CMLS, Cell. Mol. Life Sci. 55 (1999) 824–830
1420-682X/99/070824-07 $ 1.50 + 0.20/0
© Birkhäuser Verlag, Basel, 1999

CMLS Cellular and Molecular Life Sciences

The one-to-four rule and paralogues of sex-determining genes

S. Ohno

Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte
(California 91010-3000, USA), Fax + 1 818 357 1929

Abstract. Because of two successive rounds of tetraploidization at their inception, the vertebrates contain four times more protein-coding genes in their genome than the invertebrates: 60,000 versus 15,000. Consequently, each invertebrate gene has been amplified to the maximum of four paralogous genes in vertebrates: the one-to-four rule. When this rule is applied to genes pertinent to gonadal development and differentiation, the following emerged: (i) Two closely related zinc-finger transcription factor genes in invertebrates have been amplified to two paralogous groups in vertebrates. One consisted of *EGR1*, *EGR2*, *EGR3* and *EGR4*, whereas the only known paralogue of the other is *WT1*, which controls the developmental fate of the entire nephric system, and therefore of gonads. Interestingly, *EGR1* and *WT1* act as antagonists of each other in nephroblastic cells. (ii) *SF-1*, which controls the fate of two steroid hormone-producing organs, adrenals and gonads, is descended from the invertebrate *Ftz-F1* gene, and its only known paralogue is *GCNF-1*. (iii) The Y-linked *SRY*, the mammalian testis-determining gene, is a paralogue neither of *SOX3* (*SRX*) nor of *SOX9*. Its ancient origin suggests that *SRY* once became extinct in earlier vertebrates, only to revive itself in the mammalian ancestor. (iv) Inasmuch as four paralogues of one invertebrate nuclear receptor gene have differentiated to receptors of androgen, mineralocoticoid, glucocorticoid and progesterone, there should at most be four paralogous estrogen-receptor genes in the vertebrate genome. It is likely that one of them plays a pivotal role in the estrogen-dependent sex-determining mechanism so commonly found among reptiles, amphibians and fish.

Key words. One-to-four rule; vertebrates versus invertebrates; paralogous genes; revived genes.

Introduction

The somewhat satirical Roman proverb 'mutatis mutandis, ipsissima omnia' ('all the necessary changes having been made, all the things remain as before') succinctly summarizes my long-held belief on the nature of evolution. Indeed, even after profound evolutionary changes in the body form and organ types, the same gene more often than not performs the same function as before; thus, all the changes having been made, everything, in fact, remained the same. The *PAX6* gene that governs the development of all metazoan eyes can be given as the best example of the above.

Metazoan eye formation was traditionally invoked as the classical example of convergent evolution, meaning achievements of the same end by divergent genetic means. At first glance, it indeed appears so, since animals equipped with eyes have a peculiar way of showing up in unexpected branches of various phyla without warning, as it were. For example, a vast majority of the flatworms belonging to the phylum Plathelminthes are devoid of eyes. Yet planarias of the class Turbellaria possess eyes. Among the phylum Mollusca, members of the class Cephalopoda such as squids are uniformly equipped with eyes, whereas clams and mussels of the class Bivalvia are devoid of eyes with equal uniformity. The class Gastropoda, on the other hand, is a mixed bag containing some with and others without eyes. Furthermore, compound eyes of crustaceans and insects belonging to the phylum Arthropoda are very different from single eyes of other animals. In spite of all the above, it has finally been proven that metazoan eyes are invariably formed under the direction of the one specific gene: *PAX6* [1, 2]. In addition, the primacy of the *PAX6* gene in eye formation has been established by coupling the *PAX6* coding sequence to various regulatory elements in *Drosophila*. Eyes formed at various ectopic sites where *PAX6* was expressed [2].

Because of my belief on the nature of evolution, when I was informed by my distinguished immunologist colleague Edward A. Boyse in 1975 that S. S. Wachtel and G. C. Koo in his laboratory found H-Y plasma membrane antigen to be heterogametic sex-specific not only among vertebrates but also among invertebrates, my immediate reaction was that this H-Y antigen must be the long-sought-after universal primary determiner of the heterogametic sex [3].

As it turned out, it is the Y-linked *SRY* transcription regulator gene that directs mammalian testicular differentiation [4, 5]. Subsequently, however, it was found that, albeit a member of the *SOX* family of genes, *SRY* as such is not found in other classes of vertebrates. Furthermore, the *SRY*-dependent sex determination sensu stricto is not universal even among mammals, for males of a certain esoteric microtine rodent species, *Ellobius lutescens*, are devoid of *SRY* [6]; in fact, the entire Y chromosome is absent in this species, with males and females sharing the identical X0 sex chromosome constitution [6]. It follows that *SRY* is a violator of the magnum dictum of evolution, and as such *SRY* is irrelevant to the sex-determining mechanisms of birds, reptiles, amphibians and fish. In this paper, I would like to consider various genes involved in vertebrate sex determination in the broader context.

The one-to-four gene number rule among invertebrates and vertebrates

While a few of the ongoing genome projects on diverse species have finally been completed, others, too, have already yielded numerous relevant information. Table 1 shows that three invertebrate species representing the three different phyla Nemathelminthes, Arthropoda and Chordata are endowed with about the same total number of protein-coding genes in the genome, the number being 15,000 or thereabout [7]. Particularly noteworthy is the fact that a tunicate *Ciona* is an invertebrate member of the phylum Chordata to which we vertebrates also belong. In table 1, gnathostomic vertebrates are represented by a puffer fish and by our own species. As with all the extremely specialized teleost fish, the genome of a puffer fish, by a drastic secondary reduction, became a mere one-eighth the size of the mammalian genome. Yet, the total number of protein-coding genes contained in the genome of a puffer is about the same as that in the human, therefore, mammalian genome [8]. In table 1, a concensus number of

Table 1. Difference in gene numbers between invertebrates and gnathostomic vertebrates.

	Total number of gene loci	Genome size in numbers of base pairs
Invertebrates		
Phylum		
Nemathelminthes		
Nematode		
Caenorhabditis elegans	17,500 ± 1500	1.00×10^8
Phylum		
Arthropoda		
Fruit fly		
Drosophila melanogaster	13,500 ± 2500	1.65×10^8
Phylum		
Chordata		
Subphylum		
Urochordata		
Tunicate		
Ciona intestinalis	15,500 ± 2500	1.90×10^8
Gnathostomic vertebrates		
Phylum		
Chordata		
Subphylum		
Vertebrata		
Class		
Osteichthyes		
Japanese puffer		
Fugu rubripes	86,250 ± 21,500	3.90×10^8
Class		
Mammalia		
Homo sapiens	86,250 ± 21,500	3.00×10^9

Table 2. One-to-four rule as it applies to regulatory genes (modified from [10]).

	Invertebrates	Vertebrates	
	Drosophila	human	
		gene	chromosomal location
Notch		*NOTCH1*	9q34.3
epidermal-neuronal	*N*	*NOTCH2*	1q11–p13
cell-cell interaction		*NOTCH3*	19p13
receptors		*INT3*	6p21.3
Mef2		*MEF2A*	15q25
MADS box	*Mef2*	*MEF2B*	19p12
enhancing factors		*MEF2C*	5q14
		MEF2D	1q12–q23
Ras		*RRAS*	19q13
GTP binding	*RAS85D*	*HRAS*	11p15.5
oncogenes		*KRAS2*	12p12.1
		NRAS	1p13
Egr/Krox-20		*EGR1*	5q23–q31
zinc finger	*sr*	*EGR2*	10q21.1
transcription factors		*EGR3*	8p21–p23
		EGR4	2p13
Gli		*GLI*	12q13
zinc-finger transcription factors	*ci*	*GLI2*	2
glioblastoma family		*GLI3*	7p13
Src		*SRC*	20q11.2
nonreceptor tyrosine kinase	‘*Src41A*’	*YES1*	18p11
protooncogenes		*FGR*	1p36
		FYN	6q21
Src-related		*LCK*	1p34–p35
nonreceptor tyrosine kinases	*Src64B*	*LYN*	8q13
		HCK	20q11–q12
		BLK	8p22–p23
Jak		*JAK1*	1p31–p32
nonreceptor tyrosine kinases B	*hop*	*JAK2*	9p24
		JAK3	?
		TYK2	19p13.2

around 86,250 is given as the characteristic total number of protein-coding genes in all vertebrate genomes [7]. However, my estimate since 1970 has been more conservative; the realistic number for all vertebrates, excluding recent tetraploid fish and amphibians, was thought to be between 50,000 and 80,000 [9]. Inasmuch as 15,000 times 4 is 60,000, table 1 is entirely compatible with the view expressed in 1970 that gnathostomic vertebrates underwent two successive rounds of tetraploidization at their inception [9]. In short, the vertebrate genome contains four times more gene loci than the invertebrate genome. Thanks to ever-increasing genomic information, the octaploid nature of vertebrate genomes has been receiving growing support in recent years. The most revealing was the one-to-four rule proposed by Spring [10]. He pointed out that, because of the octaploid nature of vertebrate genomes, each single gene locus of *Drosophila*, as a representative of invertebrates, has been amplified, as a rule, to four paralogous genes in humans. This point is illustrated in table 2, which lists eight varieties of genes with regulatory roles [10]. Needless to say, only two or three instead of all four paralogues were occasionally encountered in the human genome. For example, it would be seen in table 2 that the gene *ci* of *Drosophila*, which encodes a zinc-finger transcription factor of the glioblastoma family, has been amplified to only three paralogues, *GLI*, *GLI2* and *GLI3*, in the human genome. Inasmuch as the human genome project is still far from completion, it is probable that a missing fourth paralogue will be found in the future. On the other hand, it will be no surprise if one or two of the original four paralogous genes have degenerated into functionless pseudo-genes. After all, the last tetraploidization event is thought to have taken place 450 million years ago at the end of the Ordovician Period. The relevance of this one-to-four rule to our understanding of interactions between various known genes involved in development and differentiation of the gonad shall now be discussed.

WT1 (11p13) and its pseudo-paralogue *EGR1* (5q23–q31)

Mice homozygously-deficient for the *Wt1* (Wilms' nephroblastoma) gene die in utero due to the absence of metanephros development, and this developmental failure apparently extends to the mesonephros as well, the absence of gonads being a necessary consequence of mesonephric failure [11]. Accordingly, *WT1* resides at the top of the regulatory hierarchy governing nephric development, of which gonadal development is a part. The vertebrate genome contains a few hundred gene loci that encode numerous families of zinc-finger transcription regulators. Of those, *WT1* is most closely related to the *Egr*/*Krox-20* family. Table 2 shows that the *sr* gene of *Drosophila* has been amplified to four *EGR*s (*early growth response* genes) in humans [10]. According to the one-to-four rule, the ready presence of *EGR1*, *EGR2*, *EGR3* and *EGR4* in the vertebrate genome implies that *WT1* is a member not of the *Egr*/*Krox-20* family itself but of its very close ally. Thus three paralogues of *WT1* are expected at most. Interestingly, it appears that whereas *WT1* exerts an inhibitory effect on nephroblast proliferation, *EGR1* promotes nephroblast proliferation by antagonizing *WT1* [12]. In view of the above, the following four experiments seem worthwhile: (i) searching for true paralogues of *WT1*; (ii) knockout of *Egr1* in mice; (iii) double knockouts of *Wt1* and *Egr1* in mice; (iv) the same knockout and double knockouts in amphibians. We recall that mesonephros persists as adult kidney in amphibians.

SF-1 (9q33), *GCNF-1* and their unidentified paralogues

In the homozygous absence of the *Sf-1* gene, mice develop neither adrenals nor gonads, two steroid hormone-synthesizing organs of vertebrates [13]. Accordingly, *SF-1* occupies the second position in the regulatory hierarchy of gonadal development. It is fitting that *SF-1* also exerts transcriptional control over genes encoding steroid hormone-synthesizing enzymes. *SF-1* is one of the genes that encode orphan nuclear receptors. These nuclear receptors without ligands are extremely ancient, as shown in an extensive phylogenic study that revealed the presence of two (COUP and FTZ-F1) types of orphan nuclear receptors in diploblastic animals lacking mesoderm of the ancient and primitive phylum Cnidaria such as a sea amenone [14]. Recall that diploblastic animals belonging to the two phyla Porifera and Cnidaria were exceptional metazoans in that they appeared considerably before the Cambrian explosion that started 530 million years ago. Table 3 shows that a COUP-type orphan nuclear receptor in *Drosophila* is encoded by the *cup* gene and that this gene has been amplified to *EAR1*, *EAR2* and *EAR3* in the human genome. Of the three *EAR*s, *EAR3* is a homologue of the chicken *COUP* gene. Inasmuch as the *COUP* gene in chicken was originally discovered as a transcriptional regulator of an ovalbumin gene, EARs, too, are not altogether unrelated to reproductive functions. The second type's namesake *FTZ-F1* in *Drosophila* is a transcriptional regulator of the *FTZ* gene that is involved in very early embryonic segmentation processes. One of its paralogues in vertebrates is *SF-1*, and the other is *GCNF-1*. Needless to say, the elucidation of *GCNF-1* function and the identification of two other paralogues of *SF-1* would be extremely rewarding.

Table 3. Nuclear receptor family.

	Invertebrates	Vertebrates
	Drosophila	human
1. Orphan		
Coup-type		*EAR1*
	cup	*EAR2*
		EAR3 (*COUP*)
FTZ-F1-type		*SF-1*
	Ftz-F1	*GCNF-1*
2. Retinoic acid		
Type A		*RARA*
	E78	*RARB*
Type X		*RXRA*
	usp	*RXRB*
		RXRG
3. Thyroid hormone, prostaglandin, vitamin D		
Thyroid hormone		*TRA*
	?	*TRB*
Prostaglandin,		*PPARA*
leukotriene	*E75*	*PPARG*
Vitamin D		*VDR*
	ECR (*ecdysone*)	*MB67*
4. Steroid hormones		
		AR (*androgen*)
	?	*MR* (*mineralocorticoid*)
		GR (*glucocorticoid*)
		PR (*progesterone*)
		ER (estrogen)
	?	*ERR1* (*estrogen*)
		ERR2 (*estrogen*)

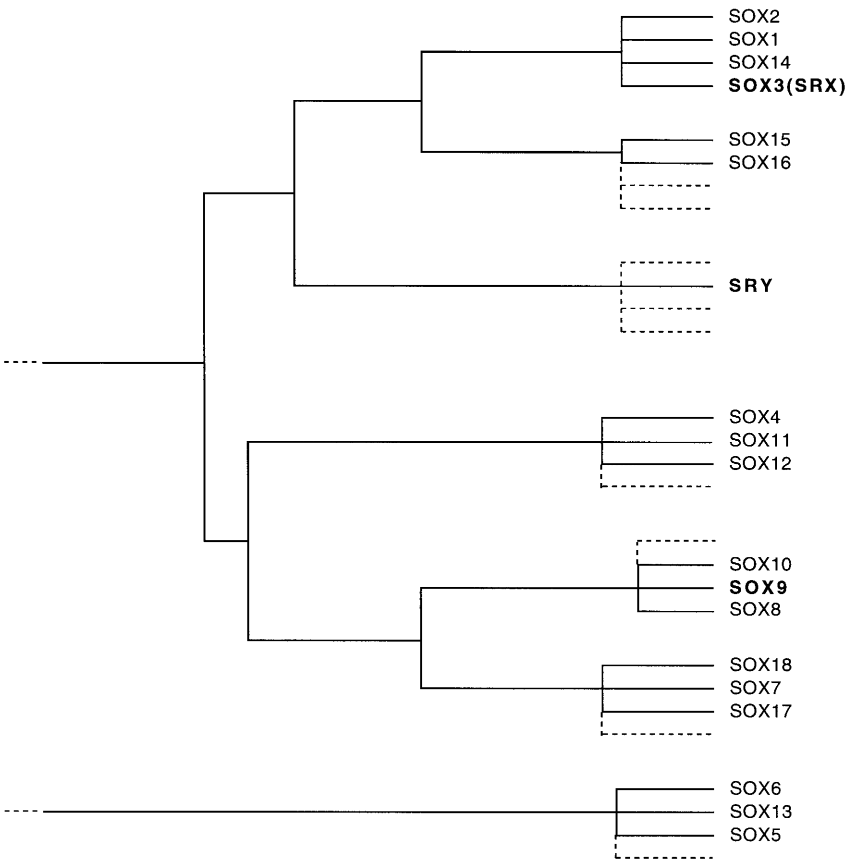

Figure 1. A highly stylized dendrogram of 19 known members of the SOX transcription factor family [15, 16]. With regard to the amino acid sequence encompassing residues 11–66 of the HMG domain, members of the same paralogous group (e.g. SOX2, SOX1, SOX14, SOX3) maintained 84% or greater identities, whereas identities between different paralogous groups were between 61% and 43%. Identities between SRY and other SOX proteins were at best 63% and at worst 47%.

The antiquity of the *SRY* (Yp11.2) gene in solitary splendor

As stated at the beginning, the *SRY* gene is found only among mammals. At first glance, this implies that *SRY* is a very recent derivative of a particular *SOX* gene paralogue. The highly schematic dendrogram of *SOX* gene families presented in figure 1 reveals otherwise. This simplified dendrogram was derived by combining data from two independent sources [15, 16]. Figure 1 shows that *SOX* genes of vertebrates, too, obey the one-to-four rule, which in turn reveals that there already were seven independent *SOX* gene loci in the invertebrate genome, one of which was an ancestor of *SRY*. Accordingly, *SRY* is closely related neither to *SOX3* nor to *SOX9*. Recall that *SOX3* is also known as *SRX*, as its human form resides on Xq26–q27. Clearly, *SOX3* and *SRY* have never been alleles of each other. Recall also that human individuals heterozygous for a defective *SOX9* gene (17q24–q25) manifest a form of osteochondro-dysplasia known as campomelic dysplasia accompanied by XY sex reversal [17]. Figure 1 shows that paralogues of *SOX9* are *SOX8* and *SOX10* but not SRY. In view of the antiquity of the *SRY* gene, an extensive search for its ancestor among various invertebrates emerges as an extremely worthwhile undertaking. If *SRY* is truly absent in vertebrate classes other than Mammalia, the *SRY* gene must be one of those genes that became extinct once and later revived, in this particular case in the mammalian ancestor. The trichrome color vision that depends upon the presence of three opsins (red, green and blue) in cone photoreceptor cells of the retina is well developed in fish as well as in birds, but color vision was secondarily lost in mammals, probably because the first mammal to emerge under the shadow of the dinosaurs was a minute nocturnal insectivore of the Cretaceous Period some 100 million years ago [18]. This loss was caused by defective mutations sustained at the two X-linked gene loci; a red opsin gene began to encode a degenerate protein, whereas a green opsin gene was extinguished. Restoration of trichrome color vision in higher primates including humans was due to rebuilding X-linked red and green opsin genes from a once degenerate red opsin gene. Indeed, a new red opsin gene and a new green opsin gene initially reappeared as alleles of a single locus [19].

The very fact that even a mammalian species can operate the sex-determining mechanism of the male heterogamety without the benefit of *SRY* [6] suggests that there is a gene which is an antagonist of *SRY*. As already noted, whereas *WT1* functions as a transcriptional repressor in nephroblastic cells, EGR1 acts as a transcription activator [12]. In the above type of antagonism, either the heterozygous deficiency of an autosomal antagonist gene or the hemizygous absence of an X-linked antagonist gene might recreate the condition normally brought about by the presence of *SRY*.

The role of sex steroid nuclear receptors in hormone-dependent sex-determining mechanisms

As Claude Pieau and colleagues point out in great detail in their own review, temperature-dependent sex-determining mechanisms widely practiced by alligators, turtles and lizards of the class Reptilia are apparent consequences of varying degrees of temperature sensitivities exhibited by individual enzymes involved in synthesis of sex steroids. At the simplest, a species whose steroid aromatase does not function well at a lower incubation temperature would produce males at that temperature simply because not enough androstenedione and testosterone are converted to oestrone and oestradiol. In fact, gonadal development of all vertebrates, excepting mammals, is under the influence of sex steroid hormones. It follows that in the vast majority of vertebrates, the pivotal role in sex determination is played by a particular branch of the above-discussed family of nuclear receptor transcription factors that utilize sex steroids as ligands. Evolutionarily as ancient as orphan nuclear receptors are retinoic acid nuclear receptors, since they were already found together with the former in diploblastic animals of the phylum Cnidaria [14]. As shown in table 3, the *Drosophila* genome as a representative of invertebrate genomes contains one ancestral gene each encoding type A and type X retinoic acid nuclear receptors. In vertebrates, both have been amplified, the former to *RARA*, *RARB* and *RARG*, and the latter to *RXRA*, *RXRB* and *RXRG* [10, 14]. Next in the evolutionary order comes a mixed bag of least-explored nuclear receptors that utilize thyroid oligopeptidic hormone, vitamin D and prostaglandins as well as leukotrienes as ligands. Ancestral genes for some but not all of these nuclear receptors in this mixed bag have been identified in the invertebrate genome. For example, *Drosophila E75* is an apparent ancestor of *PPARA* and *PPARG*, and ecdysone (insect metamorphosis hormone) receptor gene *Ecr* of *Drosophila* is ancestral to vitamin D receptor paralogues *VDR* and *MB67*, as shown in table 3 [14]. Nuclear receptors that utilize steroid hormones as ligands are present only in vertebrates, and table 3 shows that they form two independent paralogous groups. Androgen receptor *AR*, mineralocorticoid receptor *MR*, glucocorticoid receptor *GR* and progesterone receptor *PR* are four paralogues descended from one ancestral invertebrate gene. The other ancestral gene has been amplified to three known paralogous estrogen receptors in vertebrates. Such a redundancy has likely created functional differentiation among *ER* paralogues. In certain vertebrates, but not in mammals, one such *ER* paralogue might have begun to control the transcription of a set of genes responsible for differentiation toward an ovary of the indifferent gonad thereby creating the estrogen-dependent autocrine sex-determining mechanism.

1 Matsuo T., Osumi-Yamashita N., Noji S., Ohuchi H., Koyama E., Myokai F. et al. (1993) A mutation in the Pax-6 gene in rat *small* eye is associated with impaired migration of midbrain crest cells. Nature Genet. **3:** 299–304

2 Halder G., Callaerts P. and Gehring W. J. (1995) Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*. Science **267:** 1788–1792

3 Wachtel S. S., Ohno S., Koo G. C. and Boyse E. A. (1975) Possible role of H-Y antigen in primary determination of sex. Nature **257:** 235–236

4 Sinclair A. H., Berta P., Palmer M. S., Hawkins J. R., Griffiths B. L., Smith M. J. et al. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature **346:** 240–244

5 Koopman P., Gubbay J., Vivian N., Goodfellow P. and Lovell-Badge R. (1991) Male development of chromosomally female mice transgenic for *Sry*. Nature **351:** 117–121

6 Just W., Rau W., Vogel W., Akhverdian M., Fredga K., Graves J. A. M. et al. (1995) Absence of *Sry* in species of the vole *Ellobius*. Nature Genet. **11:** 117–118

7 Simmen M. W., Leitgeb S., Clark V. H., Jones S. J. M. and Bird A. (1998) Gene number in an invertebrate chordate, *Ciona intestinalis*. Proc. Natl. Acad. Sci. USA **95:** 4437–4440

8 Brenner S., Elgar G., Sandford R., Macrae A., Venkatesh B. and Aparicio S. (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. Nature **366:** 265–268

9 Ohno S. (1970) Evolution by Gene Duplication, Springer, Berlin

10 Spring J. (1997) Vertebrate evolution by interspecific hybridisation – are we polyploid? FEBS Lett. **400:** 2–8

11 Kreidberg J. A., Sariola H., Loring J. M., Maeda M., Pelletier J., Housman D. et al. (1993) WT-1 is required for early kidney development. Cell **74:** 679–691

12 Madden S. L., Cook D. M., Morris J. F., Gashler A., Sukhatme V. P. and Rauscher III F. J. (1991) Transcriptional repression mediated by the WT1 Wilms tumor gene product. Science **253:** 1550–1553

13 Luo X., Ikeda Y. and Parker K. L. (1994) A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. Cell **77:** 481–490

14 Escriva H., Safi R., Haenni C., Langlois M.-C., Saumitou-Laprade P., Stehelin D. et al. (1997) Ligand binding was acquired during evolution of nuclear receptors. Proc. Natl. Acad. Sci. USA **94:** 6803–6808

15 Wright E. M., Snopek B. and Koopman P. (1993) Seven new members of the *Sox* gene family expressed during mouse development. Nucleic Acids Res. **21:** 744

16 Stock D. W., Buchanan A. V., Zhao Z. and Weiss K. M. (1996) Numerous members of the Sox family of HMG box-containing genes are expressed in developing mouse teeth. Genomics **37:** 234–237

17 Meyer J., Suedbeck P., Held M., Wagner T., Schmitz M. L., Bricarelli F. D. et al. (1997) Mutational analysis of the *SOX9* gene in campomelic dysplasia and autosomal sex reversal: lack of genotype/phenotype correlations. Hum. Mol. Genet. **6:** 91–98

18 Ohno S. (1967) Sex Chromosomes and Sex-Linked Genes, Springer, Berlin

19 Jacobs G. H. and Neitz J. (1987) Inheritance of color vision in a New World monkey (*Saimiri sciureus*). Proc. Natl. Acad. Sci. USA **84:** 2545–2549