Early evolution of the T-box transcription factor family

Arnau Sebé-Pedrós[a,b,1], Ana Ariza-Cosano[c,1], Matthew T. Weirauch[d], Sven Leininger[e], Ally Yang[f], Guifré Torruella[a], Marcin Adamski[e], Maja Adamska[e], Timothy R. Hughes[f], José Luis Gómez-Skarmeta[c,2], and Iñaki Ruiz-Trillo[a,b,g,2]

[a]Institut de Biologia Evolutiva (Consejo Superior de Investigaciones Científicas-Universitat Pompeu Fabra), 08003 Barcelona, Spain; [b]Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, Spain; [c]Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas, Universidad Pablo de Olavide-Junta de Andalucía, 41013 Sevilla, Spain; [d]Center for Autoimmune Genomics and Etiology and Divisions of Rheumatology and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; [e]Sars International Centre for Marine Molecular Biology, 5008 Bergen, Norway; [f]Terrence Donnelly Centre and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada M5S 3E1; and [g]Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved August 13, 2013 (received for review May 24, 2013)

Developmental transcription factors are key players in animal multicellularity, being members of the T-box family that are among the most important. Until recently, T-box transcription factors were thought to be exclusively present in metazoans. Here, we report the presence of T-box genes in several nonmetazoan lineages, including ichthyosporeans, filastereans, and fungi. Our data confirm that Brachyury is the most ancient member of the T-box family and establish that the T-box family diversified at the onset of Metazoa. Moreover, we demonstrate functional conservation of a homolog of Brachyury of the protist *Capsaspora owczarzaki* in *Xenopus laevis*. By comparing the molecular phenotype of *C. owczarzaki* Brachyury with that of homologs of early branching metazoans, we define a clear difference between unicellular holozoan and metazoan Brachyury homologs, suggesting that the specificity of Brachyury emerged at the origin of Metazoa. Experimental determination of the binding preferences of the *C. owczarzaki* Brachyury results in a similar motif to that of metazoan Brachyury and other T-box classes. This finding suggests that functional specificity between different T-box classes is likely achieved by interaction with alternative cofactors, as opposed to differences in binding specificity.

origin multicellularity | premetazoan evolution | subfunctionalization | Porifera | Holozoa

Transcriptional regulation is a central aspect of animal development. Thus, deciphering the early evolution of metazoan transcription factors is vital for achieving a better understanding of the origin of animals. The T-box family of genes is among the most important developmental transcription factors present in Metazoa. This family is characterized by an evolutionary conserved DNA-binding domain of 180–200 amino acids, known as the T-box domain (1–3). Brachyury is the founding and best-characterized member of the T-box family, with well-established roles in blastopore specification, mesoderm differentiation and, in chordates, notochord formation (4–6). It has been hypothesized that the ancestral role of Brachyury was primarily that of blastopore determination and gastrulation (5, 7).

Other T-box classes include Tbx4/5, Tbx6, Tbx2/3, Eomes, and Tbx1/15/20. With only a few exceptions (8), all classes of T-box genes are widespread among bilaterian animals, with a handful being identified and studied in nonbilaterian metazoans, such as cnidarians (5, 9), ctenophores (7, 10), and sponges (11–14). T-box genes were initially thought to be specific to metazoans (13, 15), but two recent studies revealed the presence of T-box genes in nonmetazoan lineages (14, 16), including the unicellular filose amoeba *Capsaspora owczarzaki*, a close relative of animals, and the chytrid fungus *Spizellomyces punctatus*. T-box genes were not identified in any other sequenced eukaryote, suggesting that T-box genes were secondarily lost in choanoflagellates (both in unicellular and colonial species) and most fungi. Interestingly, one of the T-box genes identified in *C. owczarzaki* is a homolog of Brachyury, making it the only Brachyury gene identified outside of metazoans to date (16). However, the degree of conservation between *C. owczarzaki* and metazoan Brachyury genes and whether T-box genes are present in other unicellular lineages remained unclear.

Here, we report a taxon-wide survey of T-box genes in several eukaryotic genomes and transcriptomes, including previously undescribed genomic data from several close relatives of metazoans, such as the other known filasterean species (*Ministeria vibrans*) and several ichthyosporean taxa (17), as well as genomic data from calcarean sponges. We identify T-box genes in *M. vibrans*, in all of the ichthyosporeans, and in several early-branching Fungi. Our data pinpoints with unprecedented detail the evolutionary history of T-box transcription factors. We also confirm that Brachyury is the founding member of the T-box family and define previously undescribed classes of T-box genes.

To obtain a glimpse into the functional conservation of the earliest Brachyury genes, we perform heterologous expression experiments of the Brachyury homologs from *C. owczarzaki*, *Sycon ciliatum* (Calcarea, Porifera), and *Nematostella vectensis* (Anthozoa, Cnidaria) in *Xenopus laevis*, a well-established model system for studying Brachyury (4, 7, 18). Our data show that *C. owczarzaki* Brachyury (*CoBra*) can partially rescue *Xenopus laevis* embryos injected with a dominant negative *XBra* construct. However, *CoBra*, contrary to *S. ciliatum* Bra (*SciBra*) and *N. vectensis* Bra (*NvBra*), activates target genes known to be regulated by other T-box gene classes, but not by Brachyury. We also use protein-binding microarrays to demonstrate that

Significance

The T-box transcription factors are key players in animal development and they were considered strictly animal-specific. We show that T-box genes have instead an important premetazoan evolutionary history, being present in several nonmetazoan unicellular taxa. Notably, we find that *Capsaspora owczarzaki*, a unicellular relative of animals, has a Brachyury homolog functionally conserved with metazoans. Through experiments in *Xenopus*, we demonstrate that *C. owczarzaki* Brachyury lacks, however, the target specificity displayed by metazoan Brachyury homologs, including sponges. This suggests that the subfunctionalization of T-box classes was established at the onset of Metazoa through new interactions with cofactors, concomitantly with the diversification of the T-box family.

Author contributions: A.S.-P., T.R.H., J.L.G.-S., and I.R.-T. designed research; A.S.-P., A.A.-C., M.T.W., and A.Y. performed research; S.L., G.T., M. Adamski, and M. Adamska contributed new reagents/analytic tools; A.S.-P., A.A.-C., M.T.W., T.R.H., J.L.G.-S., and I.R.-T. analyzed data; and A.S.-P., M.T.W., T.R.H., J.L.G.-S., and I.R.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]A.S.-P. and A.A.-C. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: inaki.ruiz@multicellgenome.org or jlgomska@upo.es.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1309748110/-/DCSupplemental.

the binding specificity of *CoBra* is indistinguishable from that of metazoan Brachyury and other T-box genes. Together, our data suggest that the subfunctionalization of Brachyury and other T-box classes is due to changes in interactions with cofactors, as opposed to changes in the DNA-binding recognition motif and that this subfunctionalization occurred at the origin of the Metazoa, concomitant with the diversification of the T-box family.

Results and Discussion

Genomic Survey of T-Box Genes in Nonmetazoan Species. We have searched for T-box genes in recently sequenced eukaryotic genomes and transcriptomes, including previously undescribed genomic and/or transcriptomic data from unicellular holozoans and calcarean sponges. This genomic survey has greatly extended the number of nonmetazoan taxa in which T-box genes have been identified. Our analyses reveal that T-box genes are present in at least four fungi taxa, belonging to three different early-branching fungal lineages (19): *S. punctatus* and *Gonapodya prolifera* (Chytridiomycota), *Pyromices* sp. (Neocallimastigomycota), and *Mortierella verticillata* (Mucoromycotina), all of which have a single T-box gene (Fig. 1). No T-box genes were found in higher fungi (Dikarya), in agreement with previous surveys (13, 14, 16). This confirms that T-box transcription factors were lost during fungal evolution (16). We also identified two T-box genes in the filasterean *M. vibrans*, as well as in each of the five ichthyosporeans analyzed: seven T-box genes in *Sphaeroforma arctica*, six in *Creolimax fragrantissima*, five in *Abeoforma whisleri*, two in *Amoebidium parasiticum*, and four in *Pirum gemmata*. We did not identify T-box genes in either of the two sequenced choanoflagellates (the colonial *Salpingoeca rosetta* and the unicellular *Monosiga brevicollis*), confirming that T-box genes were also lost in this group (16). No T-box genes are present outside the opisthokonts, under the current taxon sampling.

To classify the T-box genes identified, we performed a phylogenetic analysis. The resulting tree demonstrates that all fungal T-box homologs, as well as one (*C. owczarzaki*, *M. vibrans*, *A. whisleri*, and *P. gemmata*) or several (*C. fragrantissima*, *S. arctica*, and *A. parasiticum*) homologs from both filastereans and ichthyosporeans cluster at the base of the Brachyury class (Fig. S1). This result supports the notion that Brachyury is the most ancient member of the T-box family (11, 13). Moreover, fungal, and especially filasterean, Brachyury genes have most of the T-box key DNA-binding and dimerization amino acids, as well as conserved exclusive amino acid motifs of the Brachyury class (Fig. S2). In contrast, the highly divergent T-box genes from ichthyosporeans lack most of the known functional T-box domain amino acids (Fig. S2).

Our tree also deciphers a previously undescribed class of T-boxes (Tbx7), which includes the remaining (non-Brachyury) filasterean and ichthyosporean T-box genes, as well as homologs from sponges (*S. ciliatum*, *Leucosolenia complicata*, and *Amphimedon queenslandica*) (Fig. 1 and Fig. S1). Statistical support for this clade is not high, due to the short number of amino acids of the T-box domain, but the group is recovered by both maximum likelihood and Bayesian analyses. Included in the Tbx7 group there is a *C. owczarzaki* T-box gene with two T-box DNA-binding domains, a configuration not present in any reported T-box gene. This is, however, not uncommon in other eukaryotic transcription factor families. It has been hypothesized that multiple DNA-binding domains can increase the length and diversity of DNA recognition motifs recognizable by the limited number of DNA-binding domain families (25, 26). Whether this or other explanations account for the presence of this T-box gene in *C. owczarzaki* remains to be elucidated.

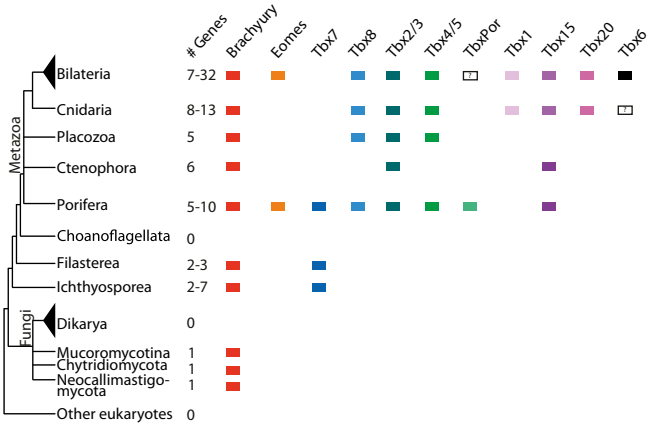

Fig. 1. Phylogenetic distribution of different T-box classes among opisthokonts. The first column indicates the minimum and maximum number of T-box genes found in each lineage. Consensus phylogenetic relationships are shown (17, 19–24). See also Fig. S1 and Dataset S1.

A Revised Evolutionary History of Metazoan T-Box Classes. Previous studies have identified T-box genes in nonbilaterian metazoans (11–14). This knowledge has enabled a reconstruction of the Urmetazoan T-box complement, which putatively included three classes (Tbx4/5, Brachyury, and a putative Tbx1/15/20), with other classes being added in a stepwise manner through the evolution of metazoans. Thus, Tbx2/3, Tbx1, Tbx15, and Tbx20 originated within eumetazoans (Cnidaria + Bilateria), whereas Tbx6 and Eomes classes originated within bilaterians. Our phylogenetic analysis, which includes previously undescribed data not only from several fungi and unicellular relatives of Metazoa, but also from two calcarean sponges, allows us to reevaluate the evolutionary history of the T-box family. Our data show that sponges, potentially the earliest-branching Metazoa (20, 21), have a much more complex complement of T-box genes than previously thought (Fig. 1). Both the homoscleromorph sponge *Oscarella carmela* and the ctenophore *Mnemiopsis leidyi* have a Tbx1/15/20 homolog, which shares with all other Tbx1/15/20 members an exclusive amino acid insertion (Fig. S2). The presumed Tbx1/15/20 homologs identified in the demosponges *A. queenslandica* and *Axinella verrucosa* (13) were previously thought to comprise a new demosponge-specific T-box class (14). We also recover this group, but surprisingly it also includes a sequence from the deuterostome *Saccoglossus kowalevski*. We have preserved the nomenclature TbxPor, following ref. 14.

We further identified a Tbx2/3 class member in the sponge *O. carmela*, as well as in the ctenophores *M. leidyi* and *Pleurobrachia pileus*. This suggests that the Tbx2/3 class was already present at the origin of animals. In agreement with previous results, we identified Tbx4/5 in most early-branching metazoans, except in ctenophores. Besides the Tbx7 (see above), we also define the group Tbx8, which to date includes only sponges (demosponges and *O. carmela*), *Trichoplax adharens*, *N. vectensis* and two bilaterians (*S. kowalevskii* and *Lottia gigantea*). Both groups appear to have been lost in some lineages during metazoan evolution. As in previous studies (7, 14), our data do not support the monophyly of Tbx6 class, but no putative orthologs were identified in early-branching metazoans. Thus, this class likely evolved later during metazoan evolution. Further, in contrast to previous reports that considered the Eomes class as a bilaterian innovation (13), we could identify homologs in the calcarean sponges *L. complicata* and *S. ciliatum*.

Finally, our results demonstrate that Brachyury is the most widely distributed class of T-box genes, with members present in all major clades: sponges (Calcarea, Demospongia, Homoscleromorpha, and Hexactinellida), ctenophores, placozoans, cnidarians, all analyzed bilaterians, and all nonmetazoan taxa with T-box family members. This suggests that Brachyury was the ancestral T-box gene from which all other classes evolved (Fig. 1). Further, at least two classes of T-box were already present at the origin of the Holozoa (Bra and Tbx7), and the Urmetazoan

T-box complement was therefore much larger than previously thought (Bra, Eomes, Tbx2/3, TbxPor, Tbx4/5, Tbx1/15/20, Tbx7, and Tbx8), suggesting a sudden diversification of T-box classes at the onset of Metazoa. Overall, our data show that T-box is an ancient transcription factor of opisthokont origin, with members present in several species belonging to five different nonmetazoan lineages (Filasterea, Ichthyosporea, and the early-branching fungi Neocallimastigomycota, Chytridiomycota, and Mucoromycotina). Evolutionarily, the T-box family is highly dynamic, with multiple secondary losses along evolution (with the exception of Brachyury, which is conserved in many lineages, but lost, for example, in *C. elegans* (8) and *A. queenslandica*), some fast-evolving members (for example, in sponges and ichthyosporeans), expansions (such as three paralogous eumetazoan classes related to the ancestral Tbx1/15/20), and major rearrangements, such as the double T-box domain found in *C. owczarzaki* (16).

Functional Conservation of *C. owczarzaki* and *S. ciliatum* Brachyury. Given its univocal phylogenetic position and the high degree of conservation at the amino acid level of *C. owczarzaki* Brachyury (Figs. S1 and S2), we decided to test its functional conservation within a metazoan context. We also included another *C. owczarzaki* T-box in our analyses (*CoTbox3*, a member of the Tbx7 class), as a non-Brachyury gene control from the same taxon. We used *Xenopus* as a model system, as it has previously been used to characterize T-box genes from early-branching metazoans (4, 7, 18).

Xenopus embryos injected with an mRNA encoding a dominant negative form of Brachyury (*XBra*_En) show defective gastrulation and impairment of muscle development (27). This phenotype is partially rescued by coinjection of *XBra* mRNA. We used embryos injected with *XBra*_En mRNA to compare the rescue capacity of *XBra*, *CoBra*, and *C. owczarzaki* Tbox3 (*CoTbx3*) mRNAs (Fig. 2). Surprisingly, both *C. owczarzaki* genes rendered a proportion of rescued embryos largely similar to those observed in embryos injected with the endogenous *XBra* (Fig. 2*A*), as determined by the general shape of the injected embryos and by in situ hybridization for the muscle gene *MyoD*. These results were confirmed by quantitative RT-PCR experiments that further demonstrated that different mesodermal-derived genes could be rescued by these coinjections (Fig. 2*B*). However, the possibility exists that the *XBra*_En construct unspecifically affects other T-box genes (because they have similar DNA-binding motifs; see below). Thus, these results should be interpreted with caution; we can only conclude that both *CoBra* and *CoTbx3* can roughly mimic endogenous *XBra* function.

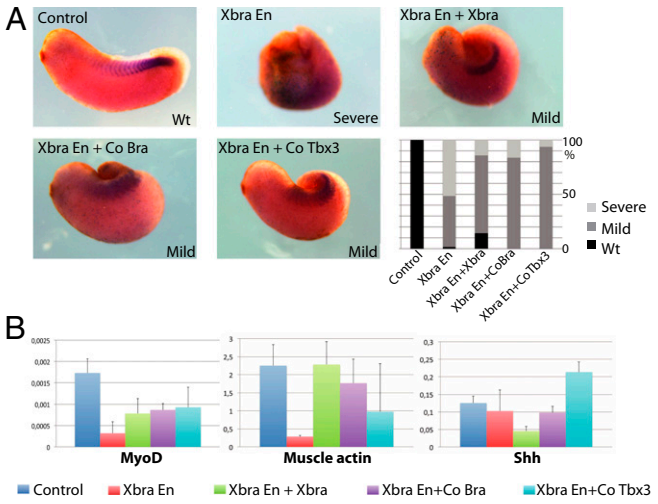

Fig. 2. *C. owczarzaki* Brachyury (*CoBra*) and Tbox-3 (*CoTbx3*) mRNAs rescue *XBra*_En injections in *Xenopus* assays. (*A*) All panels show *MyoD* expression in stage-30 embryos injected with 500 pg of the mRNA or the combinations of mRNAs indicated in the *Upper Left*. The phenotypes were classified in three categories based on the amount of trunk structures observed in the injected embryos as determined by muscle *MyoD* expression. Wild-type embryos showed complete trunk and full *MyoD* expression. Mild affected embryos showed partial reduction of the trunk with reduced *MyoD* expression domain. Severe affected embryos lacked almost all trunk tissue, and the expression of *MyoD* was hardly or not detected. *Lower Right* shows bar plots summarizing the different phenotypes observed in each case. (*B*) Quantitative RT-PCR experiments showing two muscle (*MyoD* and *Muscle actin*) and one notochord (*Shh*) genes in the different injected embryos. Controls are noninjected embryos. Bar plots represent relative expression, normalized with endogenous Histone 4 expression levels. Error bars represent SD from at least two different biological replicates.

For this reason, we next evaluated if this similar rescue potential is the consequence of the capacity of these genes to activate similar downstream target genes. It has been shown that not all T-box genes activate the same target genes. For example, in animal caps assays, Tbx6 (*VegT*) strongly activates a broad panel of mesendodermal genes such as *Wnt11*, *Wnt8*, *endodermin, Sox17, chordin*, and *pintallavis* while Brachyury strongly activates *Wnt11* and weakly activates *Sox17*, while it never activates *chordin* (7, 28, 29). This difference seems to be due to the ability of Brachyury to interact with the cofactor Smad1. This interaction, which takes place through an N-terminal domain of the Brachyury protein, allows the activation of *Xom*, a repressor of dorsal mesendodermal genes (30, 31). We therefore compared the ability of different T-box genes to activate these target genes in *Xenopus* overexpression assays. For comparison, we also included the T-box gene of the fungus *S. punctatus* (*SpBra*), the Bra gene of the cnidarian *N. vectensis* (*NvBra*) and the two Bra paralogs of the calcarean sponge *S. ciliatum* (*SciBra*1 and *SciBra*2).

Fig. 3 shows the molecular phenotypes obtained in animal caps (Fig. 3 *A* and *B*) and whole embryos (Fig. 3*C* and Fig. S3) injected with different Brachyury mRNA homologs. Metazoan homologs (*NvBra*, *SciBra*1, and *SciBra*2) showed largely similar molecular phenotypes to that previously reported for *Xenopus* Bra (7, 28), as they activated *Wnt11* but not *chordin*. However, sponge homologs, unlike cnidarian and *Xenopus* Brachyury mRNAs, were unable to activate *Sox17* and *endodermin*. In addition, *NvBra* and *SciBra*1 activated *pintallavis* and *wnt8*, respectively, indicating some difference in the regulatory potential among metazoan Brachyury homologs. Despite their ability to rescue the loss of *XBra* function (see above), a different molecular phenotype was observed in embryos injected with *C. owczarzaki* Brachyury (*CoBra*) and Tbx7 (*CoTbx3*) mRNAs, which strongly activated all mesendodermal genes. This suggests a clear boundary between metazoan and nonmetazoan Brachyury homologs, which may be explained by the ability of the metazoan Brachyury orthologs to interact with cofactors that restrict their function, such as Smad1. Interestingly, this factor is present in the genome of *S. ciliatum*, but not in the genome of *C. owczarzaki* (16). Finally, we obtained discrepant results in the different assays with the fungus homolog (*SpBra*), with no clear patterns of gene activation, neither strong activation levels. This might be explained by the fact that the fungal T-box genes, including *SpBra*, are very divergent (Fig. S2).

Several amino acid motifs have been suggested to be key determinants of the specificity of Brachyury, compared with other T-box family members. We therefore asked whether any of these motifs could account for the differences observed between metazoan and nonmetazoan homologs. Ref. 28 proposed that the presence of a Lysine in position 149 of *XBra* accounts for its differential behavior, compared with other T-box classes such as Tbx6 (*VegT*) and Eomes, which instead have an Asparagine at this position. Our alignments (Fig. S2) indicate that this position is indeed conserved in the *N. vectensis* and *S. ciliatum* Brachyury proteins. However, despite the presence of an Arginine (R) instead of a Lysine (K) in the *CoBra* protein, we do not believe that this difference alone can explain the drastic phenotypic differences we observed between metazoan Brachyury and *CoBra*,

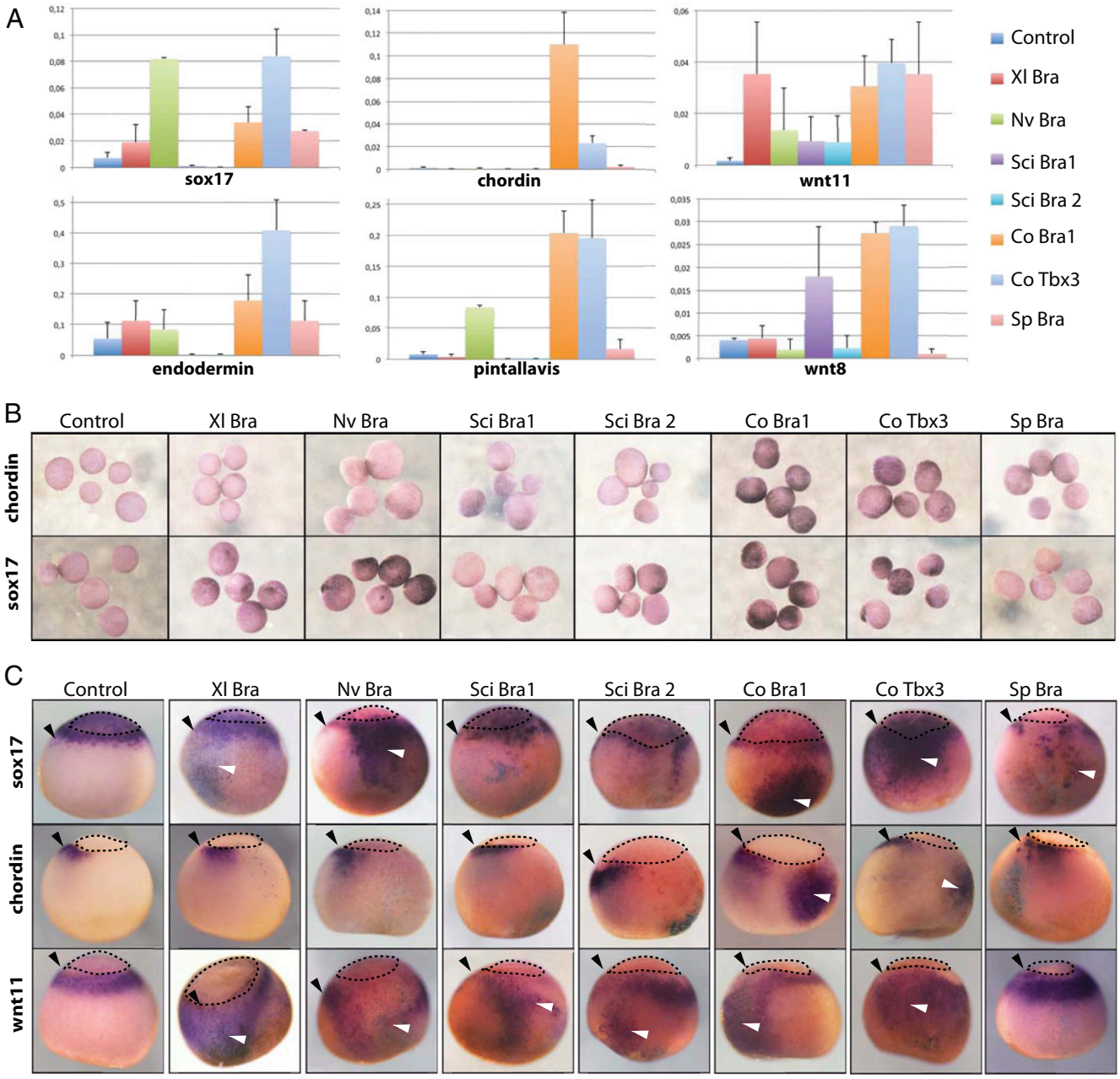

Fig. 3. Molecular phenotype of *Xenopus* animal caps or embryos injected with different Brachyury orthologs mRNAs. (*A*) Quantitative RT-PCR experiments showing several mesendoderm markers in animal caps injected with different mRNAs. Controls are noninjected embryos. Bar plots show relative expression, normalized with endogenous Histone 4. Error bars represent SD from at least two different biological replicates. (*B*) In situ hybridization to detect *chordin* and *Sox17* expression in animal caps injected with the Brachyury mRNAs indicated above each panel. (*C*) Whole-mount in situ hybridization of *Sox17*, *chordin*, and *Wnt11* genes in stage 11–12 *Xenopus* embryos injected with the Brachyury orthologs indicated above each panel. All embryos are shown in the same orientation. Dotted lines represent the closing blastopore; the black arrowhead indicates the dorsal side, and the white arrowhead highlights ectopic expression.

especially when considering that R is a hydrophilic basic amino acid, being extremely similar to Lysine, and it is very different from the neutral Asn (N) found in all other T-box classes. In fact, *SciBra2* also has an R in this position and, nonetheless, it does not activate *chordin* (Fig. 3). Ref. 30 proposed that an N-terminal domain is responsible for the interaction between Brachyury and Smad1, which would restrict its function spatially. Indeed organisms whose Brachyury lack this domain, such as *Drosophila* Bra or ascidian Bra, are unable to behave as endogenous *XBra* (31). However, the ctenophore *M. leidyi*, the sponge *S. ciliatum*, and the cnidarian *N. vectensis* Bra homologs lack the conserved N-terminal region and, nonetheless, behave similarly to *XBra*. Finally, ref. 18 proposed that a conserved motif in the C-terminal activation domain (called the R1 domain), which is present in Bilateria and Cnidaria, is responsible for Brachyury specificity. However, this domain is again not present in *M. leidyi* or *S. ciliatum* Brachyury homologs. Thus, our data, together with the results from ref. 7, suggest that the difference between metazoan and nonmetazoan Brachyury homologs in their ability to mimic endogenous *XBra* functions cannot be explained by the presence of any of these specific amino acidic motifs outside of the T domain. To gain further insights into this question, we constructed six different *XBra-CoBra* gene chimeras, combining N-terminal domain, T-box, and C-terminal domain from each

taxa (Fig. S4). While most chimeras largely behaved as *XBra*, the *CoBra-XBra-CoBra* fusion activated the same panel of downstream genes as *CoBra* (Fig. 3), although *chordin* up-regulation was limited. Therefore, N-terminal and C-terminal domains, even though they do not contain any recognizable conserved amino acidic motifs, could largely account for the metazoan Brachyury homologs specificity (including those from sponges and ctenophores). This observation is in agreement with the protein-binding microarrays (PBM) results (see below), which show no differences in DNA-binding motif specificities between *CoBra* and metazoan Brachyury homologs.

***C. owczarzaki* Brachyury Has a Conserved T-Box DNA-Binding Motif.** To further investigate the function of *C. owczarzaki* Brachyury, we determined its binding preferences using universal PBM (32, 33). Our results indicate that *CoBra* has a highly similar motif to that determined in the mouse Bra-homolog, called T (Fig. 4) (28, 34–36). Moreover, our results indicate that the T-box DNA recognition sequence is strongly conserved, both across a wide range of T-box classes [including Eomes (37), Tbx1, Tbx4, and Tbx2] and also across different organisms (Fig. 4 and Fig. S1). Thus, our data from the protist *C. owczarzaki* suggest that T-box genes have preserved a DNA recognition motif that has undergone very little change during evolutionary time, even with the diversification of the family at the origin of Metazoa. These results suggest that cooperative interactions of T-box genes with different cofactors, as opposed to differences in DNA-binding sequence recognition, are the key means through which members of this family have diverged in function. Similar findings have been reported, for example, for Hox family transcription factors (38). Moreover, it is likely that regulation of temporal expression could contribute to differences in function. The conserved binding motif also helps to explain the ability of *CoBra* to rescue endogenous *XBra* and to activate several downstream T-box targets in *Xenopus*, but without the specificity of *XBra*, probably due to the inability of *CoBra* to interact with cofactors. In sharp contrast, the Brachyury orthologs of the early-branching metazoan *S. ciliatum* can perfectly mimic the behavior of endogenous *XBra*.

Conclusions

Our data, which include several previously unreported T-box genes from sponges, fungi, ichthyosporeans, and filastereans, allow us to reconstruct T-box transcription factor family evolution with unprecedented detail. We have also analyzed the most conserved nonmetazoan Brachyury homolog known to date, that of the filose amoeba *C. owczarzaki*, a close relative of Metazoa (17, 39).

Our results demonstrate that the repertoire of T-box transcription factors in premetazoans is much richer than previously thought, with members of this family present in several fungi, ichthyosporeans, and filastereans. T-box genes evolved in the last common ancestor of all opisthokonts (Fig. 1) and were secondarily lost in higher fungi (Dikarya) and in choanoflagellates. Phylogenetic analyses and molecular signatures confirm that Brachyury is the most ancient member of the T-box genes. A new member of the family (Tbox7) evolved later, within the Holozoa clade, with members present in some ichthyosporeans, filastereans, and sponges, but was secondarily lost in other metazoans. The T-box family radiated at the origin of Metazoa in a highly dynamic scenario with some fast-evolving classes (such as Tbx7 and Tbx8) and some classes that have been secondarily lost (such as Tbx7 and TbxPor, which are only present in sponges). After this initial period, the number of classes stabilized until the emergence of Tbx1, Tbx15, and Tbx20 from a common Tbx1/15/20 ancestor at the stem of Cnidaria + Bilateria and the origin of Tbx6 at the stem of Bilateria.

The subfunctionalization of Brachyury seems to have been well-established at the very origin of the Metazoa. However, the high number of T-box classes (including some, like Tbx7 or TbxPor, that were subsequently lost), the uneven distribution of T-box classes in sponges, and the presence of fast-evolving T-box genes in sponges suggest an early scenario of fast evolution of new T-box classes.

Results of our analyses indicate that the binding specificity of Brachyury is highly conserved among metazoan and nonmetazoans, as well as between Brachyury and other T-box classes. This reinforces the idea that cofactor interactions may be responsible for the functional differences observed between different T-box classes and may also explain why the Brachyury of *C. owczarzaki*, although clearly a Brachyury ortholog, does not have the ability to interact with cofactors in a *Xenopus* heterologous context, in sharp contrast to the Brachyury of sponges, ctenophores, or cnidarians. Most likely, these restrictions were set at the origin of Metazoa with the radiation of T-box classes, as evidenced by the perfect functional mimic of *SciBra* and *M. leidyi* Bra (7) with *XBra*. In that sense, both *CoBra* and *CoTbx3* (a member of the Tbx7 class) behave as what we call "pan-Tbox" genes, activating all potential targets (like *chordin*) that will later in evolution be controlled by specific T-box classes (in the case of *chordin*, Tbx6 and Eomes). Through time, novel T-box specificities were established through the evolution of new functional interactions with different cofactors.

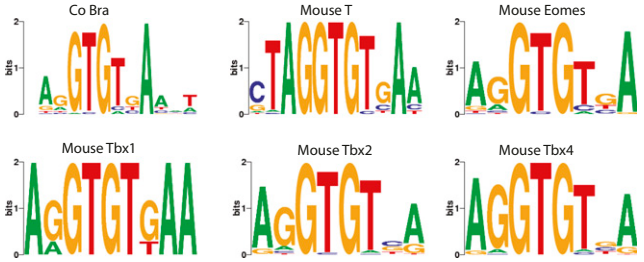

Fig. 4. *CoBra*-binding motifs derived from PBM data (*SI Methods*). For comparison, different mouse T-box classes binding motifs also derived from PBM data [except mouse T, based on SELEX (Systematic Evolution of Ligands by Exponential Enrichment) (34)]. See also Dataset S2.

Methods

Microinjection of Brachyuru Genes into *Xenopus* Embryos. The entire coding regions of Brachyury genes from different species (*CoBra*, *CoTbox3*, *SciBra*1, *SciBra*2, *NvBra*, and *SpBra*) as well as six different *CoBra-XlBra* chimeras (representing all possible N-terminal domain, T-box, and C-terminal domain combinations) were inserted into the multicloning site of pCS2+ (40). mRNAs, prepared as previously described (41), were injected in *Xenopus* embryos at two-four-cell stage in a single blastomere at 500–1,000 pg per embryo. X-Gal staining was performed as described elsewhere (42).

Histochemistry. *Xenopus* embryos were fixed in MEMFA (0.1 M 3-(N-Morpholino) propanesulfonic acid (Mops), 2 mM EGTA, 1 mM $MgSO_4$, and 3.7% (vol/vol) formaldehyde, pH7.4) for 1h at room temperature and then kept in methanol at −20 °C. Antisense RNA probes were prepared from cordin, *Wnt11*, and *Sox17β* cDNAs using dioxigenin (Roche). *Xenopus* embryos were hybridized as described (43). After immunostaining, embryos were bleached by the treatment with 10% H_2O_2 in PBS under the light for 2–3 h.

Animal Caps Assays and Quantitative RT-PCR. Animal caps were prepared as previously described (44) and cultured until stage 13 according to Nieuwkoop and Faber (45). For qRT-PCR, total RNA from 10 animals caps or five stage-30 embryos was isolated with TRIzol (Gibbco) followed by phenol/chloroform extraction. cDNA synthesis was performed using SuperScript III kit (Invitrogen) according to the manufacturer's protocol. SYBR-Green real-time PCR was performed on CFX96 BioRad Detection System using iTaq Universal SYBR-Green Supermix (BioRad). All reactions were done in duplicates in at least two different biological replicas. The expression of histone 4 (H4) was used as a control. For a list of primers used for these experiments, see Table S1.

Gene Searches and Phylogenetic Analysis. A primary search was performed using the basic local alignment sequence tool (BLAST: BlastP and TBlastN) using *Homo sapiens* and *C. owczarzaki* proteins as queries against protein, genome, and transcriptome databases with the default BLAST parameters

and an e-value threshold of e-5 at the National Center for Biotechnology Information (NCBI) and against completed or ongoing genome project databases at the Joint Genome Institute (JGI) (for *Piromyces* sp., *Gonapodya prolifera*, and other early-branching fungi available) and the Broad Institute (for *M. verticillata*, *S. rosetta*, *S. arctica*, and *S. punctatus*). For *A. whisleri*, *P. gemmata*, *A. parasiticum*, and *M. vibrans* we assembled the trace RNAseq data using the Trinity assembler. *C. fragrantissima* genomic sequences were assembled using the whole-genome shotgun (WGS) assembler (http://sourceforge.net/apps/mediawiki/wgs-assembler/index.php?title5Main_Page). In both cases, we performed local BLAST searches and annotated the sequences manually. We also performed profile hidden Markov model searches using HMMER3.0b2 (46) to confirm that we were retrieving all T-box orthologs. For details on phylogenetic analyses, see *SI Methods*.

Protein-Binding Microarrays. Details of the design and use of universal PBMs has been described elsewhere (32, 33). For further details, see *SI Methods*.

ACKNOWLEDGMENTS. We thank Ignacio Maeso, Alex de Mendoza, Hiroshi Suga, and other members of the multicellgenome laboratory for useful insights. We thank Ana Gilles and Ulrich Technau (University of Vienna) for providing the *N. vectensis* Brachyury clone. We also thank the Joint Genome Institute and the Broad Institute for making data publicly available. This work was supported by an Institució Catalana de Recerca i Estudis Avançats contract, European Research Council Starting Grant ERC-2007-StG-206883, and Grant BFU2011-23434 from the Ministerio de Economía y Competitividad (MINECO) (to I.R.-T.). A.S.-P. was supported by a pregraduate Formación Profesorado Universitario grant from MINECO and a grant from MINECO to perform a research stay at the J.L.G.-S. laboratory. J.L.G.-S. thanks the Spanish and Andalusian Governments for Grants BFU2010-14839, CSD2007-00008, and Proyecto de Excelencia CVI-3488 that funded this study. This work was also supported by a Canadian Institutes of Health Research (CIHR) grant [MOP-111007 (to T.R.H.)]. M. Adamski, M. Adamska, and S.L. acknowledge funding from the core budget of the Sars International Centre for Marine Molecular Biology. M.T.W. was supported by fellowships from CIHR and the Canadian Institute for Advanced Research Junior Fellows Genetic Networks Program.

1. Papaioannou VE, Silver LM (1998) The T-box gene family. *Bioessays* 20(1):9–19.
2. Smith J (1999) T-box genes: What they do and how they do it. *Trends Genet* 15(4): 154–158.
3. Wilson V, Conlon FL (2002) The T-box family. *Genome Biol* 3(6):3008.1–3008.7.
4. Marcellini S, Technau U, Smith JC, Lemaire P (2003) Evolution of Brachyury proteins: Identification of a novel regulatory domain conserved within Bilateria. *Dev Biol* 260(2):352–361.
5. Scholz CB, Technau U (2003) The ancestral role of Brachyury: Expression of NemBra1 in the basal cnidarian *Nematostella vectensis* (Anthozoa). *Dev Genes Evol* 212(12): 563–570.
6. Showell C, Binder O, Conlon FL (2004) T-box genes in early embryogenesis. *Dev Dyn* 229(1):201–218.
7. Yamada A, Martindale MQ, Fukui A, Tochinai S (2010) Highly conserved functions of the Brachyury gene on morphogenetic movements: Insight from the early-diverging phylum Ctenophora. *Dev Biol* 339(1):212–222.
8. Papaioannou VE (2001) T-box genes in development: From hydra to humans. *Int Rev Cytol* 207:1–70.
9. Technau U (2001) Brachyury, the blastopore and the evolution of the mesoderm. *Bioessays* 23(9):788–794.
10. Martinelli C, Spring J (2005) T-box and homeobox genes from the ctenophore Pleurobrachia pileus: Comparison of Brachyury, Tbx2/3 and Tlx in basal metazoans and bilaterians. *FEBS Lett* 579(22):5024–5028.
11. Adell T, Grebenjuk VA, Wiens M, Müller WE (2003) Isolation and characterization of two T-box genes from sponges, the phylogenetically oldest metazoan taxon. *Dev Genes Evol* 213(9):421–434.
12. Manuel M, Le Parco Y, Borchiellini C (2004) Comparative analysis of Brachyury T-domains, with the characterization of two new sponge sequences, from a hexactinellid and a calcisponge. *Gene* 340(2):291–301.
13. Larroux C, et al. (2008) Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol* 25(5):980–996.
14. Holstien K, et al. (2010) Expansion, diversification, and expression of T-box family genes in Porifera. *Dev Genes Evol* 220(9-10):251–262.
15. King N, et al. (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451(7180):783–788.
16. Sebé-Pedrós A, de Mendoza A, Lang BF, Degnan BM, Ruiz-Trillo I (2011) Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Mol Biol Evol* 28(3):1241–1254.
17. Torruella G, et al. (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol Biol Evol* 29(2): 531–544.
18. Bielen H, et al. (2007) Divergent functions of two ancient Hydra Brachyury paralogues suggest specific roles for their C-terminal domains in tissue fate induction. *Development* 134(23):4187–4197.
19. Grigoriev I, Cullen D, Goodwin S, Hibbett D (2011) Fuelling the future with fungal genomics. *Mycology* 2:192–209.
20. Philippe H, et al. (2009) Phylogenomics revives traditional views on deep animal relationships. *Curr Biol* 19(8):706–712.
21. Pick KS, et al. (2010) Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol Biol Evol* 27(9):1983–1987.
22. Shalchian-Tabrizi K, et al. (2008) Multigene phylogeny of choanozoa and the origin of animals. *PLoS ONE* 3(5):e2098.
23. Dunn CW, et al. (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452(7188):745–749.
24. Hejnol A, et al. (2009) Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Biol Sci* 276(1677):4261–4270.
25. Charoensawan V, Wilson D, Teichmann SA (2010) Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res* 38(21): 7364–7377.
26. Itzkovitz S, Tlusty T, Alon U (2006) Coding limits on the number of transcription factors. *BMC Genomics* 7:239.
27. Conlon FL, Sedgwick SG, Weston KM, Smith JC (1996) Inhibition of *Xbra* transcription activation causes defects in mesodermal patterning and reveals autoregulation of *Xbra* in dorsal mesoderm. *Development* 122(8):2427–2435.
28. Conlon FL, Fairclough L, Price BM, Casey ES, Smith JC (2001) Determinants of T box protein specificity. *Development* 128(19):3749–3758.
29. Xanthos JB, Kofron M, Wylie C, Heasman J (2001) Maternal *VegT* is the initiator of a molecular network specifying endoderm in *Xenopus laevis*. *Development* 128(2): 167–180.
30. Messenger NJ, et al. (2005) Functional specificity of the *Xenopus* T-domain protein Brachyury is conferred by its ability to interact with Smad1. *Dev Cell* 8(4):599–610.
31. Marcellini S (2006) When Brachyury meets Smad1: The evolution of bilateral symmetry during gastrulation. *Bioessays* 28(4):413–420.
32. Berger MF, et al. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* 24(11): 1429–1435.
33. Berger MF, Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc* 4(3):393–411.
34. Kispert A, Herrmann BG (1993) The Brachyury gene encodes a novel DNA binding protein. *EMBO J* 12(8):3211–3220.
35. Casey ES, O'Reilly MA, Conlon FL, Smith JC (1998) The T-box transcription factor Brachyury regulates expression of eFGF through binding to a non-palindromic response element. *Development* 125(19):3887–3894.
36. Garnett AT, et al. (2009) Identification of direct T-box target genes in the developing zebrafish mesoderm. *Development* 136(5):749–760.
37. Badis G, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324(5935):1720–1723.
38. Slattery M, et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147(6):1270–1282.
39. Ruiz-Trillo I, et al. (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. *Curr Biol* 14(22):R946–R947.
40. Turner DL, Weintraub H (1994) Expression of achaete-scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev* 8(12):1434–1447.
41. Harland R, Weintraub H (1985) Translation of mRNA injected into *Xenopus* oocytes is specifically inhibited by antisense RNA. *J Cell Biol* 101(3):1094–1099.
42. Coffman CR, Skoglund P, Harris WA, Kintner CR (1993) Expression of an extracellular deletion of Xotch diverts cell fate in *Xenopus* embryos. *Cell* 73(4):659–671.
43. Jones CM, Smith JC (1999) Wholemount in situ hybridization to *Xenopus* embryos. *Methods Mol Biol* 97:635–640.
44. Gómez-Skarmeta JL, Glavic A, de la Calle-Mustienes E, Modolell J, Mayor R (1998) Xiro, a *Xenopus* homolog of the Drosophila Iroquois complex genes, controls development at the neural plate. *EMBO J* 17(1):181–190.
45. Nieuwkoop PD, Faber J (1975) *Normal Table of Xenopus laevis (Daudin)* (Elsevier/North-Holland Publishing, Amsterdam).
46. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14(9):755–763.