View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Developmental Biology 273 (2004) 436–453

DEVELOPMENTAL BIOLOGY

www.elsevier.com/locate/ydbio

Genomes & Developmental Control

Creation of *cis*-regulatory elements during sea urchin evolution by co-option and optimization of a repetitive sequence adjacent to the *spec2a* gene

Sandeep Dayal[a,b], Takae Kiyama[a], Jeffrey T. Villinski[a], Ning Zhang[a], Shuguang Liang[a], William H. Klein[a,b,*]

[a]*Department of Biochemistry and Molecular Biology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA*
[b]*Graduate Training Program in Genes and Development, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, TX 77225, USA*

Received for publication 20 May 2004, accepted 4 June 2004
Available online 31 July 2004

Abstract

The creation, preservation, and degeneration of *cis*-regulatory elements controlling developmental gene expression are fundamental genome-level evolutionary processes about which little is known. Here, we identify critical differences in *cis*-regulatory elements controlling the expression of the sea urchin aboral ectoderm-specific *spec* genes. We found multiple copies of a repetitive sequence element termed RSR in genomes of species within the Strongylocentrotidae family, but RSRs were not detected in genomes of species outside Strongylocentrotidae. *spec* genes in *Strongylocentrotus purpuratus* are invariably associated with RSRs, and the *spec2a* RSR functioned as a transcriptional enhancer and displayed greater activity than did *spec1* or *spec2c* RSRs. Single-base pair differences at two *cis*-regulatory elements within the *spec2a* RSR increased the binding affinities of four transcription factors, SpCCAAT-binding factor at one element and SpOtx, SpGoosecoid, and SpGATA-E at another. The *cis*-regulatory elements to which these four factors bound were recent evolutionary acquisitions that acted to either activate or repress transcription, depending on the cell type. These elements were found in the *spec2a* RSR ortholog in *Strongylocentrotus pallidus* but not in RSR orthologs of *Strongylocentrotus droebachiensis* or *Hemicentrotus pulcherrimus*. Our results indicated that a dynamic pattern of *cis*-regulatory element evolution exists for *spec* genes despite their conserved aboral ectoderm expression.
© 2004 Elsevier Inc. All rights reserved.

Keywords: *cis*-regulatory element evolution; Stabilizing selection; Sea urchin evolution; Sea urchin development; Sea urchin genomics; *spec* gene; Otx; Goosecoid; CBF; GATA

Introduction

The question of how transcriptional enhancers have evolved in bilaterian genomes has attracted much recent attention mainly because changes in enhancer sequence might alter expression patterns for developmentally expressed genes (Carroll et al., 2001; Davidson, 2001). An equally important but largely unexplored corollary issue is the extent to which *cis*-regulatory elements within an enhancer can change when gene expression patterns are constrained. Orthologous genes from related species can have identical expression patterns yet their corresponding enhancers can differ substantially in *cis*-regulatory architecture (for examples, see Galis et al., 2002; Ludwig et al., 2000; Romano and Wray, 2003; Scemama et al., 2002; Takahashi et al., 1999). The differences in enhancer sequence are thought to occur by rapid fixation of functionally compensating mutations, a process called stabilizing selection (Ludwig et al., 2000; Ohta, 2003).

* Corresponding author. Department of Biochemistry and Molecular Biology, Unit 117, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Fax: +1 713 563 2968.
E-mail address: wklein@mdanderson.org (W.H. Klein).

0012-1606/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2004.06.011

The dynamic changes within the *even-skipped* stripe 2 enhancer in *Drosophila* species appear to represent an example of stabilizing selection. Cross-species comparisons reveal that multiple transcription factor-binding sites have been either lost or acquired at different positions within the *even-skipped* stripe 2 enhancer, but the net output of the enhancer is stable (Ludwig et al., 2000). This and other examples demonstrate that sequence changes within an enhancer are possible if compensatory changes maintain the constrained expression pattern of the gene. Clearly, there is flexibility for change within enhancer sequences, but the degree to which entire regulatory regions can change is open to question. We show here that enhancer evolution is not limited to small, gradual changes but can also involve large-scale alterations in *cis*-regulatory domains.

The sea urchin *spec* genes provide a novel opportunity for investigating the capacity for *cis*-regulatory element evolution and stabilizing selection. The *spec* genes encode a family of EF-hand-containing intracellular calcium-binding proteins whose expression is highly restricted to the embryonic aboral ectoderm of all sea urchin species examined to date (reviewed in Brandhorst and Klein, 1992). The sequences of the transcriptional regulatory regions controlling the expression of *spec* genes in *Lytechinus pictus* are dissimilar from those of their counterparts in *Strongylocentrotus purpuratus* (Xiang et al., 1991). These two species diverged from a common ancestor 35–50 million years ago (mya) (Lee, 2003; Littlewood and Smith, 1995). In the *S. purpuratus* genome, a repetitive sequence element termed repeat-spacer-repeat (RSR) is invariably associated with the 5′ upstream regions of *spec* genes (Gan et al., 1990b). The *spec* RSRs act as transcriptional enhancers that contribute various levels of transcriptional activity to each *spec* gene; when RSRs are deleted from *spec1* and *spec2c*, reporter gene expression is reduced 3- to 7-fold, whereas when the *spec2a* RSR is removed, there is a more than 80-fold loss of activity (Gan et al., 1990b). In addition, spatially regulated expression appears to require more than the RSRs in the case of *spec1* and *spec2c*, yet for *spec2a*, the RSR largely suffices (Gan et al., 1990a). These considerations have led us to posit that in *S. purpuratus*, *spec1* and *spec2c* have additional *cis*-regulatory elements controlling aboral ectoderm expression that lie outside of their RSR sequences (Gan et al., 1990a).

Because the *spec2a* RSR functions as a potent aboral ectoderm enhancer, we have devoted considerable effort toward identifying the critical *cis*-regulatory elements and the corresponding *trans* factors responsible for spatiotemporal expression. The chief transcriptional activator is SpOtx, a pair-rule K50 homeodomain-containing transcription factor that is expressed in all cells of the embryo at the time when *spec* genes are activated (late cleavage stage) (Gan and Klein, 1993; Gan et al., 1995; Li et al., 1997; Mao et al., 1994, 1996). Within a 78-bp fragment in the S region of the *spec2a* RSR enhancer are five tightly clustered *cis*-regulatory elements that bind to six different transcription factors (Mao et al., 1994; Yuh et al., 2001). Proximal (TAATCT) and distal (TAATCC) Otx/Goosecoid (Gsc) elements act additively as positive regulatory elements in the aboral ectoderm and endoderm by binding to SpOtx. These same elements act negatively in the oral ectoderm by binding to the pair-rule K50 homeodomain-containing transcriptional repressor SpGsc (Angerer et al., 2001). A CCAATT (CBF) element (positioned in reverse orientation as AATTGG) acts as a positive ectoderm element that binds to the *S. purpuratus* CCAAT-binding factor ortholog, SpCBF (Li et al., 2002; Yuh et al., 2001). An oral ectoderm repressor (OER) element (TTCACTG) and an endoderm repressor (ENR) element (not clearly defined, but overlapping the proximal Otx/Gsc element) act to repress *spec2a* in the oral ectoderm and endoderm, respectively (Yuh et al., 2001). We have recent preliminary evidence that the proximal but not the distal Otx/Gsc element binds to SpGATA-E, an *S. purpuratus* GATA 4/5 ortholog, which is expressed specifically in the endomesoderm and is essential for endoderm differentiation (Davidson et al., 2002). However, it is not clear yet whether the ENR, whose target site overlaps the proximal Otx/Gsc element, is in fact SpGATA-E. In summary, the *spec2a* RSR enhancer uses SpOtx, and to a lesser extent SpCBF, to effect expression in all cells of the late cleavage stage embryo. Spatial expression is conferred mainly by repression; SpGsc and OER are required for repression in the oral ectoderm, and ENR (possibly SpGATA-E) is required for repression in the endoderm. Regions distal to the RSR appear to be responsible for efficient mesenchymal cell repression, but *cis*-regulatory elements within this region remain uncharacterized.

The *cis*-regulatory mechanisms responsible for aboral ectoderm expression of *spec2a* in *S. purpuratus* appear to be substantially different from those operating on the *spec* genes in *L. pictus*. Moreover, in *S. purpuratus*, *spec2a* expression is entirely dependent on its RSR enhancer whereas *spec1* and *spec2c* rely on their RSRs to a much lesser extent. It is therefore likely that *cis*-regulatory elements required for *spec* gene expression have evolved in a complex, dynamic way and that new insights into enhancer evolution will be revealed by addressing the basic differences among the transcriptional control regions of the *spec* genes. In the current study, we focused on the RSRs and their relationship to *spec* gene expression. We determined the evolutionary origins of the RSR repetitive sequence element and traced the evolution of the *cis*-regulatory elements within the *spec2a* RSR enhancer. Our results were consistent with a model in which the RSR elements amplified in the genome of the ancestor that gave rise to the Strongylocentrotidae family, and at least a subset of these RSRs were co-opted as transcriptional enhancers. In the *S. purpuratus* clade, the *spec2a* RSR enhancer evolved two new *cis*-regulatory elements, a CBF element and an Otx/Gsc/GATA-E element. In addition to other nucleotide changes, these acquisitions optimized enhancer

function and led to an RSR-dependent mechanism for *spec2a* expression.

Materials and methods

Visualization tools for alignments (VISTA) nucleotide sequence comparison plots

spec gene 5′ -flanking sequences from *S. purpuratus* and *L. pictus* were analyzed using the VISTA Web site (at http://www.gsd.lbl.gov/vista/index.shtml). The mVISTA program was used to generate sequence comparison plots using default parameters, except the window size was set to 100 bp and the output range was set to 50–100% sequence match. In these plots, regions highlighted in pink indicate sequence identities greater than 70% (Dubchak et al., 2000; Mayor et al., 2000).

Plasmid construction and cloning

Spec RSR-luciferase constructs were generated by fusing target RSRs to the *endo16* basal promoter fragment (provided by Eric Davidson, California Institute of Technology, Pasadena, CA). The fragments were ligated into the *Nhe*I–*Bgl*II restriction enzyme sites in the multiple cloning site of the pGL3-basic firefly luciferase vector (Promega, Madison, WI). Constructs containing the 78-bp *spec2a* C fragment, SpRSR3 C fragment, and SpRSR4 R1S fragment, all of which are RSR subfragments, were generated by annealing plus- and minus-strand oligonucleotides and ligating the annealed oligonucleotides into the *endo16* basal promoter pGL3 construct. The *spec1*-mutant and *spec2a*-mutant RSR luciferase constructs were generated by creating single-point mutations in the CBF and proximal Otx/Gsc elements of the corresponding constructs using a mutagenesis kit (Quick Change Multi Site-Directed kit; Stratagene, La Jolla, CA). The correct sequences of the constructs were confirmed by DNA sequencing at The University of Texas M.D. Anderson DNA Analysis Core Facility.

Sea urchin embryo culture, microinjection, particle gun delivery, and luciferase assay

S. purpuratus and *L. pictus* were obtained from Marinus (Long Beach, CA) or Charles Hollahan (Santa Barbara, CA). Gametes were collected and prepared for injection as described previously (Mao et al., 1996). Microinjection of DNA into zygotes was performed as described previously (Gan et al., 1990b). Approximately 2000–4000 copies of *Bam*HI-linearized firefly luciferase constructs were injected per embryo in a solution containing 40% glycerol and a fivefold molar excess of *Bam*HI-digested genomic DNA. Two to four thousand copies of cytomegalovirus-Renilla luciferase plasmid (Promega) were co-injected as an internal control.

Introduction of DNA constructs into sea urchin embryos by the particle gun method was done essentially as previously described (Akasaka et al., 1995) with modifications for the PDS-1000/He Biolistic Particle Delivery System (Bio-Rad, Hercules, CA).

Luciferase activity was measured by harvesting 100–200 injected or 1×10^5 particle gun-treated embryos at the late blastula or early gastrula stage. Firefly and Renilla luciferase activities were measured sequentially using the dual-luciferase assay system (Promega), and firefly luciferase activity was normalized to that of Renilla.

Electrophoretic mobility shift assay (EMSA)

EMSAs were performed as previously described (Yuh et al., 2001). The sequences of the oligonucleotides used to generate probes are shown in Supplementary Table S1. Embryo nuclear extracts were used to detect SpOtx, SpCBF, and OER. Nuclear extracts were prepared from *S. purpuratus* 22-h blastula stage embryos, and DNA–protein complex formation was performed as previously described (Gan and Klein, 1993; Gan et al., 1995) except that the glycerol concentration of the binding buffer was reduced to 6% and in some cases a final concentration of 4% Ficoll (Fisher, Pittsburgh, PA) was added to the reaction. The glutathione S-transferase (GST)-SpGsc recombinant protein was generated as previously described (Angerer et al., 2001), and whole bacterial lysates were used as a protein source. A GST-SpGATA-E fusion construct was created by subcloning two polymerase chain reaction (PCR)-amplified SpGATA-E zinc-finger domains into the *Bam*HI–*Xho*I site of the pGEX 5T-1 vector (Pfizer, New York, NY). GST-SpGATA-E recombinant proteins were generated by expression in the Rosetta strain of *Escherichia coli* and whole bacterial lysates were used for EMSA studies. Competition assays with unlabeled oligonucleotides were performed using 5-, 10-, 250-, 500-, and 2000-fold molar excesses of competitor oligonucleotide for SpOtx; 5-, 50-, 250-, and 500-fold molar excesses for SpCBF; 1000-, 2000-, 3000-, and 4000-fold molar excesses for GST-SpGsc; and 100- and 500-fold molar excesses for SpGATA-E.

Genomic DNA isolation and gene amplification

Genomic DNA was isolated from the sperm or gonadal tissue of *L. pictus*, *S. purpuratus*, *Strongylocentrotus pallidus* (tissue provided by Christiane Biermann, University of Washington, Seattle, WA), *Strongylocentrotus droebachiensis* (animals obtained from the Marine Biological Laboratory, Woods Hole, MA), *Hemicentrotus pulcherrimus* (DNA provided by Koji Akasaka, Hiroshima University, Hiroshima, Japan), *Strongylocentrotus franciscanus* (sperm provided by Katherine Foltz, University of California at Santa Barbara, Santa Barbara, CA), and *Heliocidaris tuberculata* and *Heliocidaris erythrogramma* (DNA provided by Rudolf Raff, Indiana University,

Bloomington, IN). Sperm DNA was prepared as described previously (Gan et al., 1990b). RSR sequences were amplified from the sperm DNA from different individuals for each species. PCR was performed using degenerate oligonucleotide primers that were designed based on known *S. purpuratus* RSR sequences (Table S1). To amplify RSRs associated with *spec* genes, a degenerate primer was designed that was based on the conserved sequence in the 5′ untranslated regions of *S. purpuratus spec* genes (Table S1; Hardin and Klein, 1987). Amplification reactions were performed using an annealing temperature of 50–55°C for 30 cycles. PCR products were subcloned and the sequences were authenticated by DNA sequencing.

Genomic Southern hybridization analysis

Approximately 10 μg of genomic DNA was digested with *Eco*RI or *Hin*DIII, size fractionated on a 1% agarose gel, transferred to a nylon membrane, and hybridized overnight at 37°C with a pooled RSR probe consisting of the *spec1*, *spec2a*, and *spec2c* RSRs at equimolar concentration. To maximize detection of divergent genomic RSR sequences, we subjected the membrane to a sequential series of washes at increasing stringencies using two 30-min washes with 2× saline-sodium citrate (SSC) and 0.1% sodium dodecyl sulfate (SDS) at 37°C, 45°C, 55°C, and 65°C. The membrane was then washed at 65°C sequentially with 1× SSC plus 0.1% SDS, 0.5× SSC plus 0.1% SDS, and 0.1× SSC plus 0.1% SDS. Band intensities were detected using a STORM phosphorimager (Amersham, Piscataway, NJ).

Database analysis

We used the National Center for Biotechnology Information Trace Archive (available at http://www.ncbi.nlm.nih.gov/BLAST/mmtrace.shtml) and sea urchin genome project BAC-end sequences (available at http://sugp.caltech.edu/) to identify and download sequences with similarity to the *S. purpuratus spec2a* RSR. Sequences identical to the *spec2a* RSR were discarded. Sequence comparisons were performed using the Clustal W multiple sequence alignment at the

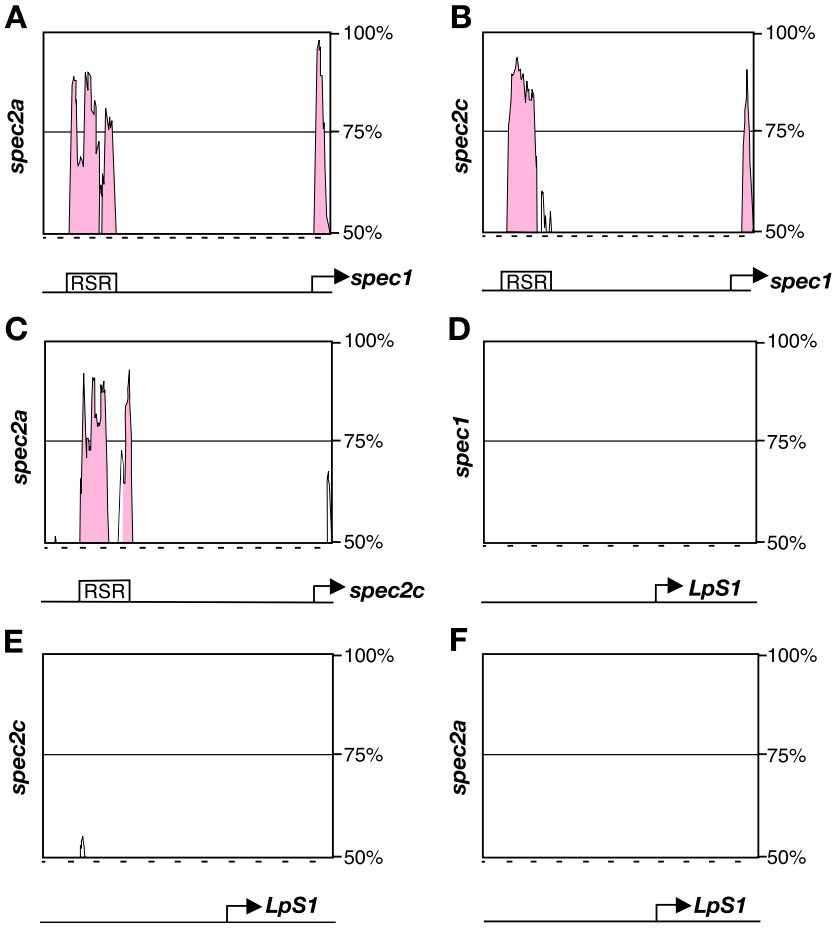

Fig. 1. Sequence comparisons of 5′ upstream regulatory regions of *spec1*, *spec2a*, *spec2c*, and *LpS1β*. VISTA outputs were based on a 100-bp sliding window and a sequence identity range of 50–100%. Relative position of the RSR is depicted below each *S. purpuratus* gene placed on the *x* axis. Regions of similarity greater than 70% are shaded in pink. (A) *spec1* versus *spec2a*. (B) *spec1* versus *spec2c*. (C) *spec2c* versus *spec2a*. (D) *LpS1β* versus *spec1*. (E) *LpS1β* versus *spec2c*. (F) *LpS1β* versus *spec2a*.

MultAlin Web site (at http://prodes.toulouse.inra.fr/multalin/multalin.html). Default DNA-5–0 parameters were used for the alignment (Corpet, 1998).

Phylogenetic tree analysis

Boundaries for the S and SR2 regions were determined by visual comparison to the *spec2a* RSR sequence, and sequences were aligned using Clustal W and verified by eye. Molecular evolutionary analyses were conducted using the MEGA phylogenetics program, version 2.1 (Kumar et al., 2001). Trees were generated using either Neighbor-joining algorithms with a Kimura-2 correction or maximum parsimony and bootstrap values were determined. Missing data were treated with either pairwise or complete deletion.

Results

Evolutionary origin of RSR elements

In the *S. purpuratus* genome, RSRs define a repetitive sequence family consisting of a few hundred interspersed, divergent members. Those associated with *spec1*, *spec2a*, and *spec2c* have been shown to behave as classical transcriptional enhancers; they can enhance transcription of reporter genes in a position- and orientation-independent fashion when tested in a sea urchin embryo gene transfer-expression system using a heterologous promoter (Gan et al., 1990a,b). The RSR upstream of *spec2a* appears to be indispensable for *spec2a* expression, whereas the RSRs associated with *spec1* and *spec2c* appear to function in cooperation with other *cis*-regulatory elements lying beyond the RSR borders. Despite their importance in *S. purpuratus spec* gene expression, we found no evidence that RSR sequences were present within the transcription regulatory region of *LpS1β*, a *spec* gene family member from *L. pictus*. Pairwise sequence comparisons between the upstream sequences of *spec1*, *spec2a*, *spec2c*, and *LpS1β* failed to reveal conserved noncoding sequences (CNSs) except for the *S. purpuratus spec* RSRs (Figs. 1A–C). In particular, we found no detectable CNSs when comparing a 762-bp upstream region from *LpS1β*, a region known to be necessary and sufficient for correct expression (Xiang et al., 1991), with any of the upstream regions from the *spec* genes (Figs. 1D–F). Moreover, the consensus sequences TAATCC/T, to which SpOtx and SpGsc bind with high affinity, were not present in the 762-bp *LpS1β* upstream region (data not shown).

The absence of an RSR element upstream of *LpS1β* could reflect a general lack of conservation of RSR elements in sea urchin genomes other than *S. purpuratus*. Indeed, it is possible that RSRs are unique to *S. purpuratus* and that other sea urchin species lack this repetitive sequence family. To gain more information on the occurrence of RSRs in the genomes of other species, we performed genomic Southern hybridization analyses. We chose species within the family Strongylocentrotidae (*S. purpuratus*, *H. pulcherrimus*, and *S. franciscanus*) and more distant species (*H. erythrogramma*, *H. tuberculata*, and *L. pictus*) (Lee, 2003). Genomes of species that diverged after the split between *S. purpuratus* and *S. franciscanus*, which occurred 13–18 mya, contained multiple copies of the RSR element, whereas genomes of species that diverged before the split between *S. purpuratus* and the *Heliocidaris* genus, which occurred 30–35 mya, had no detectable RSR elements (Fig. 2). The results suggested that the RSR repetitive family arose between 13 and 35 mya, most likely within the genome of the ancestral species that gave rise to the Strongylocentrotidae family.

Comparisons of Strongylocentrotidae RSRs

To understand the basis for variable *spec* RSR enhancer function and to determine whether non-*spec* RSR elements contained *cis*-regulatory elements like those found in the *spec* RSRs, we compared the nucleotide sequences of the three *spec* RSRs and several other RSRs mined from the *S. purpuratus* genome. In Fig. 3A, we show a representative analysis of the "S" region from ten RSR elements. We compared the S region because the enhancer activity within the *spec2a* RSR is largely confined to the *cis*-regulatory elements within this region (Mao et al., 1994; Yuh et al.,

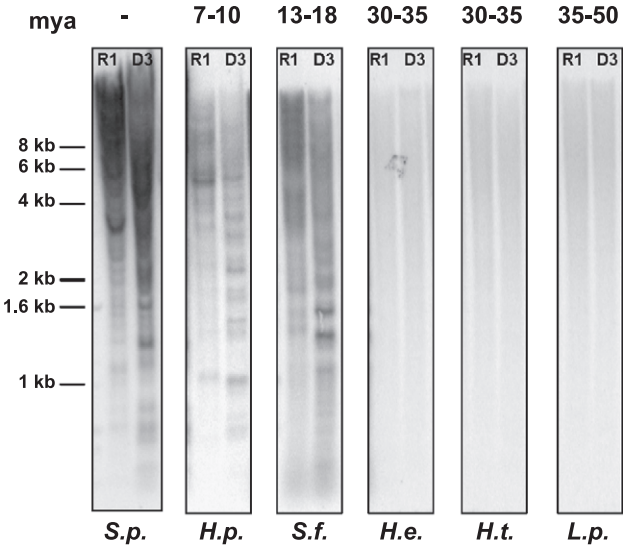

Fig. 2. Occurrence of the RSR repetitive element in sea urchin genomes. Genomic DNA from different species was digested with *Eco*RI (R1) or *Hin*DIII (D3) and used for Southern hybridization with a pooled RSR probe. The hybridized filter was washed at low stringency to maximize cross-species hybridization. Divergence times for each species from *S. purpuratus* in millions of years (mya) is indicated above each lane, and the species is indicated below each lane. S.p., *S. purpuratus*; H.p., *H. pulcherrimus*; S.f., *S. franciscanus*; H.e., *H. erythrogramma*; H.t., *H. tuberculata*; L.p., *L. pictus*.

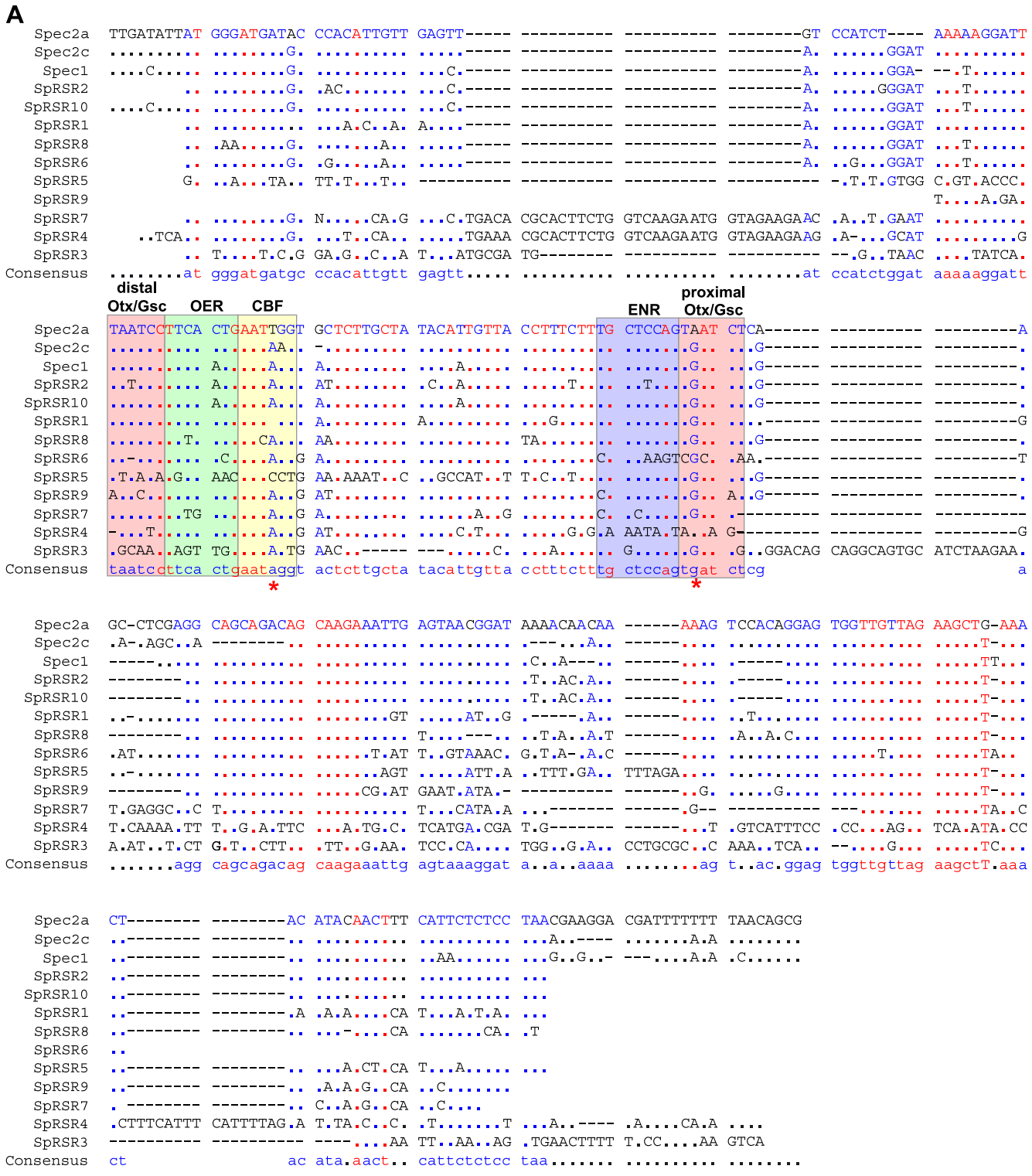

Fig. 3. Intraspecies and interspecies RSR alignments. (A) RSR elements from the *S. purpuratus* genome were identified, and their S regions were aligned. For simplicity, RSR identities were arbitrarily assigned (RSR1, RSR2, etc.). Identification of the RSRs by their Sea Urchin Genome Project trace number is available on request. (B) Alignments of S regions of RSR elements from *S. droebachiensis* (Sdroeb3-2), *S. franciscanus* (Sfran3-2), and *H. pulcherrimus* (Hpulch1-1). Sequences corresponding to defined *spec2a cis*-regulatory elements are boxed: red, proximal and distal Otx/Gsc; green, OER; yellow, CBF; blue, ENR. Dots represent nucleotides unchanged from *spec2a* and dashes indicate indels. Red nucleotides indicate that the position is greater than 90% conserved and blue nucleotides indicate that the position occurs in most cases. The asterisks below the consensus sequence indicate the newly acquired nucleotides in the *S. purpuratus spec2a* CBF and proximal Otx/Gsc elements (A and T, respectively). The lowercase letters on the last line depict the consensus sequence.

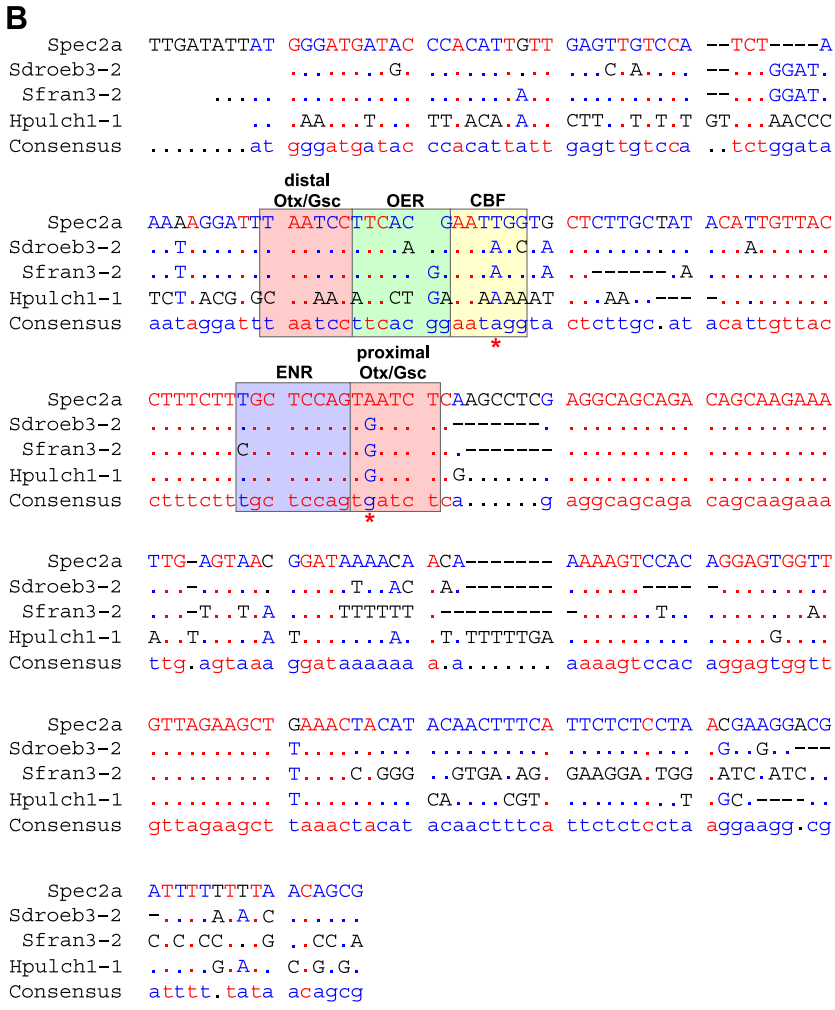

Fig. 3 (*continued*).

2001). The comparison revealed several striking features, particularly when the previously identified *cis*-regulatory elements of *spec2a* were aligned with the corresponding elements from other S regions.

As might be expected for a repetitive sequence family in the *S. purpuratus* genome, we found substantial sequence divergence among the different S regions, including many single base pair changes as well as insertions and deletions (indels) (Fig. 3A). However, several nucleotide positions were highly conserved; these positions were conserved in most cases and with greater than 90% identity in many cases. The simplest explanation for the nucleotide sequence conservation was that it reflected the ancestral state at most of the positions. A consensus S region sequence, which would be predicted to resemble the nucleotide sequence of the pre-amplified ancestral element, is shown on the bottom line of each chart in Fig. 3A.

The sequences defining the five *cis*-regulatory elements from the *spec2a* S region were substantially different from the corresponding S regions from both *spec* and non-*spec* RSR elements (Fig. 3A). The distal Otx/Gsc element was conserved in *spec1* and *spec2c*, but in SpRSR2 and SpRSR9, TAATCC changed to TATTCC and AAACCC, respectively. These changes would be predicted to significantly reduce SpOtx binding, thereby reducing enhancer activity. The OER element of *spec2a* was also altered in other S regions. In particular, TTCACTG was changed to TTCAATG in *spec1*. We also noted that a single base pair in the CBF element of *spec2a* (in the reverse orientation) was altered in almost all other S regions examined; rather than the T found in the fourth position of the AATTGG *spec2a* sequence, most other sequences, including *spec1* and *spec2c*, contained an A (AAT*A*GG for *spec1* and AAT*A*AG for *spec2c*). These results suggested that the ancestral RSR element did not contain a consensus SpCBF-binding site but that rather this site was recently acquired in the *spec2a* RSR. Another notable single base pair alteration was found in the proximal Otx/Gsc element. In this case, the TAATCT element of *spec2a* was T*G*ATCT in almost all other sequences examined, including *spec1* and *spec2c*. As with the CBF element, the proximal Otx/Gsc element appeared to be a recent *spec2a* RSR acquisition because it was not found in any other S regions that were analyzed. On the basis of the deviations from the consensus binding sites, we

predicted that the ancestral sequences found in *spec1* and *spec2c* would have lower-affinity binding sites for SpCBF (AATTGG versus AATAGG or AATAAG) and for SpOtx, SpGsc, and SpGATA-E (TAATCT versus TGATCT).

We also amplified via PCR non-*spec* RSR elements from *S. franciscanus*, *S. droebachiensis*, and *H. pulcherrimus* genomic DNA and compared their S regions with that of *spec2a* (Fig. 3B). Similar sequence divergences were found throughout the S regions of RSRs from these related species, as was observed with *S. purpuratus* RSRs. Most notably, counterparts to the CBF and proximal Otx/Gsc *cis*-regulatory elements found in the *spec2a* S region contained the identical nucleotides that were found in the *S. purpuratus* non-*spec2a* S regions, namely an A rather than a T for the CBF element (AATTGG to AATAGG) and a G rather than an A for the proximal Otx/Gsc element (TAATCT to TGATCT). The results supported the hypothesis that these two *spec2a* RSR *cis*-regulatory elements

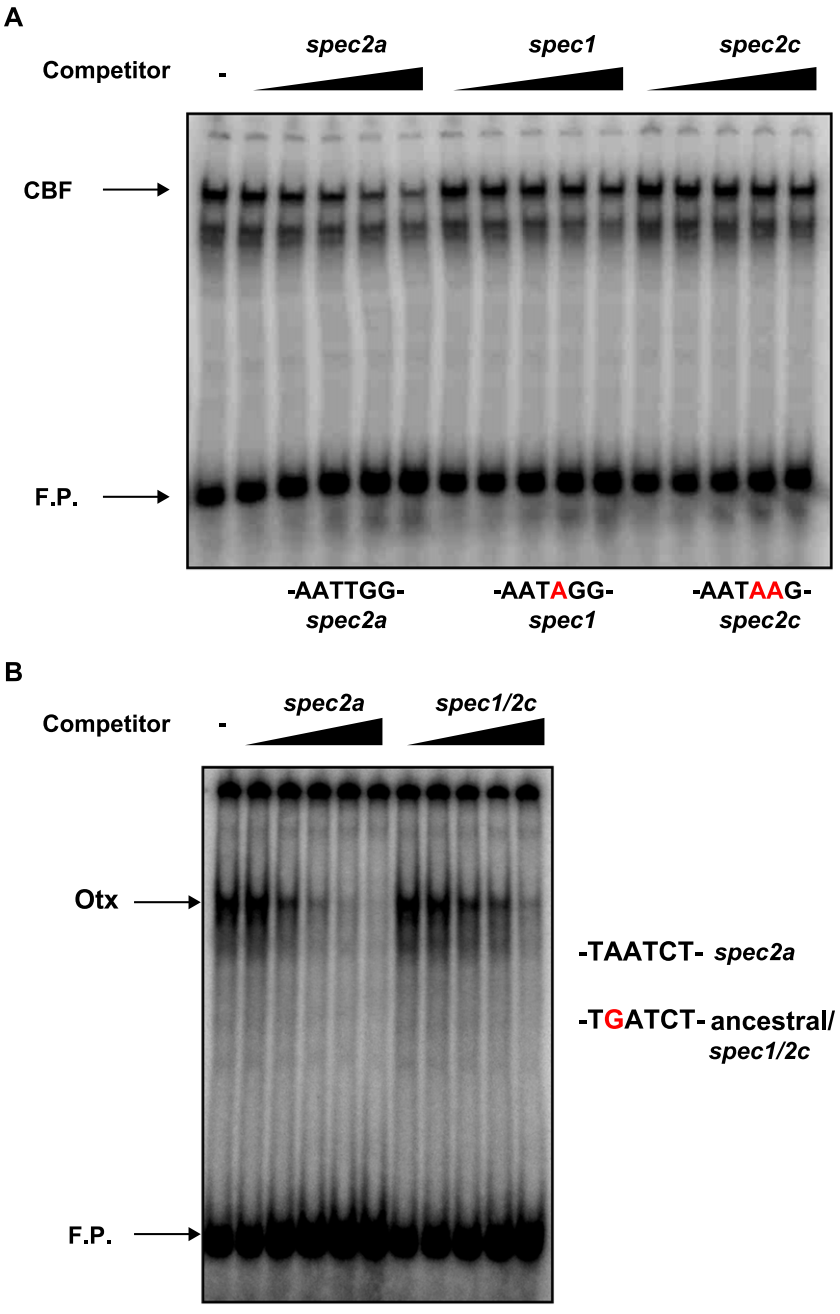

Fig. 4. DNA–protein binding affinities for SpCBF and SpOtx at CBF and proximal Otx/Gsc elements. EMSA using blastula nuclear extracts was used to detect (A) SpCBF and (B) SpOtx. A high-affinity CBF or Otx probe was used to form an SpCBF–DNA or SpOtx–DNA complex, respectively. The left lane in each panel (−) shows the respective complex and the free probe (F.P.). The remaining lanes show the extent of complex formation in the presence of excess unlabeled oligonucleotides representing the CBF or proximal Otx/Gsc site for *spec2a*, *spec1*, and *spec2c*.

were recently derived from ancestral sequences and that these ancestral sequences were conserved in a large majority of non-*spec2a* RSRs, including *spec1* and *spec2c*.

The sequence relationships among S regions from 24 different *S. purpuratus* RSRs were established by constructing an unrooted phylogenetic tree (data not shown). Although the tree was not sufficiently robust for generating a single solution, the analysis did suggest that *spec1 spec2a*, and *spec2c* S regions formed a distinct cluster group. The clustering suggested that the *spec*-associated S regions derived from a recent amplification event, probably in association with the amplification of the ancestral *spec* gene sometime during Strongylocentrotidae cladogenesis. However, as Fig. 3A showed, the *spec* S sequences were by no means identical, and the differences observed in the OER, CBF, and Otx/Gsc elements would be predicted to affect transcription factor binding and enhancer activity.

Creation of four transcription binding sites from two derived nucleotides in the spec2a RSR

The nucleotide sequence differences between the ancestral and derived CBF and proximal Otx/Gsc elements should produce measurable differences in the binding affinities of the corresponding transcription factors. EMSA was performed to determine whether SpCBF, SpOtx, SpGsc, and

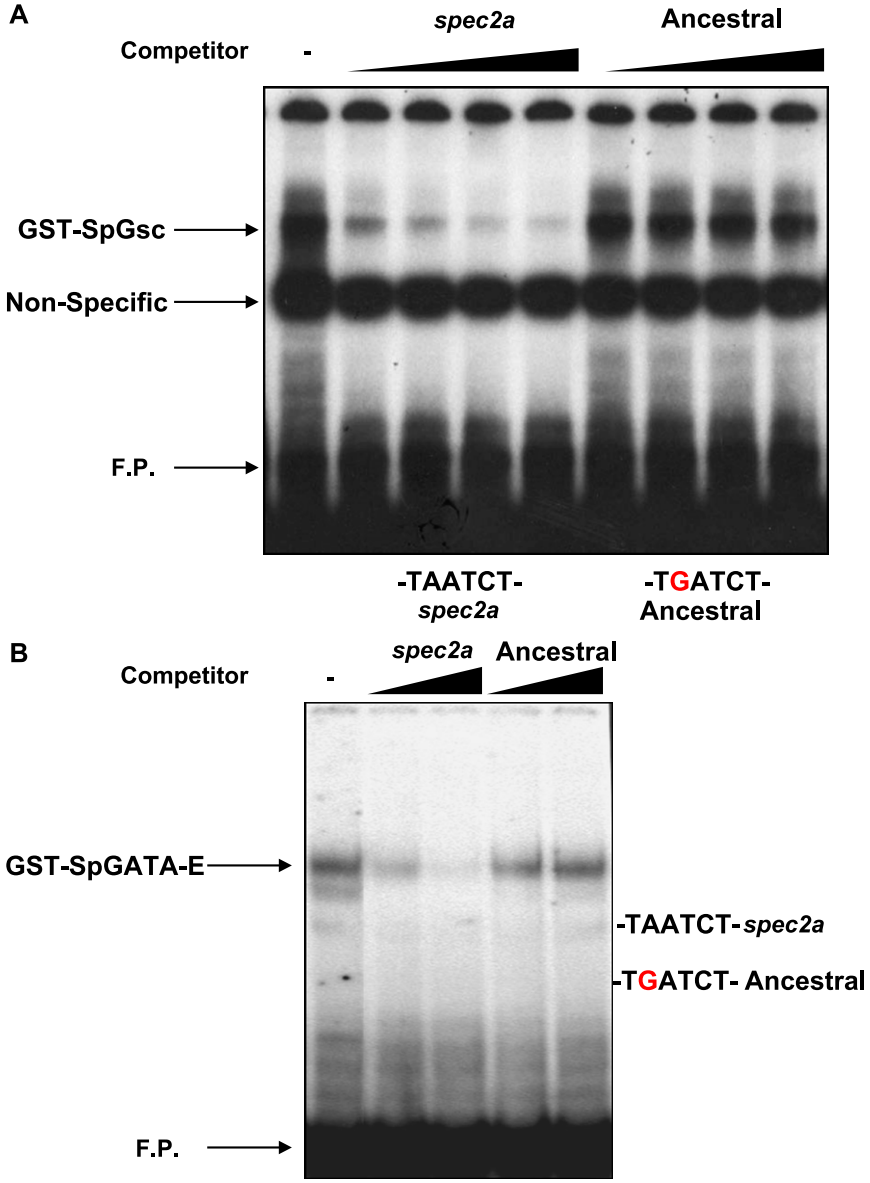

Fig. 5. DNA–protein binding affinities for SpGsc and SpGATA-E at proximal Otx/Gsc elements. EMSA using bacterial extracts was employed to detect (A) recombinant GST-SpGsc or (B) recombinant GST-SpGATA-E zinc finger domain. A high affinity Otx/Gsc or GATA probe was used to form an SpGsc or SpGATA-E complex, respectively. The left lane in each panel (−) shows the complex and the free probe (F.P.). The remaining lanes show the extent of excess unlabeled oligonucleotides representing the proximal Otx/Gsc site for *spec2a* or an ancestral sequence. The complex labeled "nonspecific" in panel A is found in extracts from uninduced cells.

SpGATA-E bound with the same or different affinities to the ancestral elements in *spec1* and *spec2c* that corresponded to the CBF and proximal Otx/Gsc *cis*-regulatory elements in *spec2a*. Blastula nuclear extracts were used as a protein source to form SpCBF– and SpOtx–DNA complexes, and recombinant GST fusion proteins were used for forming SpGsc and SpGATA-E complexes. An SpCBF–DNA complex with a consensus CBF binding site (AATTGG) oligonucleotide formed (Fig. 4A; Li et al., 2002). Complex formation was effectively inhibited by competition with the homologous oligonucleotide but much less so with an oligonucleotide containing the *spec1* (AATAGG) or *spec2c* (AATAAG) sequences (Fig. 4A). Similarly, an SpOtx complex was formed with an Otx consensus oligonucleotide (TAATCC). The formation of this complex was blocked by an oligonucleotide with the proximal Otx/Gsc sequence (TAATCT), but an oligonucleotide with the ancestral sequence (TGATCT) was much less effective (Fig. 4B).

We observed even more dramatic results with SpGsc and SpGATA-E using the same oligonucleotides we used to form SpOtx complexes. The SpGsc- and SpGATA-E complexes were strongly inhibited with the TAATCT-containing oligonucleotide but the ancestral oligonucleotide was largely ineffective (Fig. 5). In summary, the EMSA results demonstrated that the recent acquisition of the CBF element in the *spec2a* RSR generated a new binding site for SpCBF. Even more notable, the single-base pair change to generate the proximal Otx/Gsc (TAATCT) element simultaneously created a new binding site for three distinct transcription factors, SpOtx, SpGsc, and SpGATA-E. Although we previously demonstrated functional roles for SpCBF, SpOtx, and SpGsc at their binding sites within the *spec2a* RSR enhancer (Angerer et al., 2001; Li et al., 2002; Mao et al., 1996; Yuh et al., 2001), the role of SpGATA-E at the Otx/Gsc element must still be defined.

We also determined whether the nucleotide alteration in the OER element in the *spec1* RSR affected binding to OER. An OER–DNA complex was formed with blastula nuclear extracts and oligonucleotide probes containing the OER-binding site corresponding to the *spec2a* (TTCACTG) or *spec1* (TTCAATG) sequence (data not shown). The shifted bands were of approximately equal intensity, indicating that the OER factor bound to each sequence with equal affinity. This result suggests that the OER element was likely to be functionally equivalent in the *spec1*, *spec2a*, and *spec2c* RSR enhancers.

Relative enhancer activities of the RSR elements

To determine whether the sequence differences among the *spec1*, *spec2a*, and *spec2c* RSR elements led to quantifiable differences in enhancer activity, we generated constructs containing RSR sequences upstream from the *endo16* basal promoter and monitored luciferase reporter gene activity in sea urchin embryo gene transfer-expression assays. We found that all three *spec* RSR elements had substantial enhancer activity: compared with the basal promoter, the *spec1* RSR stimulated expression 17.5-fold, the *spec2c* RSR 20-fold, and the *spec2a* RSR 50-fold (Fig. 6A). Consistent with previous deletion studies (Gan et al., 1990b), the *spec2a* RSR reproducibly led to 2.5- to 3.0-fold higher activation than the two other *spec* RSRs. The inherent nucleotide sequence differences among the *spec* RSR sequences likely caused the differences in enhancer activity.

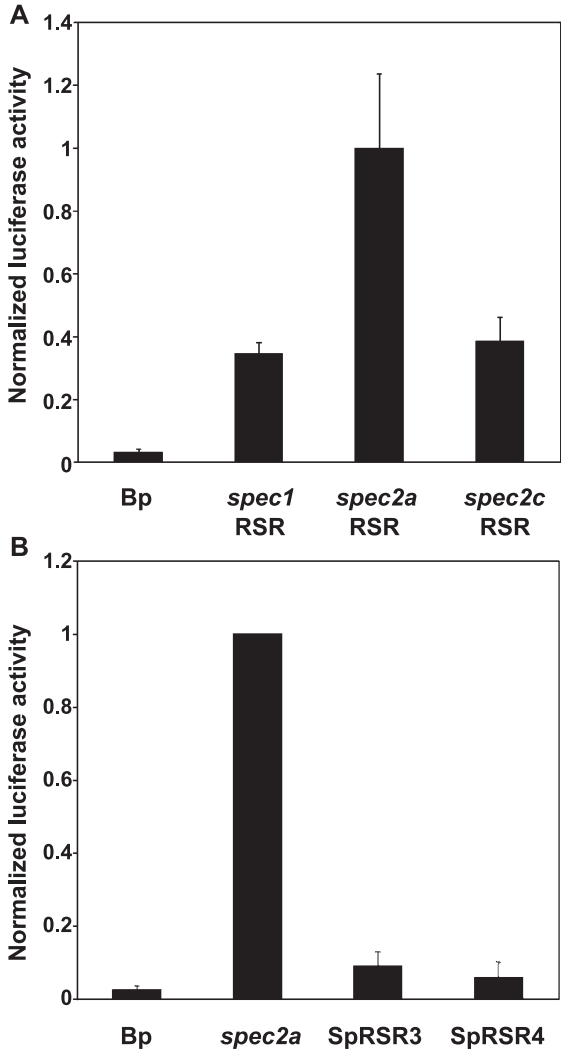

Fig. 6. Relative enhancer activities of *spec* and non-*spec* RSRs. (A) Relative enhancer strengths of *spec1*, *spec2a*, and *spec2c* RSRs. *S. purpuratus* zygotes were injected with the indicated RSR-*endo16* basal promoter:firefly luciferase construct, cultured, and harvested at the late blastula stage. (B) Relative enhancer activities of non-*spec* RSRs. The 78-bp C fragment from *spec2a*'s RSR, SpRSR3, and the R1S region of SpRSR4 were fused to the *endo16* basal promoter and the firefly luciferase gene. The constructs were introduced into *S. purpuratus* zygotes using a particle gun delivery system, cultured, and harvested at the late blastula stage. In both A and B, activity values are normalized to a co-injected cytomegalovirus-driven Renilla luciferase control. Bp indicates the construct with only the *endo16* basal promoter. Data are presented as mean ± standard deviations.

It was possible that enhancer activity would be associated with all RSR elements, perhaps by fortuitous, nonspecific mechanisms. However, this possibility was unlikely because of the extensive sequence divergence among the non-*spec* RSRs. To rule out general, nonspecific RSR enhancer activity, we tested the activity of another RSR element, SpRSR3, which was highly divergent from the *spec* RSRs and had no discernible Otx/Gsc or CCAATT consensus elements (Fig. 3A). As a negative control, we used a DNA fragment that contained only a 3′ portion of the upstream R (R1) and a 5′ portion of the S from SpRSR4 (Fig. 3A). This truncated RSR sequence was not expected to have substantial enhancer activity. We observed no greater stimulation of reporter gene expression with the SpRSR3 or the truncated SpRSR4 constructs than with the basal promoter (Fig. 6B). These results suggested that the *spec* RSRs evolved into transcriptional enhancers at or after the time they became associated with *spec* genes or that SpRSR3 had lost its original enhancer function.

Mutational analysis of the CBF and proximal Otx/Gsc cis-regulatory elements within the spec2a RSR enhancer

To address why the *spec2a* RSR was a substantially stronger enhancer than the *spec1* or *spec2c* RSRs, we generated point mutations in the *spec2a* RSR to produce the ancestral sequences at the CBF and proximal Otx/Gsc elements, namely AATAGG in place of AATTGG, TGATCT in place of TAATCT, or a combination of both mutations. We also mutated the ancestral nucleotides in the *spec1* RSR to produce the *spec2a*-derived CBF and proximal Otx/Gsc elements within the framework of the *spec1* RSR sequence.

A single-point mutation in the *spec2a* RSR proximal Otx/Gsc element reduced activity 40–50%, and a single-point mutation in the CBF element reduced activity 25–30% compared with that in the unaltered *spec2a* construct (Fig. 7A). As might be anticipated, these results implied that SpOtx and SpCBF act as positive transcription factors at the

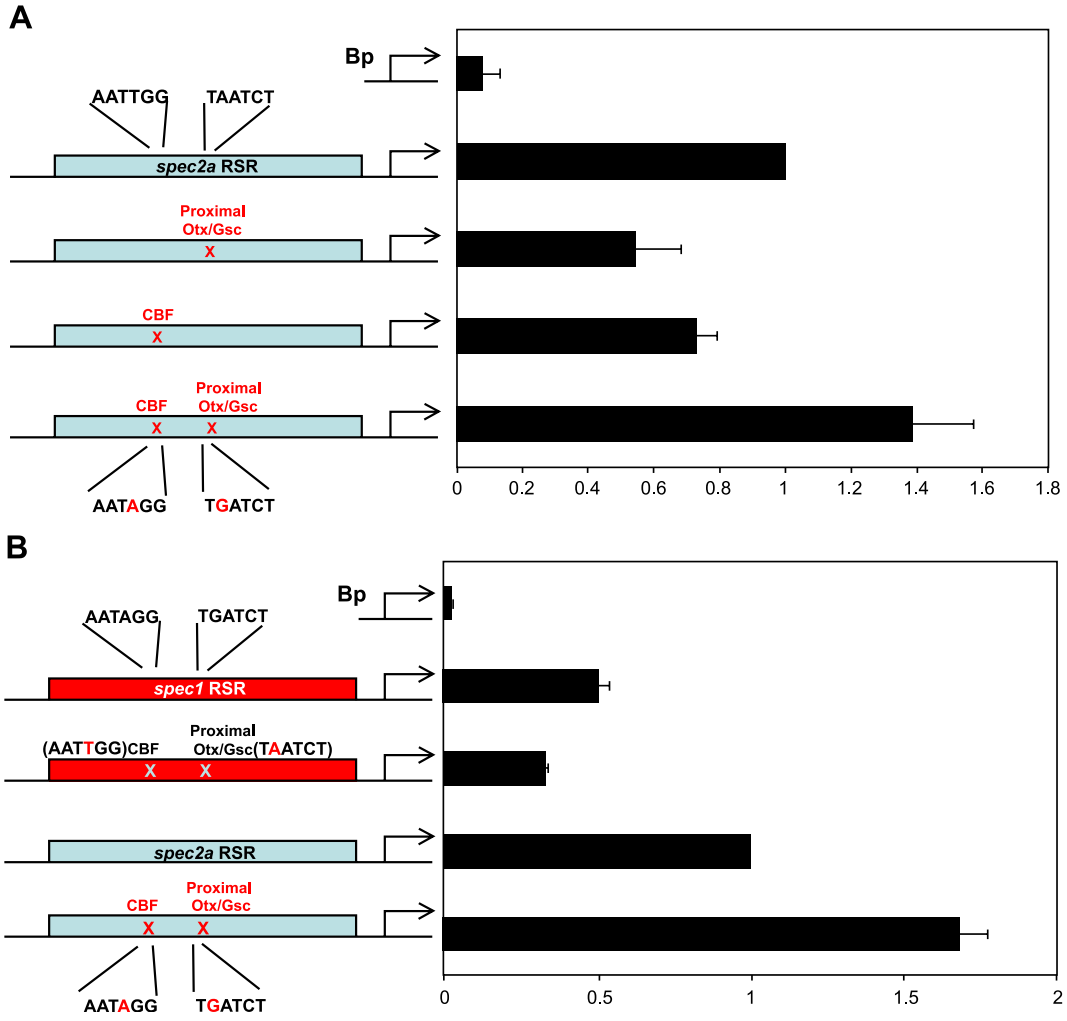

Fig. 7. Enhancer activities of mutated *spec1* and *spec2a* RSRs. (A) Relative activities associated with wild-type and mutant *spec2a* RSR-*endo16* basal promoter:luciferase constructs. Schematic representations indicate point mutations in the CBF element, proximal Otx/Gsc element, or both elements. (B) Relative activities associated with *spec1* RSR:*endo16* basal promoter:luciferase constructs. Schematic representations indicate point mutations in the CBF and proximal Otx/Gsc elements. Mutated nucleotides are shown in red.

spec2a elements and that neither factor could act as effectively at its corresponding ancestral element. However, a mutant *spec2a* construct containing both nucleotide changes exhibited a highly reproducible 40% increase in reporter expression compared with the unaltered *spec2a* construct (Fig. 7A). Moreover, we observed a reproducible 30–40% reduction in reporter gene expression with the *spec1* RSR-mutant construct compared with the unaltered *spec1* construct (Fig. 7B). Thus, the combined effect of these two nucleotide changes in the *spec2a* RSR was the acquisition of a transcriptional repression function. This result was somewhat unexpected because the *spec2a* RSR was, overall, a stronger enhancer than the *spec1* RSR (Figs. 6A and 7B). The results suggested that sequences outside of the CBF and proximal Otx/Gsc elements were responsible for the stronger overall enhancer activity of the *spec2a* RSR. Each individual nucleotide change in the *spec2a* RSR led to reduced enhancer activity, but the combined mutations led to an increase. This observation suggested that a functional interaction exists between the derived CBF and proximal Otx/Gsc elements to mediate repressor activity.

Polymorphism at the spec2a RSR locus

S. purpuratus populations are highly polymorphic, averaging 4–5% nucleotide sequence differences between individuals (Britten et al., 1978). To investigate whether the derived CBF and Otx/Gsc elements associated with the *spec2a* RSR enhancer were polymorphic or fixed in the population, we compared the DNA sequences of 10 *S. purpuratus* individuals obtained from the Pacific coasts of southern California and southern British Columbia.

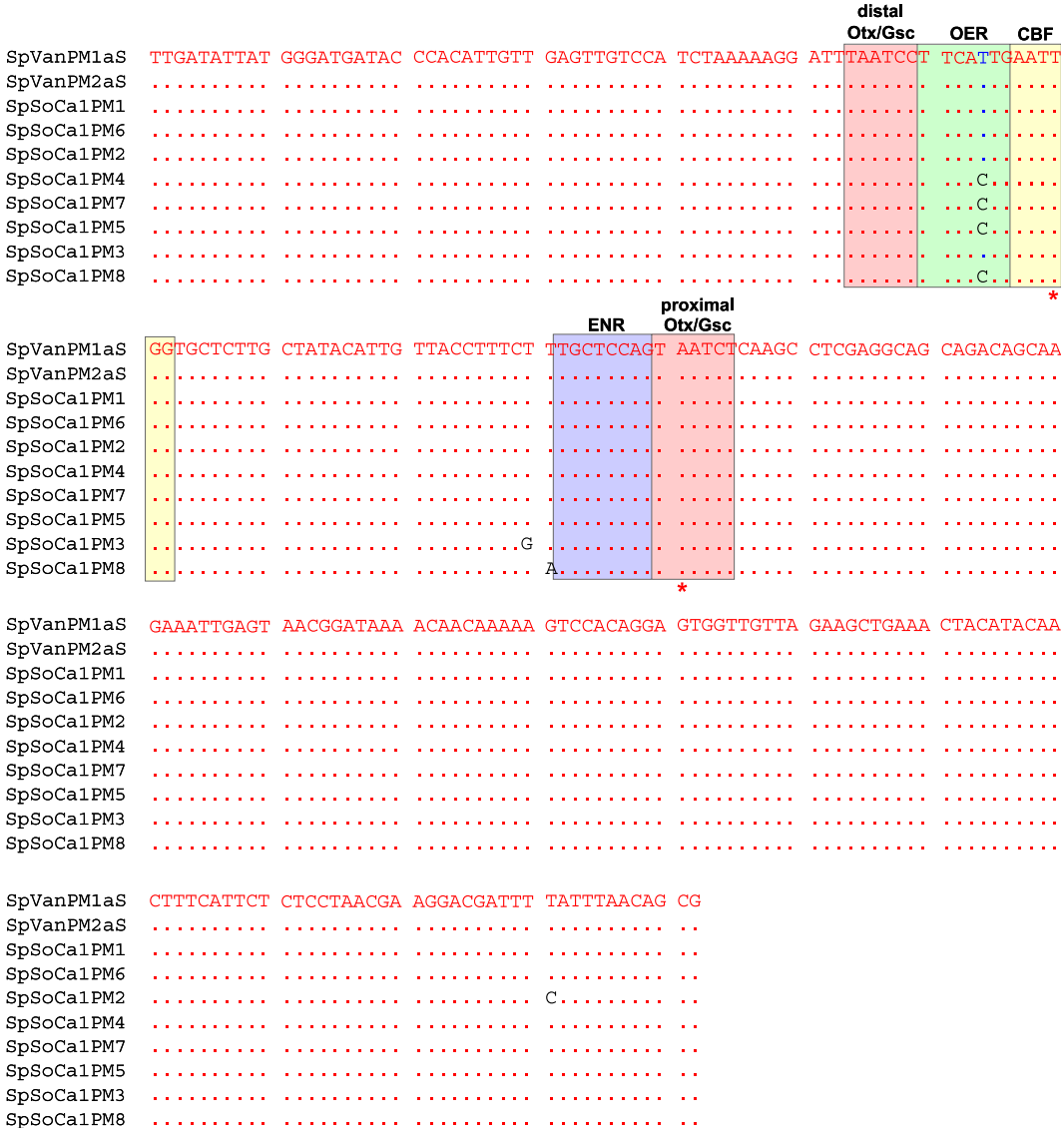

Fig. 8. Polymorphism within the *S. purpuratus spec2a* RSR. S regions from 10 *spec2a* RSR alleles are aligned with the *cis*-regulatory elements in colored boxes, and other designations are as indicated in Fig. 3. SpVanPM, *S. purpuratus* Vancouver individual; SpSoCalPM, *S. purpuratus* southern California individual.

Sequence comparisons of a non-*spec* RSR (SpRSR3) locus from eight of these individuals showed approximately 5.4% polymorphism (data not shown). In contrast, the *spec2a* RSR sequence showed only 1.0% polymorphism and, in particular, the CBF and Otx/Gsc elements were completely monomorphic (Fig. 8). Nevertheless, the *spec2a* RSR sequence was not impervious to change. For example, the OER element was polymorphic at one position (TTCACTG and TTCATTG). The derived CBF and proximal Otx/Gsc elements were fixed in the two geographically separated *S. purpuratus* populations, strongly suggesting that functional selection had occurred at these nucleotide positions.

spec2a RSRs from other Strongylocentrotid species

The occurrence of RSR elements in the genomes of other Strongylocentrotid species provided an opportunity for us to determine whether the derived CBF and proximal Otx/Gsc *cis*-regulatory elements were present in orthologous *spec2a* RSRs. It was possible that these elements were generated very recently at the time of speciation of *S. purpuratus*. Alternatively, these elements may have originated with the branching of the Strongylocentrotidae family or within some later clade that included *S. purpuratus spec2a* RSR sequences amplified by PCR, were obtained from *S. pallidus*, *S. droebachiensis*, and *H. pulcherrimus* DNA. *H. pulcherrimus spec2a* RSRs were monomorphic at both sites; all 12 individuals examined contained the ancestral TGATCT sequences, indicating that this derived element originated after *S. purpuratus* and *H. pulcherrimus* diverged approximately 7–10 mya (Fig. 9A) (Lee, 2003). The CBF site was AATTGA in almost all individuals except for the *H. pulcherrimus* allele 7, in which an indel dispersed both the OER and CBF elements (data not shown).

S. purpuratus, *S. droebachiensis*, and *S. pallidus* are very closely related species and are thought to have diverged from one another very rapidly (Biermann, 1998). We sampled two *S. droebachiensis* individuals and, as observed with the *S. purpuratus spec2a* RSRs, we found limited sequence differences between these two orthologous *spec2a* RSR alleles (data not shown). However, the ancestral TGATCT sequence was present in both individuals, indicating that the proximal Otx/Gsc element was fixed in *S. purpuratus* after the divergence of these species from a common ancestor (Fig. 9B). We also found that the two alleles had substantial differences from the *S. purpuratus*

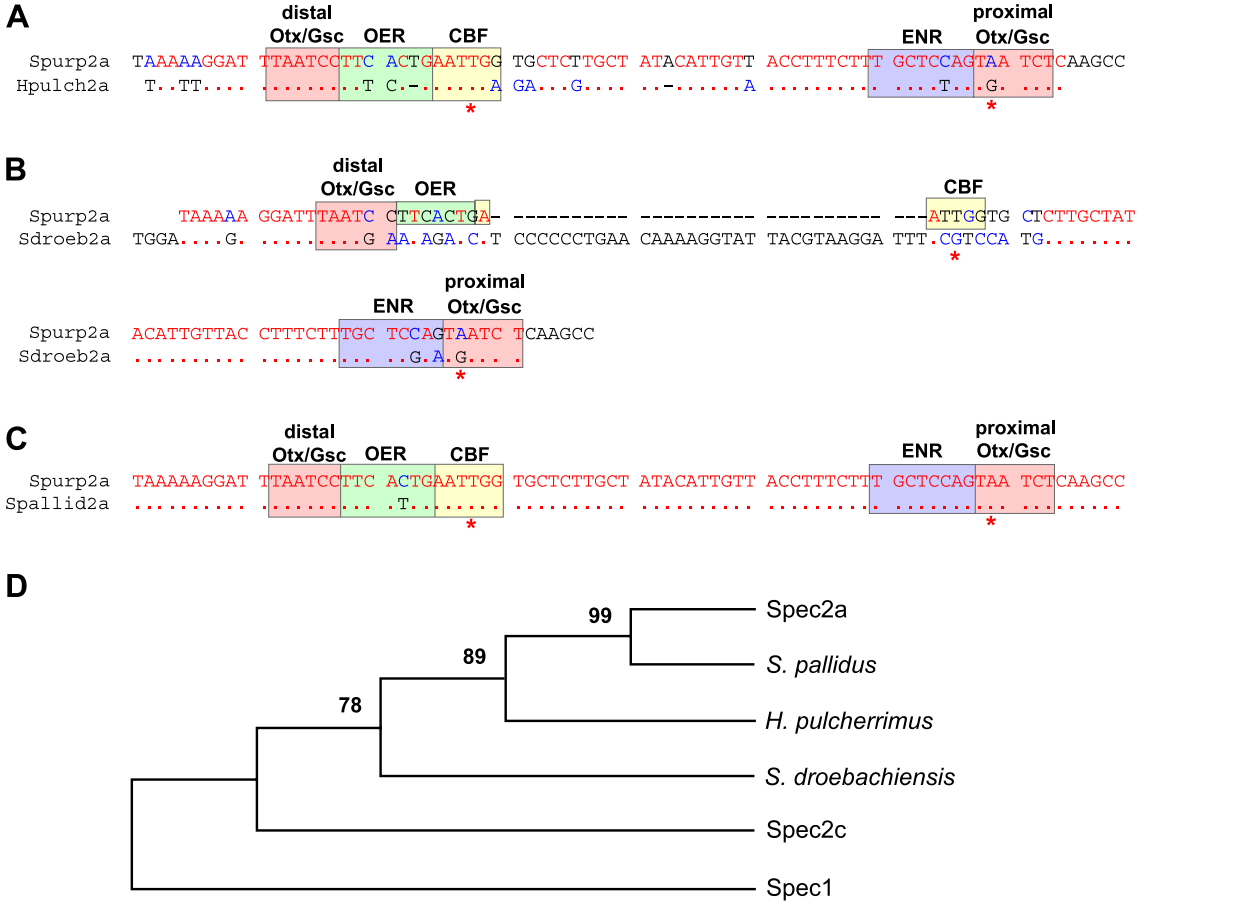

Fig. 9. Comparisons of *spec2a* RSR sequences from *S. purpuratus*, *S. pallidus*, *S. droebachiensis*, and *H. pulcherrimus*. Alignments of sequences corresponding to the *spec2a* C fragment from *S. purpuratus* and: (A) *H. pulcherrimus*; (B) *S. droebachiensis*; and (C) *S. pallidus*. (D) Maximum parsimony phylogenetic tree based on *spec2a* S-R2 sequences. The *spec2a*, *spec1*, and *spec2c* sequences are from *S. purpuratus*. Bootstrap values are shown at each node.

spec2a RSR sequence in the region corresponding to the distal Otx/Gsc-OER-CBF elements; this region was disrupted relative to the *S. purpuratus* sequence and was lacking all three of these *cis*-regulatory elements (Fig. 9B). This disruption was reminiscent of the indel in *H. pulcherrimus spec2a* RSR allele 7, although the import of these deleted sequences is not yet clear.

We sequenced two *spec2a* RSR alleles from *S. pallidus* and found that in this species, the derived CBF and proximal Otx/Gsc sites were present in both alleles (Fig. 9C). These results suggested that the derived elements originated before the separation of *S. purpuratus* and *S. pallidus*, and were subsequently fixed in both species. The relationships of these sequences were further established by constructing a *spec2a* RSR gene tree, including the sequences displayed in Figs. 9A–C. The tree was highly robust and showed the relationships among the sequences as follows: {*S. droebachiensis* [*H. pulcherrimus* (*S. purpuratus*, *S. pallidus*)]} (Fig. 9D).

Discussion

Basal and derived cis-regulatory mechanisms for spec gene expression

The *cis*-regulatory elements associated with *spec* gene expression show considerable interspecies and intraspecies variability despite the tight constraints placed on the spatiotemporal expression pattern of *spec* genes. The *spec* RSR elements behave as transcriptional enhancers and have important functions in controlling *spec* expression in *S. purpuratus*, but these repetitive sequences were not found in the genomes of sea urchins outside of Strongylocentrotidae. In particular, the *L. pictus spec* gene, *LpS1β*, appears to make use of an entirely different *cis*-regulatory mechanism for its expression. In addition, we found major differences in the nucleotide sequences and transcriptional activities of the RSR enhancers associated with the three *S. purpuratus spec* genes, *spec1*, *spec2a*, and *spec2c*. The *spec2a* RSR was a stronger enhancer than the others, and two *cis*-regulatory elements were identified in the *spec2a* RSR as novel acquisitions recently derived from ancestral sequences.

Previous results from our laboratory indicated that the *spec2a* RSR enhancer is largely sufficient for correct aboral ectoderm expression but that this was not the case for the *spec1* and *spec2c* RSR enhancers (Gan et al., 1990a). More than five kb of 5′ upstream *spec1* DNA, which included an RSR element, were unable to confer the correct expression pattern to a *lacZ* reporter gene in a sea urchin gene transfer-expression assay, suggesting that additional sequences positioned farther upstream or downstream were required (Gan et al., 1990a). The differences in nucleotide sequences and transcription factor binding affinities between the *spec1* and *spec2a* RSRs indicate that the *spec2a* RSR has been optimized over evolutionary time to become the major control region driving *spec2a* expression without the need for additional *cis* elements.

We propose an evolutionary model based on two separate *cis*-regulatory mechanisms, one basal and the other derived, to explain the interspecies and intraspecies differences in *spec* gene transcriptional control regions. This model is analogous to the "duplication–degeneration–complementation" model for stepwise gene evolution (Force et al., 1999). Our model proposes that a basal transcriptional control mechanism without an RSR enhancer is required for *spec* gene expression in *L. pictus* and other sea urchin species outside of the Strongylocentrotidae family. The basal control mechanism was partially disrupted by the acquisition of *spec* RSR enhancers during the evolution of the Strongylocentrotidae. In the case of *spec1* and *spec2c*, both basal and RSR mechanisms may still be operating. During Strongylocentrotidae cladogenesis, an RSR element with minimal enhancer function was inserted into the 5′ upstream region of the ancestral *spec* gene between the basal *cis* elements and the transcriptional start site. This insertion was tolerated because it did not disrupt the basal elements but did provide an opportunity for additional change. Optimization of the RSR enhancer over recent evolutionary time and degeneration of the basal *cis*-regulatory elements ultimately led to the complete dependence of *spec2a* on its RSR, whereas *spec1* and *spec2c* continue to rely partially on the basal mechanisms. The distinction between *spec* RSRs may have been the result of RSR positioning: in *spec2a*, the RSR element is adjacent to the transcriptional start site, whereas in *spec1* and *spec2c*, the RSRs are several kilobases upstream.

Evolution of transcriptional mechanisms responsible for aboral ectoderm-specific expression of spec2a

Although the *spec2a* RSR enhancer contains two other more distal Otx/Gsc *cis*-regulatory elements, a 78-bp fragment within the S region of the *spec2a* RSR has sufficient sequence information to activate reporter gene expression preferentially in aboral ectoderm cells and to repress it in oral ectoderm and endoderm cells (Mao et al., 1994; Yuh et al., 2001). The results presented here demonstrate that the proximal Otx/Gsc element, which binds in a sequence-specific fashion to SpOtx, SpGsc, and SpGATA-E, is unique to *spec2a*. These same factors have little or no affinity for the ancestral sequence present within all other spec RSRs, including *spec2a* RSR orthologs in *H. pulcherrimus* and *S. droebachiensis*.

The proximal Otx/Gsc element appears to have multiple, diverse functions. Oral ectoderm repression relies on an antagonistic relationship between SpOtx and SpGsc at both the proximal and distal Otx/Gsc elements (Angerer et al., 2001). The gene encoding SpGsc is largely restricted in its expression to the oral ectoderm territory, where SpGsc is thought to prevent SpOtx binding, and to recruit corepressor proteins through its Engrailed-like repression domain

(Angerer et al., 2001). Our results are consistent with the hypothesis that the proximal and distal Otx/Gsc elements act together to promote oral ectoderm repression and that this function has been recently acquired. The distal Otx/Gsc element contains a canonical consensus TAATCC sequence that binds to Otx factors with 10-fold higher affinity than the TAATCT sequence present in the proximal Otx/Gsc element does (Klein and Li, 1999). Therefore, the distal element probably plays the main role in SpOtx-mediated activation and the proximal element serves an auxiliary function. We find it intriguing that *HpArs* gene expression in *H. pulcherrimus* embryogenesis is restricted to the aboral ectoderm and is under the control of *cis*-regulatory elements that include Otx-binding sites (Kurita et al., 2003; Sakamoto et al., 1997). Oral ectoderm repression at these sites may be conferred through a Gsc factor, as is the case for *spec2a* in *S. purpuratus*.

The ENR element overlaps the proximal Otx/Gsc element within the *spec2a* S region and has been shown to bind to a blastula nuclear factor that may in fact be SpGATA-E (T. Kiyama and W. H. Klein, unpublished results). GATA factors can function as transcriptional activators or repressors, depending on the promoter and cellular environment (Letting et al., 2004; Svensson et al., 2000). However, from our previous characterization of ENR (Yuh et al., 2001), SpGATA-E is unlikely to be the sole protein involved in endoderm repression at the overlapping ENR-proximal Otx/Gsc elements. Relevant to the present study, SpGATA-E does not bind to the distal Otx element or to the ancestral TGATCT element corresponding to the proximal Otx/Gsc element (T. Kiyama and W. H. Klein, unpublished results). If SpGATA-E is involved in endoderm repression, it would imply that a single-base pair alteration from G to A led to a recently evolved *cis*-regulatory element that functions in aboral ectoderm activation (SpOtx), oral ectoderm repression (SpGsc), and endoderm repression (SpGATA-E).

Swaps of the ancestral and derived CBF and the proximal Otx/Gsc elements between *spec1* and *spec2a* RSRs suggested that the two derived elements acting together had acquired a repression function. This result is consistent with the hypothesis that SpGsc and SpGATA-E are acting as transcriptional repressors at the derived proximal Otx/Gsc element. However, this simple interpretation is complicated by the reduction in transcriptional activity observed with constructs containing an individual point mutation in either element. These results were not surprising because SpCBF and SpOtx are generally considered to function as transcriptional activators (Klein and Li, 1999; Maity and de Crombrugghe, 1998), but the increased transcriptional activity observed with the constructs containing both mutations suggests a complex interaction between these elements. Both elements are likely to function in a context-dependent manner rather than as autonomous 6-bp sequences. One possibility is that the CBF site stabilizes or promotes the oral ectoderm repression function from the OER and distal Otx/Gsc elements (Yuh et al., 2001). In this scenario, the loss of CBF function prevents oral ectoderm repression from these distal elements, a function that might be compensated by the proximal Otx/Gsc element. Similarly, functional compensation by the distal Otx/Gsc-OER elements may preserve the oral ectoderm repression activity lost by the point mutation in the proximal Otx/Gsc element. In both cases, the reduced enhancer activity may simply reflect the reduced binding of SpCBF and SpOtx. However, if both the proximal Otx/Gsc and CBF-mediated distal Otx/Gsc elements are lost, oral ectoderm repression is abrogated, resulting in an overall increase in transcriptional output. Previous studies have demonstrated that CCAAT elements can mediate transcriptional repression under particular cellular conditions, findings that are consistent with our proposed model (Wang et al., 1997). A realistic interpretation is that the *spec2a* RSR enhancer has been optimized over time as an integrated functional unit involving more than just the changes associated with the CBF and proximal Otx/Gsc elements. Indeed, many other base pair changes can be found when comparing the *spec2a* RSR sequence with other RSR sequences, and these nucleotide sequence differences might be as critical to RSR enhancer function as the differences found in the CBF and proximal Otx/Gsc elements.

Our results predict that reporter gene expression is repressed under the control of the proximal Otx/Gsc element in the presence of SpGsc and that the ancestral sequence does not have this repression activity. In addition, the proximal Otx/Gsc element would be expected to drive reporter genes specifically in aboral ectoderm (and mesenchymal cells), whereas the ancestral sequence would drive ubiquitous expression. However, attempts to perform these experiments were unsuccessful because the enhancer activity of the proximal Otx/Gsc element was too weak on its own for detecting reporter gene expression.

Time line for RSR enhancer evolution and derived cis-regulatory elements

Our interspecies sequence comparisons allowed us to trace the evolution of three genomic characteristics associated with *spec* genes and place them on a time line with respect to echinoid phylogeny (Fig. 10). First, the amplified RSR element seems to have originated in sea urchin genomes sometime after the divergence of *S. franciscanus* from *H. erythrogramma* 30–35 mya but before *S. franciscanus* diverged from *H. pulcherrimus* 13–18 mya. Second, we have recent evidence that *spec2a* orthologs appeared after the divergence of *S. franciscanus* from the other species in the Strongylocentrotidae branch, leading to *S. purpuratus* 13–18 mya. *S. franciscanus* probably has two or three *spec* genes, but these appear to be derived from a separate gene amplification event than the one that gave rise to the seven or eight member *spec* gene family found in the *S. purpuratus* genome (J. T. Villinski and W. H. Klein,

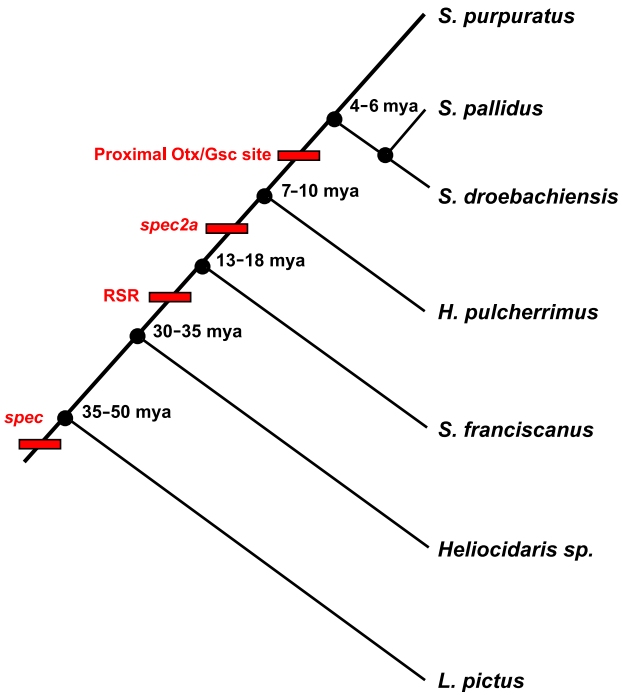

Fig. 10. Evolutionary time line for the origin of the RSR element, the *spec2a* gene, and the derived proximal Otx/Gsc element. Species divergence times from *S. purpuratus* are given at each node. The relative time points of *spec* gene origin (*spec*) RSR amplification, presence of a *spec2a* ortholog, and the G→A transition in the proximal Otx/Gsc element in the *spec2a* RSR are shown by red bars.

unpublished results). PCR analysis from the present study showed that in addition to *S. purpuratus*, *S. pallidus*, *H. pulcherrimus*, and *S. droebachiensis* all have a *spec2a* gene; the amplified fragments from genomic DNA of these species contain RSR sequences that cluster most closely with the *S. purpuratus spec2a* RSR, and the downstream primer used for PCR amplification was derived from the highly conserved *S. purpuratus spec2a* 5′ untranslated region. Third, the proximal Otx/Gsc element likely appeared after the divergence of *S. purpuratus* from *H. pulcherrimus*. In *S. purpuratus* and *S. pallidus*, the proximal Otx/Gsc sequence constitutes the major, if not only, allele in these species. We did not identify the proximal Otx/Gsc element in *S. droebachiensis*, which may indicate that this allele is present in lower frequency or has indeed been lost.

The *H. pulcherrimus* CBF site contains the derived A→T transversion, suggesting two possibilities. Either this change occurred before to the divergence of *H. pulcherrimus* and *S. purpuratus* or it arose convergently in the *H. pulcherrimus spec2a* RSR. The *H. pulcherrimus spec2a* CBF site (AATTGA), however, differs from *S. purpuratus* in one nucleotide (AATTGG). The function of the CBF site is undetermined in *H. pulcherrimus*, but this nucleotide may reduce the binding affinity of CBF. Indeed, a point mutation from AATTGG to AATTGC reduced the relative SpCBF-binding affinity 10-fold (Li et al., 2002). One allele contained an indel that disrupted the OER and CBF sites, indicating that these elements are not required for *spec2a* transcription in this haplotype. Together, these results suggest that the elements regulating *H. pulcherrimus spec2a* are different from those regulating *S. purpuratus spec2a*.

The inferred molecular phylogenies of several Strongylocentrotid species were recently reported (Biermann et al., 2003; Lee, 2003) and found to be as follows: ([*S. franciscanus*, *Strongylocentrotus nudas*], {*H. pulcherrimus* [*S. purpuratus*, (*S. droebachiensis*, *S. pallidus*)]}). Comparisons of the *spec2a* RSR sequences generated a gene tree in which the *S. purpuratus spec2a* RSR grouped more closely with that of *S. pallidus* than did the *S. droebachiensis spec2a* RSR. However, we cannot rule out that the identified RSR element from either *S. pallidus* or *S. droebachiensis* is not the *spec2a* RSR ortholog. It is possible that these RSRs are close relatives of *spec2a* RSRs, thus compromising our interpretation. Alternatively, the sequence divergence may reflect the evolution of functionally integrated compensatory *cis*-regulatory elements within the *S. droebachiensis spec2a* RSR. In this scenario, the allele containing the proximal Otx/Gsc site may have been present with some frequency in the ancestral species, but was retained at high frequency only in *S. purpuratus* and *S. pallidus*, and not in *S. droebachiensis*.

Mechanisms of cis-regulatory evolution under stabilizing selection

Recent observations have revealed an unappreciated dynamism in transcriptional regulatory processes, and several studies have demonstrated complete or nearly complete changes in *cis*-regulatory domains that are tolerated in the face of maintaining stable expression patterns (Galis et al., 2002; Ludwig et al., 2000; Scemama et al., 2002). A recent comparative study using the sea urchin *endo16* gene is particularly relevant (Romano and Wray, 2003). Extensive changes have evolved in the *cis*-regulatory elements driving the *endo16* gene in *S. purpuratus* and *Lytechinus variegatus*, although the endomesoderm expression pattern has been largely conserved. Reciprocal reporter gene studies indicate that changes in *cis*-elements and *trans* factors have evolved to regulate *endo16* expression. Sequence comparisons reveal that the proximal region of the regulatory domain, Module A, is conserved, whereas distal regions, containing several other critical modules in the *S. purpuratus endo16* regulatory region, are not alignable. These results suggest that some elements have been preserved but others have not, or at the very least, the *cis*-regulatory domain has been subjected to dramatic rearrangement (Eric Davidson, personal communication).

In conclusion, in sea urchin genomes, the regulatory regions upstream of *spec* genes have undergone two types of evolutionary change: a large-scale change was caused by an insertion into an ancestral genome of an RSR repetitive element, and optimization of the minimal RSR enhancer

activity by small-scale changes within the *spec2a* RSR, including two single-nucleotide mutations led to two new *cis*-regulatory elements. The results presented here add to a growing body of information illustrating the dynamic properties of transcriptional regulatory regions in bilaterian genomes during evolution.

Acknowledgments

We thank the other members of our laboratory for technical advice and ideas. The work was supported by a National Institute of Child Health and Development (NICHD) grant (HD22619) to WHK, and by the Robert A. Welch Foundation. SD and JTV were supported in part by an NICHD Pre- and Postdoctoral Training grant (HD072325). The University of Texas M. D. Anderson DNA Analysis Core Facility is supported in part by National Institute of Cancer grant CA16672.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2004.06.011.

References

Akasaka, K., Nishimura, A., Hijikata, K., Iuchi, Y., Morokuma, J., Takahashi, M., Morikawa, H., Shimada, H., 1995. Introduction of DNA into sea urchin eggs by particle gun. Mol. Mar. Biol. Biotechnol. 4, 255–261.

Angerer, L.M., Oleksyn, D.W., Levine, A.M., Li, X., Klein, W.H., Angerer, R.C., 2001. Sea urchin goosecoid function links fate specification along the animal–vegetal and oral–aboral axes. Development 128, 4393–4404.

Biermann, C.H., 1998. The molecular evolution of sperm bindin in six species of sea urchins (Echinoida: Strongylocentrotidae). Mol. Biol. Evol. 15, 1761–1771.

Biermann, C.H., Kessing, B.D., Palumbi, S.R., 2003. Phylogeny and development of marine model species: strongylocentrotid sea urchins. Evol. Dev. 5, 360–371.

Brandhorst, B.P., Klein, W.H., 1992. Territorial specification and control of gene expression in the sea urchin embryo. Semin. Dev. Biol. 3, 175–186.

Britten, R.J., Cetta, A., Davidson, E.H., 1978. The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. Cell 15, 1175–1186.

Carroll, S.B., Grenier, J.K., Weatherbee, S.D., 2001. From DNA to Diversity. Blackwell Science, Malden, MA.

Corpet, F., 1998. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16, 10881–10890.

Davidson, E.H., 2001. Genomic Regulatory Systems: Development and Evolution. Academic Press, San Diego, CA.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev. Biol. 246, 162–190.

Dubchak, I., Brudno, M., Loots, G.G., Mayor, C., Pachter, L., Ruben, E.M., Frazer, K.A., 2000. Active conservation of noncoding sequences revealed by 3-way species comparisons. Genome Res. 10, 1304–1306.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.

Galis, F., van Dooren, T.J., Metz, J.A., 2002. Conservation of the segmented germband stage: robustness or pleiotropy? Trends Genet. 18, 504–509.

Gan, L., Klein, W.H., 1993. A positive *cis*-regulatory element with a bicoid target site lies within the sea urchin *spec2a* enhancer. Dev. Biol. 157, 119–132.

Gan, L., Wessel, G.M., Klein, W.H., 1990a. Regulatory elements from the related *spec* genes of *Strongylocentrotus purpuratus* yield different spatial patterns with a lacZ reporter gene. Dev. Biol. 142, 346–359.

Gan, L., Zhang, W., Klein, W.H., 1990b. Repetitive DNA sequences linked to the sea urchin *spec* genes contain transcriptional enhancer-like elements. Dev. Biol. 139, 186–196.

Gan, L., Mao, C.A., Wikramanayake, A., Angerer, L.M., Angerer, R.C., Klein, W.H., 1995. An orthodenticle-related protein from *Strongylocentrotus purpuratus*. Dev. Biol. 167, 517–528.

Hardin, P.E., Klein, W.H., 1987. Unusual sequence conservation in the 5′ and 3′ untranslated regions of the sea urchin *spec* mRNAs. J. Mol. Evol. 25, 126–133.

Klein, W.H., Li, X., 1999. Function and evolution of Otx proteins. Biochem. Biophys. Res. Commun. 258, 229–233.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe, AZ, USA.

Kurita, M., Kondoh, H., Mitsunaga-Nakatsubo, K., Shimotori, T., Sakamoto, N., Yamamoto, T., Shimada, H., Takata, K., Akasaka, K., 2003. Utilization of a particle gun DNA introduction system for the analysis of *cis*-regulatory elements controlling the spatial expression pattern of the arylsulfatase gene (*HpArs*) in sea urchin embryos. Dev. Genes Evol. 213, 44–49.

Lee, Y.H., 2003. Molecular phylogenies and divergence times of sea urchin species of Strongylocentrotidae, Echinoida. Mol. Biol. Evol. 20, 1211–1221.

Letting, D.L., Chen, Y.Y., Rakowski, C., Reedy, S., Blobel, G.A., 2004. Context-dependent regulation of GATA-1 by friend-of-GATA-1. Proc. Natl. Acad. Sci. U. S. A. 101, 476–481.

Li, X., Chuang, C.K., Mao, C.A., Angerer, L.M., Klein, W.H., 1997. Two Otx proteins generated from multiple transcripts of a single gene in *Strongylocentrotus purpuratus*. Dev. Biol. 187, 253–266.

Li, X., Bhattacharya, C., Dayal, S., Maity, S., Klein, W.H., 2002. Ectoderm gene activation in sea urchin embryos mediated by the CCAAT-binding factor. Differentiation 70, 109–119.

Littlewood, D.T., Smith, A.B., 1995. A combined morphological and molecular phylogeny for sea urchins (Echinoidea: Echinodermata). Philos. Trans. R. Soc. Lond., B Biol. Sci. 347, 213–234.

Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403, 564–567.

Maity, S.N., de Crombrugghe, B., 1998. Role of the CCAAT-binding protein CBF/NF-Y in transcription. Trends Biochem. Sci. 23, 174–176.

Mao, C.A., Gan, L., Klein, W.H., 1994. Multiple Otx binding sites required for expression of the *Strongylocentrotus purpuratus spec2a* gene. Dev. Biol. 165, 229–242.

Mao, C.A., Wikramanayake, A.H., Gan, L., Chuang, C.K., Summers, R.G., Klein, W.H., 1996. Altering cell fates in sea urchin embryos by overexpressing SpOtx, an orthodenticle-related protein. Development 122, 1489–1498.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I., 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16, 1046–1047.

Ohta, T., 2003. Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. Genetica 118, 209–216.

Romano, L.A., Wray, G.A., 2003. Conservation of *Endo16* expression in sea urchins despite evolutionary divergence in both *cis* and *trans*-acting components of transcriptional regulation. Development 130, 4187–4199.

Sakamoto, N., Akasaka, K., Mitsunaga-Nakatsubo, K., Takata, K., Nishitani, T., Shimada, H., 1997. Two isoforms of orthodenticle-related proteins (HpOtx) bind to the enhancer element of sea urchin arylsulfatase gene. Dev. Biol. 181, 284–295.

Scemama, J.L., Hunter, M., McCallum, J., Prince, V., Stellwag, E., 2002. Evolutionary divergence of vertebrate Hoxb2 expression patterns and transcriptional regulatory loci. J. Exp. Zool. 294, 285–299.

Svensson, E.C., Huggins, G.S., Dardik, F.B., Polk, C.E., Leiden, J.M., 2000. A functionally conserved N-terminal domain of the friend of GATA-2 (FOG-2) protein represses GATA4-dependent transcription. J. Biol. Chem. 275, 20762–20769.

Takahashi, H., Mitani, Y., Satoh, G., Satoh, N., 1999. Evolutionary alterations of the minimal promoter for notochord-specific *Brachyury* expression in ascidian embryos. Development 126, 3725–3734.

Wang, Q., Zambetti, G.P., Suttle, D.P., 1997. Inhibition of DNA topoisomerase Hα gene expression by the p53 tumor suppressor. Mol. Cell. Biol. 17, 389–397.

Xiang, M.Q., Ge, T., Tomlinson, C.R., Klein, W.H., 1991. Structure and promoter activity of the LpS1 genes of *Lytechinus pictus*. Duplicated exons account for LpS1 proteins with eight calcium binding domains. J. Biol. Chem. 266, 10524–10533.

Yuh, C.H., Li, X., Davidson, E.H., Klein, W.H., 2001. Correct expression of *spec2a* in the sea urchin embryo requires both Otx and other *cis*-regulatory elements. Dev. Biol. 232, 424–438.