

Research article

Open Access

## Comparative genomics of vertebrate Fox cluster loci

Karl R Wotton and Sebastian M Shimeld\*

Address: Department of Zoology, University of Oxford, The Tinbergen Building, South Parks Road, Oxford, OX1 3PS, UK

Email: Karl R Wotton - karl.wotton@balliol.ox.ac.uk; Sebastian M Shimeld\* - sebastian.shimeld@zoo.ox.ac.uk

\* Corresponding author

Published: 24 October 2006

Received: 25 July 2006

BMC Genomics 2006, 7:271 doi:10.1186/1471-2164-7-271

Accepted: 24 October 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/271>

© 2006 Wotton and Shimeld; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Vertebrate genomes contain numerous duplicate genes, many of which are organised into paralogous regions indicating duplication of linked groups of genes. Comparison of genomic organisation in different lineages can often allow the evolutionary history of such regions to be traced. A classic example of this is the Hox genes, where the presence of a single continuous Hox cluster in amphioxus and four vertebrate clusters has allowed the genomic evolution of this region to be established. Fox transcription factors of the C, F, LI and QI classes are also organised in clusters in both amphioxus and humans. However in contrast to the Hox genes, only two clusters of paralogous Fox genes have so far been identified in the Human genome and the organisation in other vertebrates is unknown.

**Results:** To uncover the evolutionary history of the Fox clusters, we report on the comparative genomics of these loci. We demonstrate two further paralogous regions in the Human genome, and identify orthologous regions in mammalian, chicken, frog and teleost genomes, timing the duplications to before the separation of the actinopterygian and sarcopterygian lineages. An additional Fox class, FoxS, was also found to reside in this duplicated genomic region.

**Conclusion:** Comparison of loci identifies the pattern of gene duplication, loss and cluster break up through multiple lineages, and suggests *FoxS1* is a likely remnant of Fox cluster duplication.

### Background

The Fox genes are members of the forkhead/winged helix family of transcription factors, characterised by a 110 amino acid DNA binding domain [1]. Fox genes have been identified in the genomes of animals and fungi, but not plants. Animals appear to have more Fox genes than fungi, with four genes identified in *Saccharomyces* and *Schizosaccharomyces*, at least 15 in the cnidarian sea anemone *Nematostella vectensis*, 17 in *Drosophila melanogaster*, 29 in *Ciona intestinalis* and over 40 in the human and other vertebrate genomes [1-4]. Phylogenetic analysis of the forkhead domains has led to placement of most of these genes into 20 subclasses named *FoxA* to *FoxS*, with a

small number of 'orphan' genes of unclear relationships defying classification [5,6]. These studies show vertebrate genomes contain more Fox genes than the genomes of other animals, suggesting an expansion that mirrors the increase in gene numbers seen for other gene families [7]. Two competing theories for how this increase in gene number took place are that two whole genome duplications (WGD; referred to as the 2R hypothesis) occurred early in the vertebrate lineage [8,9], or that continuous gene duplications occurred throughout vertebrate evolution [10-14]. The 2R hypothesis predicts that a 1:4 ratio of genes should have been present in an ancestral vertebrate when compared to an invertebrate, and that in molecular

phylogenetic trees these paralogues would adopt a specific topology reflecting the history of duplication [11]. While the identification of blocks of paralogous genes has been interpreted as evidence in favour of WGD [15-24], paralogous gene sets often adopt different topologies [11], and multiple independent duplications could also explain the observed data.

Previously it has been reported that the human representatives of four Fox subclasses, *FoxC*, *FoxF*, *FoxL1* and *FoxQ1*, are localised to two small regions of the human genome; specifically, the genes *FOXC2*, *FOXF1* and *FOXQ1* are found on chromosome 6 within 70 kb, while the genes *FOXL1*, *FOXQ1*, *FOXF2* are found on chromosome 16 within 300 kb [1,25,26]. Recently, we have shown that the amphioxus orthologues of these genes are found clustered in one region of the amphioxus genome, suggesting block duplication of this region in the vertebrates underlies the evolution of the chromosome 6 and 16 loci [25]. Under the 2R hypothesis, however, four such loci would be predicted. This discrepancy could be explained by gene loss.

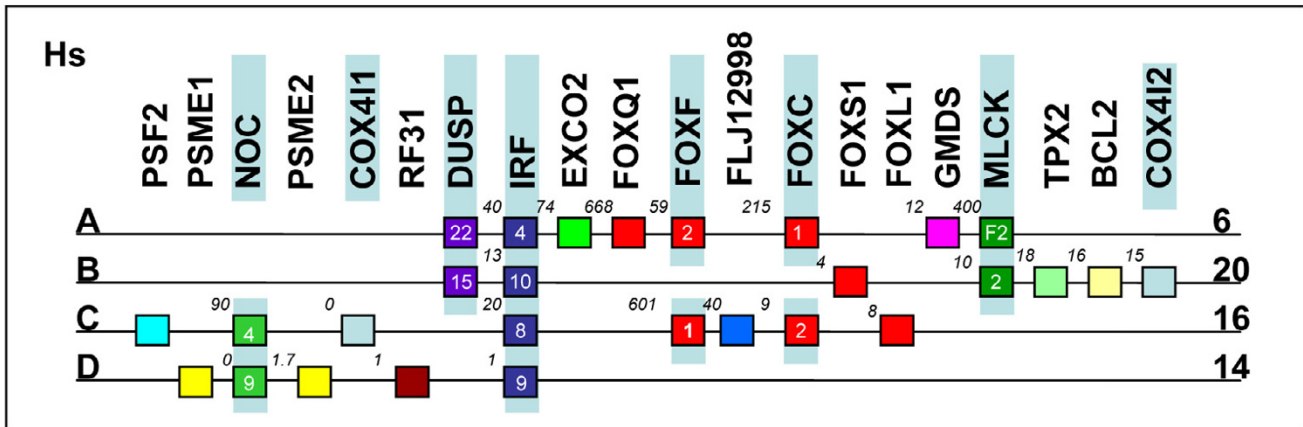
To investigate this, we carried out a detailed comparative genomic survey of the human chromosome 6 and 16 Fox cluster loci. Genes around each locus were identified, and their evolutionary history established by database searches and molecular phylogenetics. This revealed two additional human loci containing genes paralogous to those found adjacent to the clustered Fox genes. All four loci were then compared to the genomes of other vertebrates, including other mammals and *Xenopus tropicalis*,

*Gallus gallus*, *Fugu rubripes*, *Danio rerio* and *Tetraodon nigroviridis*. This allowed us to identify orthologous regions in all these animal genomes, and to show teleost genomes have undergone further duplications. By comparing these loci, we are now able to establish the pattern of clustered Fox gene duplication and loss through multiple vertebrate lineages. Furthermore, a previously orphan gene, *FoxS1*, was identified as a likely remnant of cluster duplication.

**Results**

**Identification of human Fox cluster paralogons**

Gene maps pertaining to this section are summarised in Figure 1. Our searches identified 5 gene families that showed evidence of genomic co-localisation with the Fox clusters; the Neighbour of COX (NOC) genes, the Cytochrome C Oxidase Subunit 4 (COX4) genes, the Dual Specificity Phosphatase (DUSP) genes, the Interferon Regulatory Factor (IRF) genes and the Myosin Light Chain Kinase (MLCK) genes. Phylogenetic analyses of these families are shown in Additional Files 1 and 2. To summarise, the NOC genes are found on the chromosome 16 and 14 (NOC4 and NOC9) and the COX4 genes are found on chromosome 16 and 20 (COX411 and COX412 respectively) though their exact positions relative to the other genes differ. DUSP genes are found next to the chromosome 6 and 20 clusters (DUSP22 and DUSP15). Each locus on chromosome 6, 20, 16 and 14 contains an IRF gene (IRF4, IRF10, IRF8 and IRF9 respectively). The MLCK genes are found by the chromosome 6 (MLCKF2) and 20 (MLCK2) clusters while a third gene, MLCKF1 is found on chromosome 16, though this last gene is not tightly linked



**Figure 1**  
**Human paralogons of the FOX cluster loci.** A schematic representation of paralogy deduced from analysis of the human genome using NCBI map viewer Hs Build 36.1 (Nov 01, 2005) [36]. Gene names are shown above each gene, while numbers in the boxes indicate subfamily. Letters on the far left indicate paralogon designation, while numbers on the far right indicate which chromosome the cluster is found on. Numbers between genes are approximate intergenic distances in Kb. Background shading indicates paralogy.

to the rest of the cluster (and hence is not shown on Figure 1). Furthermore, another Fox gene, *FOXS1*, which has been previously regarded as an orphan [5,6], was identified adjacent to *IRF10* on chromosome 20. We therefore conclude regions of chromosomes 14 and 20 shown in Figure 1 are the likely remaining paralogous genomic regions (paralogons) of the chromosome 6 and 16 Fox cluster regions.

#### **Orthologous paralogons in mammals, chicken, *Xenopus* and teleosts**

Conserved genomic regions to the four human loci described above were identified in the *M. musculus*, *X. tropicalis*, *G. gallus*, *F. rubripes*, *D. rerio* and *T. nigroviridis* genomes. These are summarised in Figure 2. We also examined chimp, dog and cow genomes. Gene organisation appeared similar to that in mouse and human (within the limits of the quality of genome assembly for these species; data not shown), and hence we have omitted them from the following sections. Further details of the phylogenetic analyses that confirm orthology and paralogy relationships are shown in the additional files that accompany this report. Below, we summarise the salient features of each orthologous set. In addition to the paralogous genes described in humans above, we also utilised other genes present in these genomic regions to re-enforce evidence of orthology.

#### **Genomic regions orthologous to human chromosome 6**

This region consists of *DUSP22*, *IRF4*, *EXCO2*, *FOXQ1*, *FOXF2*, *FOXC1*, *GMD5* and *MLCKF2*. Genomic regions orthologous to human chromosome 6 are summarised in Figure 2A. A region of mouse chromosome 13 (and of chimp chromosome 6 and dog chromosome 35, data not shown) showed the order and intergenic distance of these genes to be comparable in these mammals. In the chicken, a region of chromosome 2 is orthologous to the same region, and includes the linked *FOXF2* and *FOXQ1* genes. A chicken *FOXC1* gene was identified in a previous study [27], however its genomic position is currently unknown. We also note a large gap in the sequence between chicken *FOXF2* and *GMD5*, and suggest chicken *FOXC1* may reside here. In *X. tropicalis* this region is contained on two scaffolds; *dusp22*, *irf4* and *exco5* are on scaffold 211, and *foxq1*, *foxf2*, *foxc1*, *gmd5* and *mlckf2* on scaffold 95. High quality sequence is present for 150 kb and 230 kb past the final genes on each respective scaffold. However, as the gap between *Exoc2* and *FoxQ1* can exceed 500 kb in mammals, this does not preclude the linkage of these regions in the *X. tropicalis* genome.

A region in the *D. rerio* genome was identified on chromosome 20, extending from *irf4* to *mlckf2* and included identical gene organisation but relatively small intergenic distances when compared to tetrapods. *foxq1*, *foxf2* and

*foxc1b* are present in the expected positions. An additional contig containing linked *irf4*, and *dusp22* orthologs was also identified. In the *T. nigroviridis* genome, two Fox clusters orthologous to human chromosome 6 were identified, one on chromosome 15 (extending from *foxq1* to *mlckf2*) and one on unassigned scaffold 14546. An additional unassigned scaffold contains *irf4a* and *dusp22*, and could also be part of the chromosome 15 region as sequence quality is low adjacent to *foxq1*. This is supported by the *F. rubripes* arrangement, in which the orthologous genes are found on one scaffold.

In summary, we found evidence for single genomic regions orthologous to the Fox cluster region of human chromosome 6 in tetrapods, while in teleost fish we found evidence for the presence of two such regions, in keeping with the additional genome duplication proposed to have occurred in this lineage [28-31].

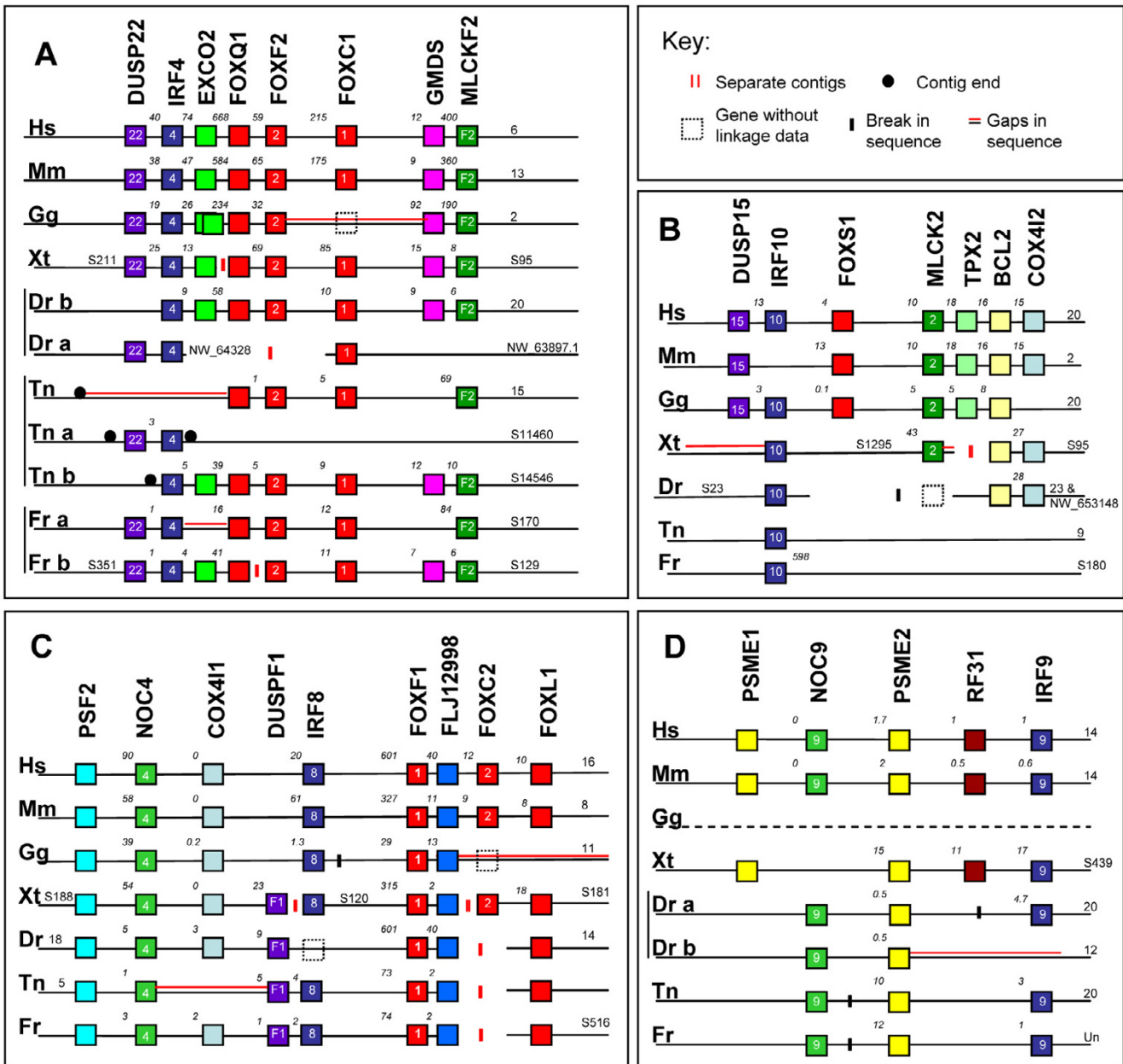
#### **Genomic regions orthologous to human chromosome 20**

This region consists of *DUSP15*, *IRF10*, *FOXS1*, *MLCK2*, *TPX2*, *BCL2* and *COX4I2*. Genomic regions orthologous to human chromosome 20 are summarised in Figure 2B. A region of mouse chromosome 2 (and of chimp chromosome 20, cow chromosome 13 and dog chromosome 24, data not shown) showed the order and intergenic distance of these genes to be comparable in these mammals. An exception is that no *Irf10* gene was identified in this region of the mouse genome. In the chicken a single orthologous region was identified on chromosome 20, and included a *FOXS1* orthologue. However we were unable to identify a *COX4I2* gene. In *X. tropicalis*, two scaffolds (1295 and 95) were identified as orthologous to this human region; these do not overlap in gene content, hence we infer they derive from one region of orthology. *foxs1* was not found in the *X. tropicalis* genome.

Evidence for orthologous regions in teleost genomes is weaker. A tenuous orthology region may exist in the *D. rerio* genome, as *bcl2* and *cox4* genes exhibit tight linkage on chromosome 23. An *irf10* orthologue is also on chromosome 23, but at 44.65 Mb from the other genes, this linkage is unlikely to be significant. In *T. nigroviridis* and *F. rubripes*, we were only able to identify *irf10* orthologues.

#### **Genomic regions orthologous to human chromosome 16**

This region contains *PSF2*, *NOC4*, *COX4I1*, *IRF8*, *FOXF1*, *FLJ12998*, *FOXC2* and *FOXLI*. Genomic regions orthologous to human chromosome 16 are summarised in Figure 2C. Investigation of mammalian genomes showed regions of chromosome 8 of the mouse (and of chimp chromosome 16, cow chromosome 18 and dog chromosome 5, data not shown) to be comparable in gene order and intergenic distance to the genes on human chromosome 16. Chromosome 11 of the chicken genome has a



**Figure 2**  
**Genomic organisation of human Fox cluster paralogs and putative orthologous counterparts.** Genomic organisation of human (Hs) Fox cluster paralogs and the putative orthologous counterparts we have identified in *Mus musculus* (Mm), *Xenopus tropicalis* (Xt), *Gallus gallus* (Gg), *Danio rerio* (Dr), *Tetraodon nigroviridis* (Tn) and *Fugu rubripes* (Fr). Panels A, B, C and D depict orthologous genomic regions to the human regions on chromosomes 6, 20, 16 and 14 respectively, as shown in Figure 1. This is a schematic diagram and not to scale. Colour coding indicates orthologous genes inferred by molecular phylogenetics (see additional files). Numbers at the ends of each line indicate chromosome number, or, where this is not available, scaffold number, with the latter indicated by S or NW. Numbers above lines indicate approximate distance in Kb between genes. A parallel red line indicates gaps in the sequence, while breaks in the sequence are indicated by double vertical black lines at the site of inversions. Double vertical red lines indicate separate contigs that we placed on a line due to their gene content; this does not imply they have been shown to be physically linked. Dashed boxes represent presence of a gene but no linkage information. Black circles indicate scaffold ends.

region orthologous to human chromosome 16, comprising of *PSF2*, *NOC4*, *COX4I1*, *IRF8* and *FOXF1*. However part of this region is inverted in comparison to mammalian genomes. Similarity past *FOXF1* extends to a region containing *FLJ12998*, which is found between *FOXF1* and *FOXC2* in the human genome. A chicken *FOXL1* is yet to be identified, however the sequence quality upstream of *FoxF1* in the chicken genome is poor.

Orthologous regions in the *X. tropicalis* genome are recovered from three separate scaffolds: s188 containing *psf2*, *noc4*, *cox411* and a *dusp* family member. s120 contains *irf8*, *foxf1* and *flj12998* and s181 contains the *foxc2* and *foxl1* genes. Truncation of scaffolds 188 and 120 at the *irf8* end do not preclude close linkage, however good sequence quality past *flj12998* and before *foxc2* suggest these regions, if linked, are minimally 3.5 Mb apart.

In *D. rerio*, an orthologous region on chromosome 18 matches the human arrangement from *psf2* to a region with similarity to *flj12998* with the exceptions that *D. rerio* (like *X. tropicalis* but unlike chicken and mammals) also has a *dusp* gene in this region, and that *irf8* is absent and instead found on a contig yet to be assigned to a chromosome. Sequence quality past *foxf1* appears to be high, however no *foxc2* has been found in the genome of these teleosts. *foxc2* is present in the basal Actinopterygian *A. calva* (see additional files), showing loss of *foxc2* to be specific to teleosts. A *foxl1* gene has been identified on chromosome 14. The arrangement in *T. nigroviridis* and *F. rubripes* is similar except *irf8* is included in these clusters.

In summary, we find good evidence for orthologous regions to the human chromosome 16 Fox cluster region in the genomes of tetrapods and teleosts. However teleosts appear to have split this region between *flj12998* and *foxl1*. An alternate explanation for this apparent split is they are remnants of the proposed teleost tetraploidy, with reciprocal gene loss yielding non-overlapping gene sets. Higher quality genome assemblies will be needed to distinguish between these possibilities.

#### Genomic regions orthologous to human chromosome 14

This region consists of *PSME1*, *NOC9*, *PSME2*, *RF31* and *IRF9*. Genomic regions orthologous to human chromosome 14 are summarised in Figure 2D. Regions identified on mouse chromosome 14 (and on chimp chromosome 14, dog chromosome 8 and cow chromosome 10, data not shown), suggest preservation of this orthologous region in mammals. We were unable to identify a similar orthologous region (or any of the genes it contains) in the chicken genome, however a similar region was identified in the genome of *X. tropicalis*, suggesting the chicken has either lost these genes, or they are not included in the current assembly and associated sequence data. The *D. rerio*

genome contains an *irf9*, *noc9*, *psme2* orthology group on chromosome 20, with an inversion between *irf9* and *psme2*. A second region on chromosome 12 includes a *noc9* gene linked to a *psme2* gene. The *T. nigroviridis* and *F. rubripes* genomes also contain a single region including *irf9*, *noc9* and *psme2*, but with an inversion between *noc9* and *psme2*.

## Discussion

### Human paralogs and orthologous regions in other vertebrates: gene duplication and gene loss

Here we report the results of a comparative analysis of vertebrate Fox cluster loci. First we identified four putative paralogs in the human genome; on chromosome 6 (with *DUSP22*, *IRF4*, *FOXQ1*, *FOXF2*, *FOXC1*, *MLCKF2* linkage); on chromosome 20 (with *DUSP15*, *IRF10*, *FOXS1*, *MLCK2*, *COX4I2* linkage); on chromosome 16 (with *NOC4*, *COX4I1*, *IRF8*, *FOXF1*, *FOXC2*, *FOXL1* linkage); and on chromosome 14 (with *NOC9*, *IRF9* linkage). Of these, that on chromosome 14 has the weakest supporting evidence, though IRF gene phylogenetics and the preservation of *NOC9* and *IRF9* linkage in other genomes supports its status as a paralogous region.

We also identified putative orthologous regions in the genomes of other vertebrates. Organisation in all mammals examined was very similar to that in human. In the tetrapods *G. gallus* and *X. tropicalis*, gene organisation was consistent with the ancestral tetrapod genome including a single copy of all four regions [32]. Some lineage specific changes are also inferred, and one paralogon is currently missing in *G. gallus*. The situation in the teleost genomes is more complex, with two putative orthologous regions identified for some of the loci. The teleost lineage is hypothesised to have undergone an additional genome duplication [28-31], and the pattern of orthologous regions we observed is consistent with this, though this would imply significant gene loss from several duplicated regions. Consistent with this we note that as high as 85% gene loss have been suggested for the teleosts following WGD [33].

### The evolutionary relationship of FoxS1

*FoxS1* has previously been named in humans as *FKHL18* [34] and in mouse as *Fkh3* [35], however its phylogenetic relationship to Fox genes in more distant taxa has remained unresolved [5,6]. Our study extends the number of known *FoxS1* genes to chimp, cow, dog and chicken orthologues, indicating the gene is at least as old as the mammal and bird lineage divergence. The consistent placement of *FoxS1* in a genomic region paralogous to the clustered Fox genes suggests its evolutionary origin from the original Fox gene cluster. This is because the presence of orthologous regions in tetrapod and teleosts imply the origin of these paralogs, including the *FoxS1* gene, via

duplication prior to the separation of the actinopterygian and sarcopterygian lineages. Assuming this, it is possible the original pre-duplication Fox gene cluster included five genes (*FoxC*, *FoxL1*, *FoxF*, *FoxQ1* and *FoxS1*), and that these have been fragmented by duplication and reciprocal gene loss. A literal interpretation of Fox gene phylogeny supports this view, as the *FoxS1* genes group together, well separated from the other Fox subclasses. However, Fox gene phylogenies (including published accounts [5,6] and our study) are necessarily based on the relatively short sequence of the forkhead domain. Rapid divergence could lead to long branch attraction artefacts in such phylogenies, and hence we consider it possible that *FoxS1* is paralogous to *FoxC*, *FoxL1*, *FoxF* or *FoxQ1*, and that this relationship has become obscured.

Furthermore, no *foxs1* gene was recovered from the genomes of the teleosts or of *X. tropicalis*. We are currently unable to determine if this gene has been lost from these lineages, or is present but has not yet been sequenced. If the latter, the sequence quality in *X. tropicalis* and *T. nigroviridis* renders it unlikely it exists in the orthologous region of these genomes. This would raise a third possibility, that *FoxS1* had independently translocated into this genomic region after the separation of the amphibian and amniote lineages. Sequencing of further vertebrate genomes may help resolve this issue.

#### **Fox cluster break-up in vertebrates**

Retention of linkage between clustered Fox genes observed in humans, amphioxus and some insects has been hypothesised to have been constrained by co-ordinated regulation, analogous to that suggested for the homeobox genes [25]. Our results show the retention of Fox clusters in mammals. The chicken genome assembly is currently of insufficient quality to determine if the Fox clusters are intact, however in *X. tropicalis*, while *foxq1-foxf2-foxc1* linkage is maintained, the *foxf1* and *foxc2-foxl1* genes are separated by at least 3.5 Mb, and could lie on different chromosomes.

In all three teleost genomes we found evidence for *foxq1-foxf2-foxc1* linkage, however again the *foxf1-foxc2-foxl1* cluster appears broken, and *foxc2* appears to have been lost. In conclusion, our data do not disprove the possibility of constraint on the organisation of the *FoxQ1-FoxF2-FoxC1* genes, but suggest such constraints, if they exist, are less likely on the organisation of the *FoxF1-FoxC2-FoxL1* genes.

In summary, our data are consistent with the origin of paralogous regions via duplication of large regions of DNA. The timing of the duplications is consistent with that proposed for three of the genome duplications suggested to have occurred in vertebrate evolution, but also consistent

with independent block duplication of this genomic region. Based on the pattern of duplication, we have constructed a model indicating how the Fox cluster locus evolved in vertebrates (Figure 3). Duplication of the whole region to yield four copies is inferred to have occurred by the base of bony vertebrates. Considerable gene loss is then inferred before the radiation of bony vertebrates. Subsequent losses and inversions in different vertebrate lineages are indicated in the figure.

#### **Conclusion**

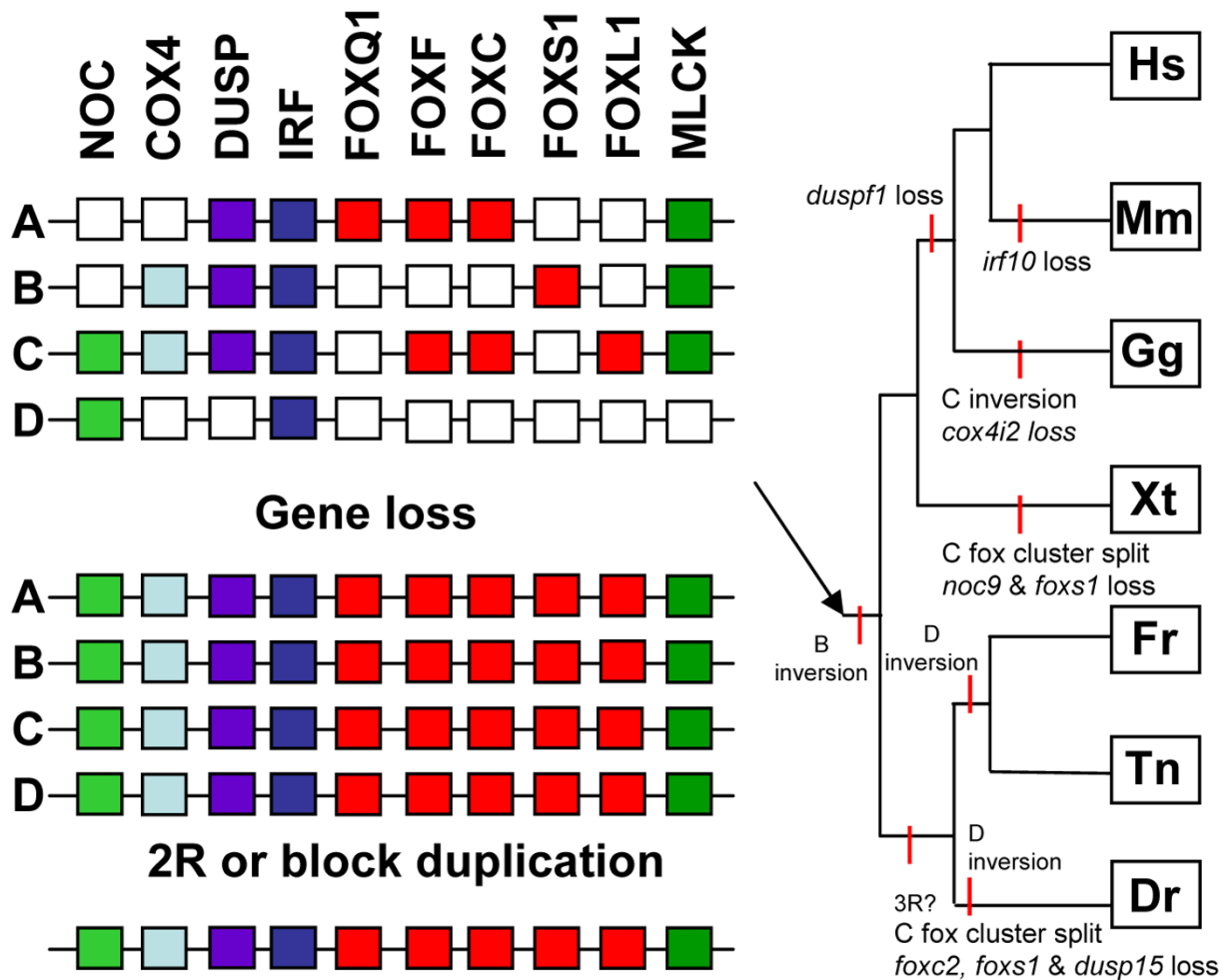
Comprehensive analysis of the human Fox cluster loci has identified 4 paralogs, 3 containing Fox genes and one Fox-less 'cryptic paralogon'. Extension of this comparison to other vertebrates has allowed a model for the evolution of this genomic region to be constructed, indicating patterns of gene duplication, loss, inversion and Fox cluster break up. To obtain a more complete picture of the evolution of the Fox clusters we note the importance of including basal actinopterygian, chondrichthian and agnathan data in future analyses to gain a greater understanding of the evolutionary history and timing of these events.

#### **Methods**

Regions surrounding the human FOX cluster loci were searched using NCBI Map Viewer build 36.1 [36]. To identify further paralogous genes, BLAST searches [37] with putative protein sequences were conducted against the human genome, and putative positive targets further characterised by molecular phylogenetics to resolve orthologous and paralogous relationships. Each of the genes from the putative human paralogs was cross referenced by BLAST against the following NCBI genome assemblies; *Pan troglodytes* (build 1.1), *Bos taurus* (build 2.1), *Canis familiaris* (build 2.1), *Mus musculus* (build 3.5), *Gallus gallus* (build 1.1) and *Danio rerio* (Zv4) on NCBI map viewer. The *Xenopus tropicalis* and *Fugu rubripes* genomes were searched at the Joint Genome Institute genome portal [38], and the *Tetraodon nigroviridis* genome searched by BLAT via the Genoscope server [39]. Putative target sequences were further analysed by molecular phylogenetics, to establish orthologous and paralogous relationships and to reaffirm our original classification of human sequences. For molecular phylogenetic analyses, protein sequences were first aligned, (with invertebrate outgroup sequences included where possible) using ClustalX [40]. Phylogenetic analyses were carried out using Neighbor Joining and maximum likelihood implemented by ClustalX, and by PHYML [41,42].

#### **Authors' contributions**

KRW conducted the database searches and phylogenetic analyses, and helped draft the manuscript. SMS helped draft the manuscript. Both authors read and approved the final manuscript.



**Figure 3**

**A model for gene duplication and loss in the Fox cluster loci during vertebrate evolution.** Two initial rounds of duplication of the whole region are inferred to have occurred by the base of bony vertebrates, and this is consistent with both the 2R hypothesis and multiple block duplications. Considerable gene loss is then inferred before the radiation of bony vertebrates. Subsequent losses and inversions in different vertebrate lineages are indicated on the tree, with capital letters indicating the paralogon (nomenclature as shown in Figures 1 and 2) affected. A third round of genome duplication (3R) is indicated in the teleosts.

## Additional material

### Additional File 1

*Gene family introduction and phylogeny interpretation. Additional file 1 gives a brief introduction to each gene family, followed by a summary of the molecular phylogenetic analyses and their interpretation. This is followed by a table giving accession numbers or other identifiers for the sequences used in these analyses. A figure showing NCBI Map Viewer data of paralogous regions of the human genome is also included.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-271-S1.pdf>]

### Additional File 2

*Phylogenetic trees. Phylogenetic trees of the Fox, MLCK, DUSP, COX, NOC and IRF families used in this analysis.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-271-S2.pdf>]

## Acknowledgements

We thank Peter Holland and Dave Ferrier for their discussions on this project and John Mulley for supplying the *Amia calva* genomic DNA, and for comments on the manuscript. This work was supported by the BBSRC.

## References

- Carlsson P, Mahlapuu M: **Forkhead transcription factors: key players in development and metabolism.** *Dev Biol* 2002, **250**:1-23.
- Magie CR, Pang K, Martindale MQ: **Genomic inventory and expression of Sox and Fox genes in the cnidarian *Nematostella vectensis*.** *Dev Genes Evol* 2005, **215**:618-630.
- Lee HH, Frasch M: **Survey of forkhead domain encoding genes in the *Drosophila* genome: Classification and embryonic expression patterns.** *Dev Dyn* 2004, **229**:357-366.
- Yagi K, Satou Y, Mazet F, Shimeld SM, Degnan B, Rokhsar D, Levine M, Kohara Y, Satoh N: **A genomewide survey of developmentally relevant genes in *Ciona intestinalis* III. Genes for Fox, ETS, nuclear receptors and NFkappaB.** *Dev Genes Evol* 2003, **213**:235-244.
- Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM: **Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria.** *Gene* 2003, **316**:79-89.
- Kaestner KH, Knochel W, Martinez DE: **Unified nomenclature for the winged helix/forkhead transcription factors.** *Genes Dev* 2000, **14**:142-146.
- Holland PW: **More genes in vertebrates?** *J Struct Funct Genomics* 2003, **3**:75-84.
- Holland PW, Garcia-Fernandez J, Williams NA, Sidow A: **Gene duplications and the origins of vertebrate development.** *Dev Suppl* 1994:125-133.
- Ohno S: **Evolution by gene duplication.** New York, Springer-Verlag; 1970.
- Friedman R, Hughes AL: **Pattern and timing of gene duplication in animal genomes.** *Genome Res* 2001, **11**:1842-1847.
- Hughes AL: **Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history.** *J Mol Evol* 1999, **48**:565-576.
- Hughes AL, Friedman R: **2R or not 2R: testing hypotheses of genome duplication in early vertebrates.** *J Struct Funct Genomics* 2003, **3**:85-93.
- Hughes AL: **Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1.** *Mol Biol Evol* 1998, **15**:854-870.
- Martin A: **Is tetralogy true? Lack of support for the "one-to-four rule".** *Mol Biol Evol* 2001, **18**:89-93.
- Dehal P, Boore JL: **Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate.** *PLoS Biol* 2005, **3**:e314.
- Lundin LG: **Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse.** *Genomics* 1993, **16**:1-19.
- Pebusque MJ, Coulier F, Birnbaum D, Pontarotti P: **Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution.** *Mol Biol Evol* 1998, **15**:1145-1159.
- Gibson TJ, Spring J: **Evidence in favour of ancient octaploidy in the vertebrate genome.** *Biochem Soc Trans* 2000, **28**:259-264.
- Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H: **Evidence of en bloc duplication in vertebrate genomes.** *Nat Genet* 2002, **31**:100-105.
- Castro LF, Holland PW: **Chromosomal mapping of ANTP class homeobox genes in amphioxus: piecing together ancestral genomes.** *Evol Dev* 2003, **5**:459-465.
- Luke GN, Castro LF, McLay K, Bird C, Coulson A, Holland PW: **Dispersal of NK homeobox gene clusters in amphioxus and humans.** *Proc Natl Acad Sci U S A* 2003, **100**:5292-5295.
- Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M, Lehrach H: **New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes.** *Genome Res* 2003, **13**:1056-1066.
- Castro LF, Furlong RF, Holland PW: **An antecedent of the MHC-linked genomic region in amphioxus.** *Immunogenetics* 2004, **55**:782-784.
- Vienne A, Rasmussen J, Abi-Rached L, Pontarotti P, Gilles A: **Systematic phylogenomic evidence of en bloc duplication of the ancestral 8p11.21-8p21.3-like region.** *Mol Biol Evol* 2003, **20**:1290-1298.
- Mazet F, Amemiya CT, Shimeld SM: **An ancient Fox gene cluster in bilaterian animals.** *Curr Biol* 2006, **16**:R314-6.
- Kaestner KH, Bleckmann SC, Monaghan AP, Schlondorff J, Mincheva A, Lichter P, Schutz G: **Clustered arrangement of winged helix genes *flh-6* and *MFH-1*: possible implications for mesoderm development.** *Development* 1996, **122**:1751-1758.
- Buchberger A, Schwarzer M, Brand T, Pabst O, Seidl K, Arnold HH: **Chicken winged-helix transcription factor *cFKH-1* prefigures axial and appendicular skeletal structures during chicken embryogenesis.** *Dev Dyn* 1998, **212**:94-101.
- Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y: **Genome duplication, a trait shared by 22000 species of ray-finned fish.** *Genome Res* 2003, **13**:382-390.
- Taylor JS, Van de Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:1661-1679.
- Wittbrodt J MASM: **More genes in fish?** *Bio-Essays* 1998, **20**:511-515.
- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH: **Zebrafish *hox* clusters and vertebrate genome evolution.** *Science* 1998, **282**:1711-1714.
- Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H: **Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431**:946-957.
- Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M: **Gene Loss and Evolutionary Rates Following Whole Genome Duplication in Teleost Fishes.** *Mol Biol Evol* 2006, **23**:1808-1816.
- Cederberg A, Betz R, Lagercrantz S, Larsson C, Hulander M, Carlsson P, Enerback S: **Chromosome localization, sequence analysis, and expression pattern identify FKHL 18 as a novel human forkhead gene.** *Genomics* 1997, **44**:344-346.



35. Kaestner KH, Lee KH, Schlondorff J, Hiemisch H, Monaghan AP, Schutz G: **Six members of the mouse forkhead gene family are developmentally regulated.** *Proc Natl Acad Sci U S A* 1993, **90**:7628-7631.
36. **NCBI Map Viewer** [<http://www.ncbi.nlm.nih.gov/mapview>]
37. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
38. **Joint Genome Institute genome portal** [<http://genome.igi-psf.org/>]
39. **Genoscope** [<http://www.genoscope.cns.fr>]
40. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
41. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
42. Guindon S, Lethiec F, Duroux P, Gascuel O: **PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference.** *Nucleic Acids Res* 2005, **33**:W557-9.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

