.
33. Barrionuevo F, Scherer G (2010) SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int J Biochem Cell Biol* 42(3):433–436.
34. King CY, Weiss MA (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc Natl Acad Sci USA* 90(24):11990–11994.
35. Weiss MA, Ukiyama E, King CY (1997) The SRY cantilever motif discriminates between sequence- and structure-specific DNA recognition: Alanine mutagenesis of an HMG box. *J Biomol Struct Dyn* 15(2):177–184.
36. Phillips NB, et al. (2004) Sry-directed sex reversal in transgenic mice is robust with respect to enhanced DNA bending: Comparison of human and murine HMG boxes. *Biochemistry* 43(22):7066–7081.
37. Ukiyama E, et al. (2001) SRY and architectural gene regulation: The kinetic stability of a bent protein-DNA complex can regulate its transcriptional potency. *Mol Endocrinol* 15(3):363–377.
38. Gontan C, et al. (2009) Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors. *J Cell Biol* 185(1):27–34.
39. Poulat F, et al. (1995) Nuclear localization of the testis determining gene product SRY. *J Cell Biol* 128(5):737–748.
40. Malki S, Boizet-Bonhoure B, Poulat F (2010) Shuttling of SOX proteins. *Int J Biochem Cell Biol* 42(3):411–416.
41. Desclozeaux M, et al. (1998) Phosphorylation of an N-terminal motif enhances DNA-binding activity of the human SRY protein. *J Biol Chem* 273(14):7988–7995.
42. Jordan BK, Jain M, Natarajan S, Frasier SD, Vilain E (2002) Familial mutation in the testis-determining gene SRY shared by an XY female and her normal father. *J Clin Endocrinol Metab* 87(7):3428–3432.
43. Pannetier M, et al. (2006) Goat SRY induces testis development in XX transgenic mice. *FEBS Lett* 580(15):3715–3720.
44. Bergstrom DE, Young M, Albrecht KH, Eicher EM (2000) Related function of mouse SOX3, SOX9, and SRY HMG domains assayed by male sex determination. *Genesis* 28(3–4):111–124.
45. Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW (1979) Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest Urol* 17(1):16–23.
46. Knowler KC, et al. (2007) Characterisation of urogenital ridge gene expression in the human embryonal carcinoma cell line NT2/D1. *Sex Dev* 1(2):114–126.
47. Booth LN, Tuch BB, Johnson AD (2010) Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* 468(7326):959–963.
48. Monod J (1972) *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*, trans Wainhouse A (Vintage Books, London), 1st Ed, p 199.
49. Quinn A, Koopman P (2012) The molecular genetics of sex determination and sex reversal in mammals. *Semin Reprod Med* 30(5):351–363.
50. Atanesyan L, Günther V, Dichtl B, Georgiev O, Schaffner W (2012) Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biol Chem* 393(1–2):63–70.
51. Kasimiotis H, et al. (2000) Sex-determining region Y-related protein SOX13 is a diabetes autoantigen expressed in pancreatic islets. *Diabetes* 49(4):555–561.
52. Persengiev SP, Saffer JD, Kilpatrick DL (1995) An alternatively spliced form of the transcription factor Sp1 containing only a single glutamine-rich transactivation domain. *Proc Natl Acad Sci USA* 92(20):9107–9111.
53. Tobaben S, Varoqueaux F, Brose N, Stahl B, Meyer G (2003) A brain-specific isoform of small glutamine-rich tetrapeptide repeat-containing protein binds to Hsc70 and the cysteine string protein. *J Biol Chem* 278(40):38376–38383.
54. Denny P, Swift S, Connor F, Ashworth A (1992) An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J* 11(10):3705–3712.
55. Griffiths R, Tiwari B (1993) Primers for the differential amplification of the sex-determining region Y gene in a range of mammal species. *Mol Ecol* 2(6):405–406.
56. Berta P, et al. (1990) Genetic evidence equating SRY and the testis-determining factor. *Nature* 348(6300):448–450.
57. Arakawa Y, Nishida-Umehara C, Matsuda Y, Sutou S, Suzuki H (2002) X-chromosomal localization of mammalian Y-linked genes in two XO species of the Ryukyu spiny rat. *Cytogenet Genome Res* 99(1–4):303–309.
58. Bianchi NO (2002) Akodon sex reversed females: The never ending story. *Cytogenet Genome Res* 96(1–4):60–65.
59. Pearson CE, Nichol Edamura K, Cleary JD (2005) Repeat instability: Mechanisms of dynamic mutations. *Nat Rev Genet* 6(10):729–742.
60. Walker FO (2007) Huntington's disease. *Lancet* 369(9557):218–228.
61. van Eyk CL, et al. (2011) Perturbation of the Akt/Gsk3- β signalling pathway is common to *Drosophila* expressing expanded untranslated CAG, CUG and AUUCU repeat RNAs. *Hum Mol Genet* 20(14):2783–2794.
62. Taylor JP, Hardy J, Fischbeck KH (2002) Toxic proteins in neurodegenerative disease. *Science* 296(5575):1991–1995.
63. Miller KE, Lundrigan BL, Tucker PK (1995) Length variation of CAG repeats in Sry across populations of *Mus domesticus*. *Mamm Genome* 6(3):206–208.
64. Eldredge N, Gould SJ (1972) Punctuated equilibria: An alternative to phyletic gradualism. *Models in Paleobiology*, ed Schopf TMJ (Freeman Cooper, San Francisco), pp 82–115.
65. Marchal JA, Acosta MJ, Bullesos M, Díaz de la Guardia R, Sánchez A (2003) Sex chromosomes, sex determination, and sex-linked sequences in Microtidae. *Cytogenet Genome Res* 101(3–4):266–273.
66. Wilson MA, Makova KD (2009) Genomic analyses of sex chromosome evolution. *Annu Rev Genomics Hum Genet* 10:333–354.
67. Polanco JC, Koopman P (2007) Sry and the hesitant beginnings of male development. *Dev Biol* 302(1):13–24.
68. Albrecht KH, Young M, Washburn LL, Eicher EM (2003) Sry expression level and protein isoform differences play a role in abnormal testis development in C57BL/6J mice carrying certain Sry alleles. *Genetics* 164(1):277–288.
69. Waddington CH (1959) Canalization of development and genetic assimilation of acquired characters. *Nature* 183(4676):1654–1655.
70. Wilson DS, Wilson EO (2007) Rethinking the theoretical foundation of sociobiology. *Q Rev Biol* 82(4):327–348.
71. Giese K, Pagel J, Grosschedl R (1994) Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. *Proc Natl Acad Sci USA* 91(8):3368–3372.
72. Li B, et al. (2006) SRY-directed DNA bending and human sex reversal: Reassessment of a clinical mutation uncovers a global coupling between the HMG box and its tail. *J Mol Biol* 360(2):310–328.
73. Hughes JF, et al. (2012) Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* 483(7387):82–86.
74. Tomiuk J, et al. (2007) Repeat expansion in spinocerebellar ataxia type 17 alleles of the TATA-box binding protein gene: An evolutionary approach. *Eur J Hum Genet* 15(1):81–87.
75. Sears KE, Goswami A, Flynn JJ, Niswander LA (2007) The correlated evolution of Runx2 tandem repeats, transcriptional activity, and facial length in *carnivora*. *Evol Dev* 9(6):555–565.
76. Chakraborty S, et al. (2006) a-Actinin 4 potentiates myocyte enhancer factor-2 transcription activity by antagonizing histone deacetylase 7. *J Biol Chem* 281(46):35070–35080.
77. Murata C, Yamada F, Kawauchi N, Matsuda Y, Kuroiwa A (2010) Multiple copies of SRY on the large Y chromosome of the Okinawa spiny rat, *Tokudaia muenninki*. *Chromosome Res* 18(6):623–634.