Development 130, 4187-4199

doi:10.1242/dev.00611

Conservation of *Endo16* expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation

Laura A. Romano and Gregory A. Wray*

Department of Biology, Duke University, Durham, NC 27708, USA
*Author for correspondence (e-mail: gwray@duke.edu)

Accepted 24 April 2003

SUMMARY

Evolutionary changes in transcriptional regulation undoubtedly play an important role in creating morphological diversity. However, there is little information about the evolutionary dynamics of cis-regulatory sequences. This study examines the functional consequence of evolutionary changes in the *Endo16* promoter of sea urchins. The *Endo16* gene encodes a large extracellular protein that is expressed in the endoderm and may play a role in cell adhesion. Its promoter has been characterized in exceptional detail in the purple sea urchin, *Strongylocentrotus purpuratus*. We have characterized the structure and function of the *Endo16* promoter from a second sea urchin species, *Lytechinus variegatus*. The *Endo16* promoter sequences have evolved in a strongly mosaic manner since these species diverged ~35 million years ago: the most proximal region (module A) is conserved, but the remaining modules (B-G) are unalignable. Despite extensive divergence in promoter sequences, the pattern of *Endo16* transcription is largely conserved during embryonic and larval development. Transient expression assays demonstrate that 2.2 kb of upstream sequence in either species is sufficient to drive GFP reporter expression that correctly mimics this pattern of *Endo16* transcription. Reciprocal cross-species transient expression assays imply that changes have also evolved in the set of transcription factors that interact with the *Endo16* promoter. Taken together, these results suggest that stabilizing selection on the transcriptional output may have operated to maintain a similar pattern of *Endo16* expression in *S. purpuratus* and *L. variegatus*, despite dramatic divergence in promoter sequence and mechanisms of transcriptional regulation.

Key words: Echinoderm, *Endo16*, Evolution, Promoter, Sea urchin, Transcription

INTRODUCTION

Comparative studies have revealed that the level, timing and spatial expression of genes is subject to change during evolution. In many instances, a change in gene expression has been correlated with a particular change in the phenotype of an organism at an anatomical, physiological or behavioral level (e.g. Dudareva et al., 1996; Sinha and Kellogg, 1996; Averof and Patel, 1997; Schulte et al., 1997; Stern, 1998; Hariri et al., 2002). However, few studies have examined the molecular mechanisms by which patterns of gene expression have evolved both within and between closely related species. Changes in transcriptional regulation undoubtedly play a central role in generating different patterns of gene expression (Raff, 1996; Doebley and Lukens, 1998; Wray and Lowe, 2000; Carroll et al., 2001; Davidson, 2001). Changes in promoter sequence or in the activity of transcription factors can alter gene expression, which may have functional consequences during development (e.g. Stockhaus et al., 1997; Singh et al., 1998). Many human polymorphisms in promoter sequences affect transcription and are correlated with phenotypic consequences (Rockman and Wray, 2002). Alternatively, changes in transcriptional regulation can serve to maintain patterns of gene expression over evolutionary time scales (Piano et al., 1999; Ludwig et al., 2000).

Studying the evolution of transcriptional regulation requires a system in which one or more promoter sequences have been characterized in detail using biochemical and functional approaches (Wray et al., 2003). Most importantly, this system must be amenable to functional analysis of promoter sequences in multiple, closely related species. To date, relatively few studies have analyzed the functional consequence of evolutionary changes in transcriptional regulation (Franks et al., 1988; Li and Noll, 1994; Ludwig et al., 1998; Ludwig et al., 2000; Shashikant et al., 1998; Singh et al., 1998; Crawford et al., 1999; Takahashi et al., 1999; Shaw et al., 2002; Tumpel et al., 2002). In this regard, sea urchins provide an outstanding system in which to study the evolution of transcriptional regulation. Eggs can be obtained in large quantities and develop synchronously upon fertilization, facilitating the collection of material for biochemical analyses. This has enabled researchers to characterize several promoter sequences in exceptional detail

including *CyIIIa* (Calzone et al., 1988; Theze et al., 1990; Wang et al., 1995; Kirchhamer and Davidson, 1996; Calzone et al., 1997; Coffman et al., 1996; Coffman et al., 1997) and *Endo16* (Yuh et al., 1994; Yuh et al., 1996; Yuh and Davidson, 1996; Yuh et al., 1998; Yuh et al., 2001a). Transient expression assays have proven remarkably successful for functional analysis of these promoter sequences in multiple species (reviewed by Kirchhamer et al., 1996). Moreover, the evolutionary history of sea urchins and other echinoderms is well characterized, allowing for interpretation of data in a phylogenetic context (Littlewood and Smith, 1995).

The *Endo16* gene was originally isolated from *Strongylocentrotus purpuratus* by screening a gastrula stage cDNA library (Nocente-McGrath et al., 1989). In *S. purpuratus*, *Endo16* is initially expressed throughout the vegetal plate of the hatched blastula (Nocente-McGrath et al., 1989; Ransick et al., 1993). *Endo16* expression is downregulated in primary mesenchymal cells (PMCs) as they migrate away from the center of the vegetal plate to form the larval skeleton. During gastrulation, *Endo16* is expressed throughout the invaginating archenteron. *Endo16* expression is then downregulated in secondary mesenchymal cells (SMCs) as they migrate away from the anterior tip of the archenteron to form various cell types, including pigment cells, muscle cells and coelomocytes. At the end of gastrulation, *Endo16* expression is downregulated in the anterior third of the archenteron, which corresponds to the prospective foregut, as well as the posterior third of the archenteron, which corresponds to the prospective hindgut. *Endo16* expression thereby becomes restricted to the midgut of the pluteus larva.

Transient expression assays demonstrated that 2.2 kb of sequence immediately upstream of the transcriptional start site is sufficient to drive *Endo16* expression (Yuh et al., 1994). Approximately 56 sites of specific DNA/protein interactions were mapped within this 2.2 kb region (Yuh et al., 1994) (Fig. 1A). These binding sites are clustered into six functionally distinct modules, which contribute in specific ways to the regulatory output of the *Endo16* promoter (Yuh et al., 1996; Yuh and Davidson, 1996) (Fig. 1B). The most proximal region of the promoter, module A, activates transcription in the vegetal plate and archenteron. Module B acts synergistically with module A to elevate levels of transcription in these regions. The activity of module A declines during gastrulation, and module B is responsible for maintaining *Endo16* expression in the midgut of the pluteus larva. The binding sites responsible for shifting the spatial control of *Endo16* expression to module B have been identified (Yuh et al., 2001a) (Fig. 1C). The most distal region of the promoter, module G, acts synergistically with modules A and B to increase the rate of transcription by ~4.2-fold throughout embryonic and larval development. Modules DC, E and F serve to confine *Endo16* expression to the endoderm: module DC represses transcription in PMCs, while modules E and F repress transcription in ectoderm adjacent to the vegetal plate. Finally, module A serves to communicate the integrated output of all modules to the basal promoter.

The biochemical and functional studies described above, when combined with the experimental advantages of sea urchins, creates an excellent opportunity to analyze promoter evolution. We have therefore characterized the *Endo16* promoter from a second sea urchin species, *Lytechinus variegatus*. Our results reveal a surprisingly strong dissociation between structure and function in this cis-regulatory system and provide insights into the evolutionary mechanisms that have operated on the *Endo16* promoter during the past 35 million years.

MATERIALS AND METHODS

Preparation of cultures

L. vareigatus adults were collected by Jennifer Keller at the Duke Marine Laboratory (Beaufort, NC) or Susan Decker (Hollywood, FL), and maintained in an aquarium at room temperature. *S. purpuratus* adults were obtained from Marinus (Long Beach, CA) or Charles Hollahan (Santa Barbara, CA), and maintained in an aquarium at 9°C. Gametes were obtained by injecting adults with 0.55 M KCl. Following fertilization, the eggs were cultured at room temperature (*L. variegatus*) or 9°C (*S. purpuratus*) in artificial seawater until the desired stages.

Isolation of full-length *LvEndo16* cDNA

RNA was isolated from gastrula-stage embryos using RNA STAT-60 (Tel-Test “B”, Friendswood, TX) and treated with DNase (Gibco BRL, Gaithersburg, MD). Reverse transcription (RT) was performed according to the instructions provided by the SuperScript Reverse Transcription kit (Gibco BRL). After the addition of a poly(A) tail, the cDNA was used to perform 5′ and 3′ RACE PCR. Primers were based on a partial cDNA sequence previously reported by Godin et al. (Godin et al., 1997) (GenBank Accession Number U89340). PCR products obtained by 5′ and 3′ RACE PCR were gel purified and ligated into pGEM-T vector (Promega, Madison, WI). Plasmid DNA was purified from transformed DH5α cells (Gibco BRL) and sequenced using an ABI Prism 3700 DNA Analyzer (PE Applied Biosystems, Foster City, CA). Sequences were assembled using Sequencher software (Gene Codes, Ann Arbor, MI).

Whole-mount in situ hybridization

Antisense and sense RNA probes were synthesized according to the instructions provided by the DIG RNA Labeling Kit (SP6/T7) (Roche, Indianapolis, IN) and stored in hybridization buffer (50 ng/μl) at –70°C. Sea urchin embryos were cultured to various stages of development and fixed for 2 hours in a solution containing 2.5% glutaraldehyde, 0.14 M NaCl and 0.2 M phosphate buffer, pH 7.4. The embryos were rinsed twice for ~15 minutes with buffer containing 0.3 M NaCl and 0.2 M phosphate buffer, pH 7.4, and dehydrated through 70% ethanol. Whole-mount in situ hybridization was performed using a protocol based on that of Zhu et al. (Zhu et al., 2001) with several modifications. One important modification was extending the incubation with PBST containing 5% sheep serum to ~16 hours at 4°C. Images were recorded using a SPOT camera (Diagnostic Instruments, Sterling Heights, MI).

Isolation of *LvEndo16* promoter and intron 1

Genomic DNA was isolated from sperm by phenol-chloroform extraction followed by ethanol precipitation. *LvEndo16* promoter sequence was obtained according to the instructions provided by the Universal GenomeWalker Kit (Clontech, Palo Alto, CA). In order to extend as far as 2.2 kb upstream of the transcriptional start site, three DNA walks were performed. Two rounds of amplification were performed for each DNA walk using nested primer pairs. Each promoter fragment was cloned and sequenced as described above. It is important to note that the promoter fragments overlapped by at least 50-100 bp. A 2337 bp sequence was assembled from overlapping fragments using Sequencher software. *LvEndo16* intron sequence was amplified by PCR using primers flanking the position at which the first intron was predicted to occur based on the *S. purpuratus* sequence

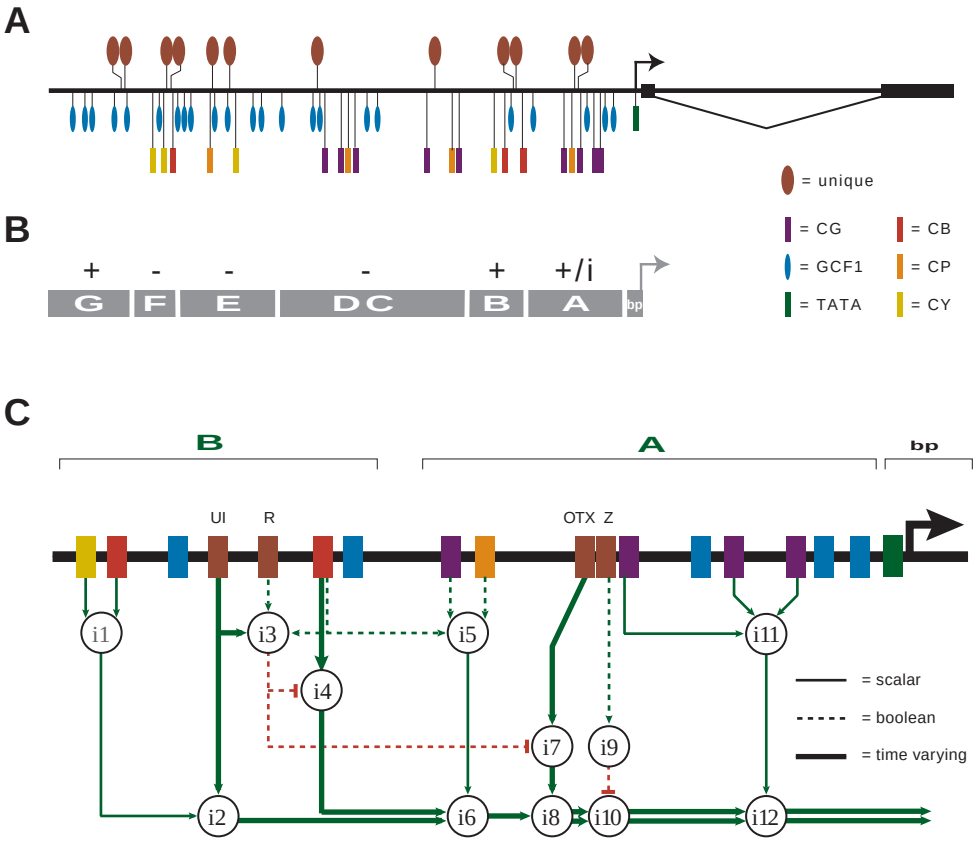

Fig. 1. Schematic representation of the *SpEndo16* promoter. (A) Relative position of 56 binding sites within the 2.2 kb region that has been shown to drive *SpEndo16* expression (Yuh et al., 1994). Twelve unique factors (brown ovals) each interact with only one binding site, six 'common factors' (colored rectangles) interact with a few identical (or nearly identical) binding sites, and the structural protein GCF1 (blue ovals) interacts with 23 sites in the *SpEndo16* promoter. [Figure adapted from Yuh and Davidson (Yuh and Davidson, 1996)]. (B) Binding sites within the *SpEndo16* promoter are clustered into six functionally distinct modules that serve to activate (+) or repress (–) transcription. (C) Logic circuit diagram showing interactions between binding sites within modules A and B of the *SpEndo16* promoter based on transient expression assays [Figure adapted from Yuh et al. (Yuh et al., 2001a). Note that binding sites in modules A and B interact extensively.

(GenBank Accession Number L34680). The sequence of the 5′ primer was 5′ AATGCGGAAGGAACTTTTTTGCTT and of the 3′ primer was 5′ GAAAGATCAAAGTCGGGAATCAT. The 468 bp product was cloned and sequenced as described above.

Sequences were aligned by ClustalX using default parameters (Thompson et al., 1997). This alignment was not significantly improved by reducing the gap penalty. Sequence similarity was calculated as the frequency of matching nucleotides for various regions of the *Endo16* locus, excluding indels (insertions and deletions). At the present time, there are no generally accepted measures of sequence similarity that incorporate indels. Seqcomp analyses were performed to detect a specified number of matching nucleotides (f) in a sliding window of size N in a manner similar to Sonnhammer and Durbin (Sonnhammer and Durbin, 1995). Empirical work by Yuh et al. (Yuh et al., 2002) supports the calculations by Brown et al. (Brown et al., 2002) showing that random matches are expected at or below a 0.7 threshold, but none above 0.75 for a 20 bp window. A seqcomp analysis of the *LvEndo16* and *SpEndo16* promoter sequences was performed at a threshold (f) of 0.8 and a window size (N) of 20 bp. Seqcomp analyses of the *LvEndo16* promoter sequence with BAC sequence from *S. purpuratus* (Sp127I21_S) and of the *SpEndo16* promoter sequence with BAC sequence from *L. variegatus* (Lv199M10_L) also were performed at a threshold (f) of 0.8 and a window size (N) of 100 bp. BAC sequences were obtained from the Sea Urchin Genome Project (http://sugp.caltech.edu:7000/resources/). Results of the seqcomp analyses were visualized on a dot plot and feature map using FamilyRelations (Brown et al., 2002). Similar results were obtained using identical parameters in the mVISTA program developed by Mayor et al. (Mayor et al., 2000) (not shown).

Microinjection

Endo16 promoter sequence was amplified by PCR as a single fragment (2,305 bp, *S. purpuratus;* 2,159 bp, *L. variegatus*) from genomic DNA using primers with restriction sites added to their 5′ ends in order to facilitate directional cloning. For *S. purpuratus*, the sequence of the 5′ primer was 5′ GCGCGAATTCGTCGGTGA-CCTAATTTCCCTTGTT, and of the 3′ primer was 5′ GCGCGG-ATCCCATCGTCTCAAAAATTAG. For *L. variegatus*, the sequence of the 5′ primer was 5′ GCGCGAATTCGAGCTTGTCAATGAGGG-TAATTTT and of the 3′ primer was GCGCGGATCCCGACCAAG-CAAAAAAGTTCC. The PCR products were cloned and sequenced as described above. The promoter fragments were excised from the pGEM-T vector (Promega) by restriction digestion with *Eco*RI and *Bam*HI, and ligated into digested pEGFP-1 vector (Clontech). The ligation products were cloned and sequenced as described above. Promoter constructs were verified by restriction digestions and sequencing using primers based on the pEGFP-1 sequence. Prior to microinjection, the *SpEndo16*-GFP and *LvEndo16*-GFP promoter constructs were linearized upstream of the promoter fragment with *Sac*I, and gel purified.

Eggs were de-jellied by incubating in artificial sea water, pH 5.0 for 3.5 minutes (*S. purpuratus*) or 1.5 minutes (*L. variegatus*). The eggs were then transferred to plastic petri dishes coated with protamine sulfate. *S. purpuratus* eggs were fertilized prior to microinjection in artificial sea water containing 0.2% PABA to prevent hardening of the fertilization envelope. Eggs were microinjected using a PLI-100 picospritzer (Medical Systems, Greenvale, NY) under an Axiovert S100 inverted microscope (Zeiss, Jena, Germany). Approximately 1500 molecules of linearized plasmid DNA were injected per egg in a 2 pl volume of solution containing a fivefold molar excess of *Hin*dIII-digested genomic DNA, as well as 0.12 M KCl and 30% glycerol. Following microinjection, the *L. variegatus* eggs were fertilized. Fertilized eggs were cultured at 9°C (*S. purpuratus*) or room temperature (*L. variegatus*) until the desired stages. Embryos and larvae were observed under a Axioskop MOT II

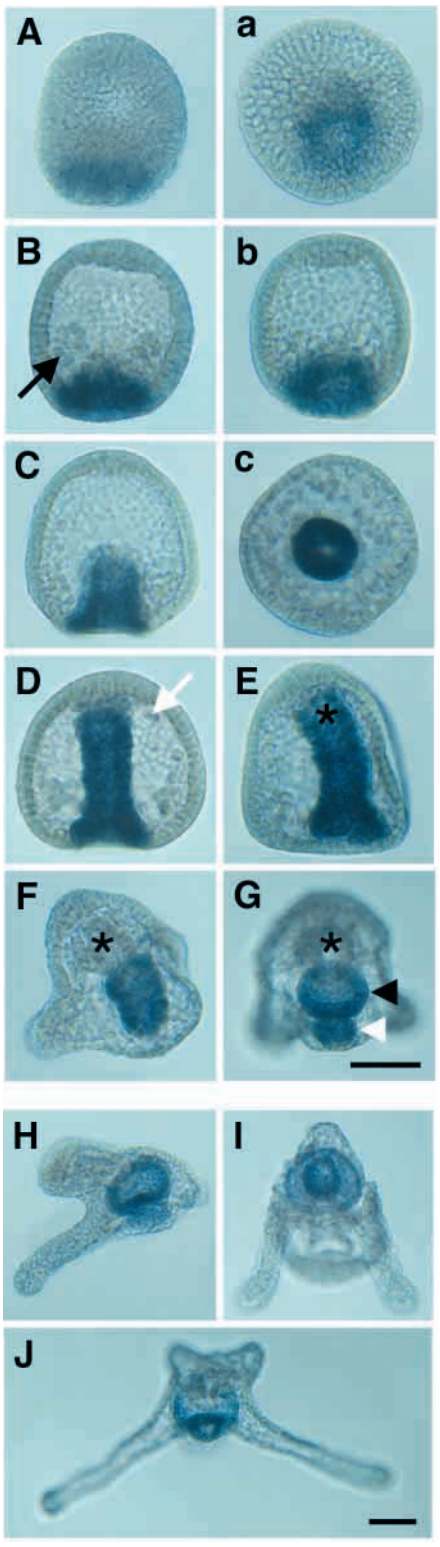

Fig. 2. Whole-mount in situ hybridization showing *LvEndo16* transcription. At the hatched blastula (A) and mesenchyme blastula (B) stages (lateral views), *LvEndo16* is expressed throughout the vegetal plate. Vegetal views (a,b) reveal that the PMCs (black arrow), which are derived from the center of the vegetal plate, do not express *LvEndo16*. As gastrulation proceeds, *LvEndo16* is expressed throughout the invaginating archenteron, as seen in lateral (C) and vegetal (c) views. Near the end of gastrulation, *LvEndo16* expression still extends throughout the archenteron (D). Expression is downregulated in SMCs (white arrow) as they ingress and migrate away from the tip of the archenteron. A lateral view (E) reveals that *LvEndo16* expression also is downregulated in the anterior third of the archenteron (prospective foregut, asterisk) as it bends to make contact with the oral ectoderm. *LvEndo16* continues to be expressed in the middle third (prospective midgut) and posterior third (prospective hindgut) of the archenteron. Lateral (F) and aboral (G) views show that *LvEndo16* expression is completely extinguished in the prospective foregut, but is maintained in the prospective midgut (black arrowhead) and hindgut (white arrowhead) at the prism stage. *LvEndo16* expression persists in both the midgut and hindgut of the pluteus larva until at least the four-arm stage (H-J). Scale bars: ~50 μm for A-G; 100 μm for H-J.

microscope (Zeiss) equipped for fluorescence microscopy. Images were recorded using a Hamamatsu digital camera (Model #C4742-95-12R) (Hamamatsu City, Japan) and analyzed using Openlab 2.2.4 (Improvision, Lexington, MA). *S. purpuratus* embryos were cultured at 9°C and therefore, developed more slowly than *L. variegatus* embryos; however, images were recorded at equivalent developmental stages for both species.

RESULTS

Characterization of *LvEndo16* expression by whole mount in situ hybridization

Full-length *LvEndo16* cDNA sequence was obtained by 5′ and 3′ RACE PCR using primers based on a partial cDNA sequence previously reported by Godin et al. (Godin et al., 1997). The full-length *LvEndo16* cDNA sequence is 4544 bp in length and encodes a protein that consists of 1485 amino acids (data not shown). Whole-mount in situ hybridization was performed using an antisense riboprobe corresponding to nucleotides 1-943 of the coding region. No expression was observed in embryos or pluteus larvae that were hybridized with the corresponding sense riboprobe as a negative control (data not shown).

LvEndo16 is initially expressed throughout the vegetal plate of the hatched blastula (Fig. 2A). *LvEndo16* expression is downregulated in PMCs as they ingress into the blastocoel (Fig. 2B). The PMCs lie at the center of the vegetal plate, so that *LvEndo16* expression appears as a ring when viewed from the vegetal pole (Fig. 2a,b). During gastrulation, *LvEndo16* is expressed throughout the invaginating archenteron (Fig. 2C), and continues to appear as a ring when viewed from the vegetal pole (Fig. 2c). *LvEndo16* expression is downregulated in SMCs as they migrate away from the anterior tip of the archenteron (Fig. 2D). *LvEndo16* expression thus remains restricted to the endoderm throughout gastrulation (Fig. 2C,D). This pattern of *Endo16* expression during embryonic development is conserved between *S. purpuratus* and *L. variegatus* (Fig. 3).

By the end of gastrulation, *LvEndo16* expression is downregulated in the anterior third of the archenteron, the prospective foregut (Fig. 2E). This decline of *LvEndo16* expression in the prospective foregut occurs as the archenteron bends to make contact with the oral ectoderm. *LvEndo16* continues to be expressed in the middle third of the archenteron, the prospective midgut (Fig. 2E). *LvEndo16* expression also continues to be expressed in the posterior third of the archenteron, the prospective hindgut (Fig. 2E). By the

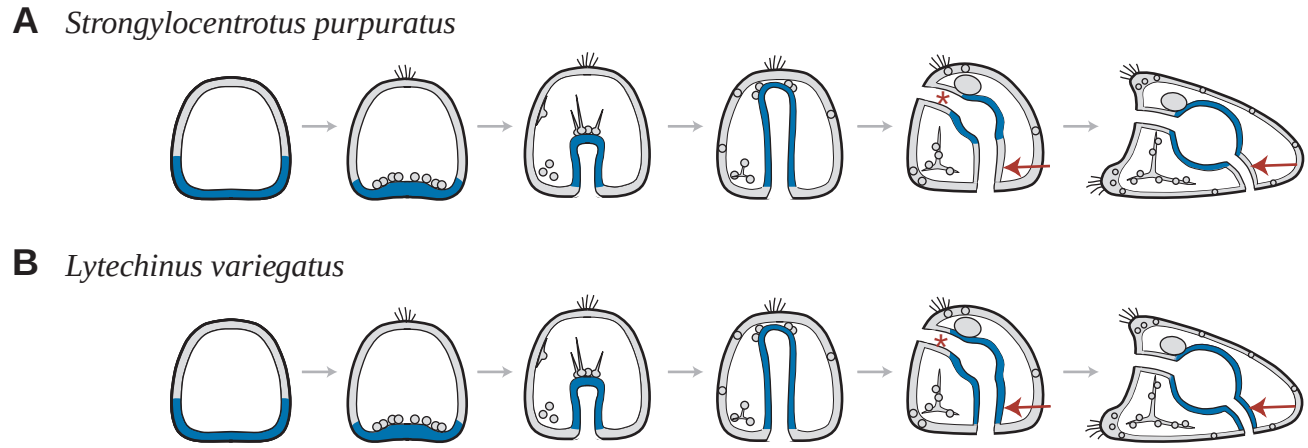

Fig. 3. Schematic comparison of *Endo16* transcription in *S. purpuratus* and *L. variegatus*. The pattern of *Endo16* expression (shown in blue) is relatively conserved between *S. purpuratus* (A) and *L. variegatus* (B). (Asterisks indicate prospective foregut.) However, *Endo16* expression is downregulated in the posterior third of the archenteron (prospective hindgut) only in *S. purpuratus*. (Arrows indicate hindgut.) *SpEndo16* expression persists in the midgut, while *LvEndo16* expression persists in both the midgut and hindgut of the pluteus larva.

time that the post-oral arms begin to extend from the pluteus larva, *LvEndo16* expression in the prospective foregut has completely disappeared (Fig. 2F,G). However, *LvEndo16* expression persists in both the midgut and hindgut of the pluteus larva until at least the four-arm stage (Fig. 2H-J). This persistent transcription in the hindgut constitutes a difference in the pattern of *Endo16* expression between *S. purpuratus* and *L. variegatus* during larval development (Fig. 3).

Characterization of the *LvEndo16* promoter

SpEndo16 expression can be driven by only 2.2 kb of sequence immediately upstream of the transcriptional start site (Yuh et al., 1994). In the present study, 2337 bp of *LvEndo16* sequence was assembled from overlapping fragments generated by a series of 'walks' upstream of the transcriptional start site (Fig. 4) (GenBank Accession Number AY292383). The *LvEndo16* promoter sequence then was amplified as a single fragment (~2.2 kb) that included the basal promoter, and cloned into the promoterless pEGFP-1 vector. The *LvEndo16* promoter sequence was inserted upstream of the EGFP gene to create a reporter construct referred to as *LvEndo16*-GFP.

Microinjection of *LvEndo16*-GFP into *L. variegatus* eggs drives GFP expression in a pattern that recapitulates the results of whole-mount in situ hybridization described above (Fig. 2). Fluorescence was consistently observed in a few cells located in the vegetal plate of the hatched blastula (Fig. 5A). These cells contributed to fluorescent patches within the invaginating archenteron (Fig. 5B). Fluorescence was maintained in the midgut of the pluteus larva until at least the four-arm stage (Fig. 5C,D). It is important to note that fluorescence also was observed in the hindgut (Fig. 5D), consistent with the fact that the endogenous gene is expressed in this region of the endoderm in *L. variegatus* but not *S. purpuratus* (Fig. 3). Ectopic fluorescence was rarely detected in the ectoderm, PMCs or SMCs. Furthermore, no fluorescence was detected upon microinjection of a promoterless construct containing the EGFP gene into *L. variegatus* eggs as a negative control. These results indicate that the 2.2 kb upstream fragment contains most or all of the *LvEndo16* promoter region.

Microinjection of DNA into sea urchin eggs produces mosaic expression (Arnone et al., 1997). In our hands, this method produced between one and six patches of fluorescent cells per embryo in which fluorescence was detected. We estimate that microinjection of *LvEndo16*-GFP into *L. variegatus* eggs produced fluorescence in ~10% of the resulting embryos. These numbers are smaller than those reported by Arnone et al. (Arnone et al., 1997) in their studies of the *sm50* and *cyIIa* genes in *S. purpuratus* perhaps because we used a different GFP vector to create fusion proteins. It is also possible that the efficiency of transient incorporation may differ between species. Because of the mosaic incorporation, it is difficult to quantitate the results of these experiments in terms of cell types expressing GFP. In contrast to CAT assays in which the level of transcription within a batch of embryos can be precisely measured, these experiments serve to define the spatial pattern of *LvEndo16* expression. In this regard, we focused on studying the spatial specificity of cis-regulatory elements, as has been carried out in several previous studies (e.g. Ludwig et al., 1998; Takahashi et al., 1999; Spitz et al., 2001; Tumpel et al., 2002; Yuh et al., 2001b). Future work using CAT reporter constructs will allow us to explore the kinetics of *LvEndo16* transcription as was done for the *SpEndo16* promoter after its initial characterization by Yuh et al. (Yuh et al., 1994).

Evolutionary analysis of the *Endo16* promoter

Alignment of the *Endo16* promoter sequences revealed that module A, the most proximal ~350 bp of the promoter, is well conserved between *S. purpuratus* and *L. variegatus* (Fig. 6). By contrast, upstream modules B through G are not conserved (sequence not shown). Although sequences upstream of module A were difficult to align, it is clear that modules B-G are significantly more divergent than module A. Specifically, module A contains only 11 indels (insertions and deletions), ranging from 1-5 bp in length, whereas the best alignment of modules B through G contains considerably more indels, ranging from 1 to 18 bp in length.

In order to further understand the significance of promoter

```
          CTGTCTT ATCTTCTGTC CTTCTTATTA CATAATATCA AATAATGTTA TATCAGAGTT
-2280 GACTGACTGT GACTGAACCT AACCATGGCC ATACGCAGAC TTCTTTTCAG GGGGGGGGCG
-2220 GGGGAGGACA CGTGAAAATT TTCAAGGAAA CTCGAAACTC GAGGGTGCGA AGCGCTCGAG
-2160 CTTGTCAATG AGGGTAATTT TATCCTTTCT AAAGTGGCAT TGAAGGATTT CGTGCACTTA
-2100 CTGAGAAAAT TTGGGGTAAA AAATGAAAGG TTTCAAAATT TCAAGGGCAA TTGCTCCTGC
-2040 CCCCCCCCCA CTGCGTTCGC TCTCTATCCA ATAATCAGTG CGTATCAATA GGAGGGAAGG
-1980 GGGCTCTTTC ATGCCTAATT CGATTGGGTG TCCTTGTAAG GGAAGTGAAA CGGCAAACTA
-1920 AAAAAATATT ATTAGCGATA TAGTAATGAT AATGATAACA TTCGAAACAT AATCAATATC
-1860 AGTAACTCAC TACTACTACT ACTACTACTA CTACTACTAC TACTACTACT ACTACTACTA
-1800 CTACTACTAC TACTACTACT ACTACTACTA CTACTACTAC TACTACTACT ACTACTACTA
-1740 TACTACTACT ACTACTACTA CTACTACTAT ACTACTACTA CTACTACTAC TACTACTATA
-1680 CTACTACTAC TACTACTACT ACTACTACTA CTACTACTAC TACTACTACT AATACTACTT
-1620 CTACTATTAC TGTAAAAGGA ATGAGGGGGG GGGGCAACTA ATGCGCATGC TCCGACGGGT
-1560 TTTGTATTCA AAACCTTTTT AGTTTCTTTG ATTATTATTA TTGTTATCAG TATAATATTA
-1500 TCGTTATTAT TATTATTATC ATTATTATTT ACTGTTATCA TTATTATTAT CATTATTATT
-1440 ATGATCATCA TTATAATCAT TATTATCACC TTTATCATTA TTTTATCATC ATCATTATTA
-1380 TTTACATAAA AAATTACAAG GTATTCTCCT TTGTAAGAAA AATCAAGGTG CTACACAACG
-1320 AGGCATCAAT CACATTTCTA GGGCAGAAAA GTAAACTGGT ATGTAATATC ATAAAATATA
-1260 GGAACAACAA TAACGAAATA GAAAATGGCA CTTTCAATTG CGTTTTTGAT TTTATATTCA
-1200 AAATGAAATC ATGAACAACT GAAAAAATGG CACACACTTT ATTATTTTCC AGCCGATTTC
-1140 ACTGAAATTA TTGTGTTTTG CGTGTTAGTT TGGCTCCAAC TATTCATTTG TATGATAAAA
-1080 ATGTGAACTC ATATCGCGGA GCAGTACTAC ATGTACGGTG TTATCTTCTT TGCTATGCTA
-1020 GGCTATAAGA TACAGGACTA TGCTTCTTTC TTTCCATCAT AAGCGTAGTT AAACTAAACA
 -960 CATAATTTCA GTGAAAAATA CCGAGATTTT AGTCCACGAA ACGCATTTTT TTTTATAAAT
 -900 GGCTTTTAGA ATGATGTTCG TTCTATGGAT AAATATTAAT GTTGATTAGA TTCCTAGTCT
 -840 TAAAAAGTCA AAAGCATGAT ACACTACGGT ACACAGCAAA AACTGTGGTG TTAACCGGTG
 -780 TACATAGAGG ACCACACCAG TTATTTTACA CCGGTATTAA ATTGGTGGTG TTAGTTTTAC
 -720 ATCTATAGGT TTTATTACAA CACCTTTGGT TGTTACATTT ACACTCTTTG GTGTTATGTT
 -660 CAATCTTTAG GGTGTAATAT TAACACCTCA GGGTGTGGTC CACTATTAGC ACCAACTGGT
 -600 GTCAGTTTGA ACACCACAGT TTTTCCAGTG TACCGATTAC AGCATATAGT AAACAGAATC
 -540 AATTGCTCAG TTTATGACTA AGAGGTTTTC TCGTTAATGG TGTGCCTTTA CTTAACTAAA
 -480 GTACAAAGAT TCGAAGTTGT TTGTGATTTC TCGGAGAGAT TAATATAAAC TGATCTTAGC
 -420 TGAAACTAAT CACTCTTTGT TTCTCAGAAA GATTTAAACT CGTAAACAAA AGCACCTTGT
 -360 ATGCGAACTG CTCATTACAA GTTCAGTATT GACAGAGACC GTATCGAATT AACATGCGAA

   module B ◄──► module A

 -300 AAAGGCTTAA TTTCCTCTTA AAGTACCTGT TTATTCCAAT AAATGTCTTT GTACAACTCA
 -240 AAACGCCAAC ATTCTCGAAG CAATAAAACT GTTGACCAAA GTGATGGCTC CATTATCATC
 -180 CCAAAGGATT AAGTGATCAA ACTACCACCA AATTATATCA CGTCAAAGTC ATCCCATCCC
 -120 CAGGTTCTAT TGTTTGAGTT TAGGCTTCTG AATGTTATAC CAAAGACAAA AGAGGTGTAA
  -60 CTTTGCCCCC CTTTGATTCG GAGCGGAGGG TTAAATAGAG TTAGACCGAC CGGGTTGGGT

      ──► transcription initiation      ──► translation initiation

   +1 CATAATATAT CAAATTTTAG AAAA[ATG]CGG AAGGAACTTT TTTGCTTGGT CGCCGTGCTG
      GCCGGGGCGC GGTCGAAGCC TAC
```

Fig. 4. *LvEndo16* promoter sequence. Shown here is the sequence from –2373 to +83 relative to the transcriptional start site. This sequence includes the promoter, the 5′ UTR, and the first exon; +83 is the position of the first intron. A microsatellite consisting of TAC repeats from –1632 to –1850 is underlined. The ATG start codon is boxed.

divergence, sequence similarity was calculated for various regions of the *Endo16* locus between the two species. Nucleotide identity within module A is 73%, which is comparable with nucleotide identity within the coding sequence. This indicates a similar level of functional constraint on the evolution of these two regions of the locus. As expected, nucleotide identity within binding sites (86%) is higher than within non-binding site nucleotides (69%) of module A. There is a decline in sequence similarity upstream of module A: 55% in module B, and less than 50% within modules DC-G. The first intron, which should be evolving neutrally due to the fact that it contains no functional binding sites (Yuh et al., 1994), has a sequence similarity of 54% (sequence not shown). Thus, modules B-G appear to be evolving neutrally as well.

Surprisingly, none of the binding sites identified within modules B through G of the *SpEndo16* promoter can be identified in the *LvEndo16* promoter, nor in the 5′ UTR, first intron, or coding sequence (Fig. 7A,B). It is important to bear in mind that more than one nucleotide can often fit the consensus sequence for a particular binding site. For example, the *SpEndo16* promoter contains multiple binding sites for GCF1 and CG. The sequences for many of these binding sites differ slightly within *S. purpuratus*, but still fall within a well-defined consensus sequence (Yuh et al., 1998). Several programs, including PipMaker (Schwartz et al., 2000), were employed to search for binding sites in the *LvEndo16* promoter. Other regions of the locus were also examined in both the 5′ and 3′ orientation, as there can be drastic changes in the order and spacing of binding sites during the evolution of cis-regulatory elements (Wray et al., 2003). It remains possible that variants of binding sites from modules B-G occur within the *LvEndo16* promoter, but if so, they have diverged considerably in sequence and perhaps relative position. In any case, such sites were not detected using algorithms to search for consensus sequences based on the *SpEndo16* promoter.

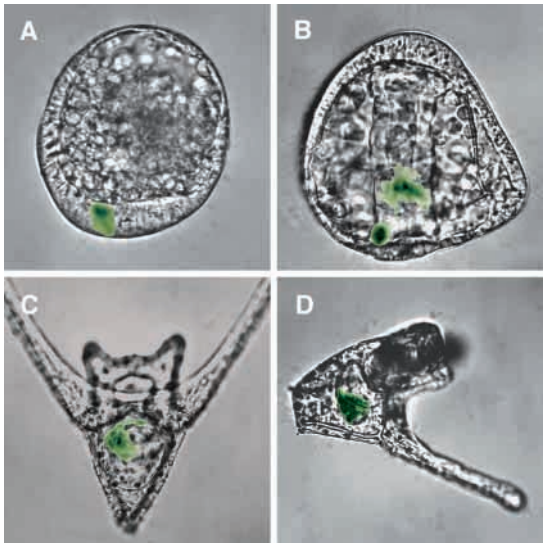

Fig. 5. Transient expression assays of the 2.2 kb upstream sequence injected into *L. variegatus* eggs. (A) Microinjection of a *LvEndo16*-GFP reporter construct resulted in fluorescence in the vegetal plate at the mesenchyme blastula stage. (B) During gastrulation, fluorescence is detected in the archenteron. (C) A ventral view showing fluorescence in the midgut of the pluteus larva. (D) A lateral view showing fluorescence in both the midgut and hindgut of the pluteus larva.

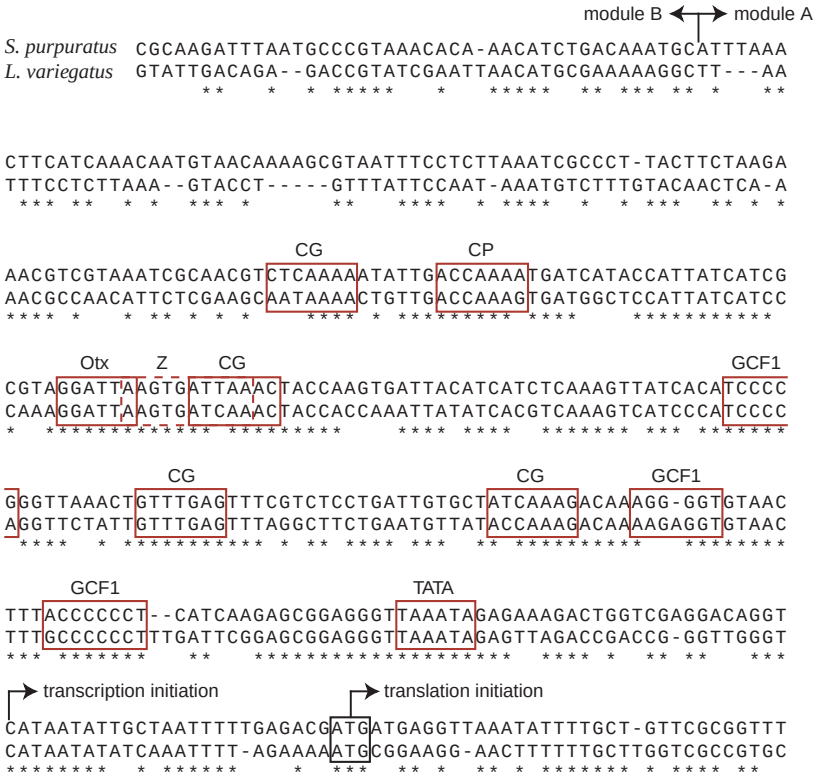

Fig. 6. Alignment of module A of the *Endo16* promoter from *S. purpuratus* and *L. variegatus*. Sequences extend upstream 335 bp and 345 bp relative to the transcriptional start site for *L. vareigatus* and *S. purpuratus*, respectively. (Asterisks indicate a nucleotide match.) Transcription factor binding sites identified in module A of the *SpEndo16* promoter are outlined by a red box. The Otx and Z binding sites occur only once within the *SpEndo16* promoter, although there are multiple binding sites for the proteins CG, CP and GCF1.

These findings are illustrated by a dot plot (Fig. 7C) and a series of feature maps (Fig. 7D-F) generated by FamilyRelations to visualize the results of a seqcomp analysis (Brown et al., 2002). Seqcomp is a relatively new program for comparative analyses that has been optimized for large sequences and can identify conserved sequences of a defined length without regard to spacing or orientation, a capability that is particularly important when examining non-coding regions. First, a pairwise comparison of the *SpEndo16* and *LvEndo16* promoter sequences was performed using a threshold of 0.8 and a window size of 20. In the case of the dot plot, the *LvEndo16* and *SpEndo16* promoter sequences are shown on the *x*- and *y*-axes, respectively, with regions of aligned sequence indicated as dots. Most of the dots occur in the upper, right corner of the graph, corresponding to module A of the *Endo16* promoter (Fig. 7C). In the feature map, the *SpEndo16* and *LvEndo16* promoter sequences are parallel with one another and red lines indicate regions of conservation. Most of the lines occur at the right end of the feature map, once again corresponding to module A of the *Endo16* promoter (Fig. 7D).

To test the possibility that modules B-G are separated from module A by a large insertion in the 5′ flanking region in *L. variegatus*, we compared the known *Endo16* promoter sequences with BAC sequences containing the *Endo16* locus. Modules B-G do not appear to be located further upstream of the isolated 2.2 kb sequence in *L. variegatus*, as evidenced by a pairwise comparison of the *SpEndo16* promoter sequence with a ~22 kb BAC sequence from *L. variegatus* that contains the *LvEndo16* locus. In this case, the analysis was performed using a threshold of 0.8 and a larger window size of 100 in order to avoid noise from repetitive elements. The feature map shows only one region of strong conservation that corresponds to module A of the *Endo16* promoter (Fig. 7E). The same parameters were applied to a pairwise comparison of the *LvEndo16* promoter sequence with a ~50 kb BAC sequence from *S. purpuratus* that contains the *SpEndo16* locus. In this case, the feature map shows two regions of conservation that correspond to module A of the *Endo16* promoter as well as a microsatellite consisting of TAC repeats (Fig. 7F).

Reciprocal injection of the *Endo16* promoter

To investigate whether there have been evolutionary changes in the set of transcription factors that bind to the *Endo16* promoter, reciprocal cross-species transient expression assays were performed. These experiments tested whether the *SpEndo16* promoter can drive correct expression in *L. variegatus* and whether the *LvEndo16* promoter can drive correct expression in *S. purpuratus*. *Endo16* promoter sequence from one species (donor) was microinjected into the egg of the other species (host), and GFP expression was observed in the resulting embryos and larvae by fluorescence microscopy. The pattern of GFP expression was interpreted in the context of the expression and sequence data obtained for each species, as well as data from microinjection of the *Endo16* promoter into eggs of the same species. As described above, microinjection of *LvEndo16*-GFP into *L. variegatus* eggs produced a pattern of GFP expression that recapitulated the results of in situ hybridization (Fig. 8J-L). Microinjection of *SpEndo16*-GFP into *S. purpuratus* eggs produced a nearly identical pattern of GFP expression; however, no fluorescence was observed in the hindgut (Fig. 8A-C). This latter result is consistent with studies by Yuh et al. (Yuh et al., 1994). No fluorescence was detected upon microinjection of a promoterless construct into eggs of either species as a negative control.

Microinjection of *SpEndo16*-GFP into *L. variegatus* eggs resulted in fluorescence in a few cells located in the vegetal plate of the hatched blastula (Fig. 8G). Patches of fluorescent cells were later observed in the invaginating archenteron (Fig. 8H), consistent with the pattern of *Endo16* expression as characterized by in situ hybridization in each species (Nocente-McGrath et al., 1989; Ransick et al., 1993). Fluorescence was maintained in the midgut of the pluteus larva until at least the four-arm stage (Fig. 8I). However, fluorescence was not observed in the hindgut, where *Endo16* is normally expressed in *L. variegatus* (Fig. 2H-J). Interestingly, fluorescence was consistently observed in SMCs during gastrulation (Fig. 8H).

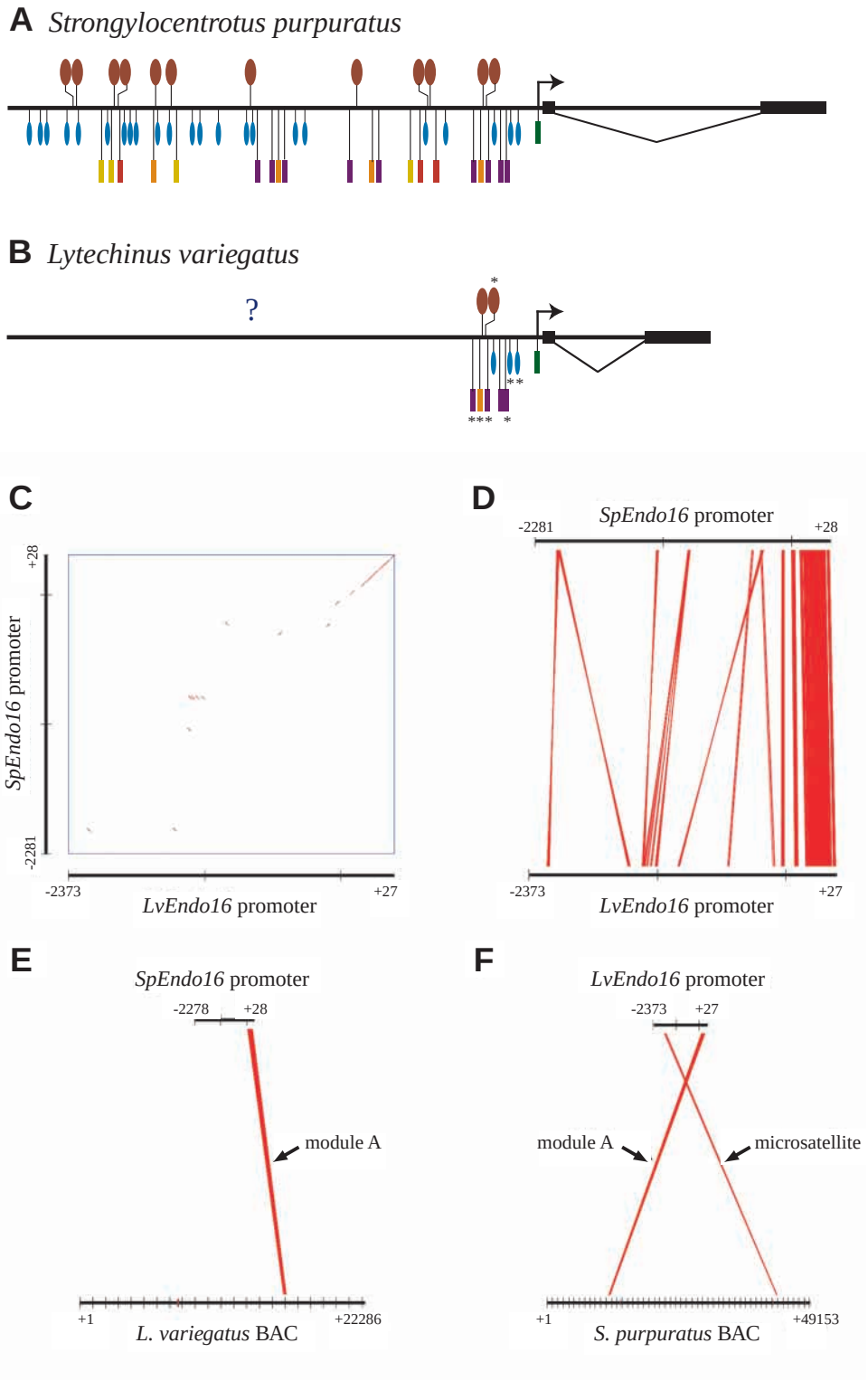

Fig. 7. Schematic representation of the *Endo16* promoter in *S. purpuratus* (A) and *L. variegatus* (B). The *LvEndo16* promoter sequence indicates only those binding sites identified in module A of *S. purpuratus*. Results from transient expression assays indicate that additional binding sites required for *LvEndo16* expression are likely to occur in the 2.2 kb region, but have not yet been identified. (An asterisk indicates that a nucleotide substitution or indel occurs within a binding site compared to the *Endo16* promoter sequence in *S. purpuratus*.) A dot plot (C) and feature maps (D-F) were generated by FamilyRelations based on a seqcomp analysis of the *Endo16* promoter (Brown et al., 2002). Alignment of the *SpEndo16* and *LvEndo16* promoter sequences is noted in the upper right corner of a dot plot (C), corresponding to module A. This is also evident at the right of a feature map (D). In neither case is there convincing evidence for sequence similarity upstream of module A. This result is supported by pairwise comparisons of the *Endo16* promoter sequence with BAC sequence from the opposite species. Only one region of conservation corresponding to module A is detected in a pairwise comparison of the *SpEndo16* promoter sequence and a BAC sequence from *L. variegatus* that contains the *LvEndo16* locus (E). The reciprocal analysis revealed two regions of conservation, corresponding to module A as well as a microsatellite consisting of TAC repeats (F).

At later stages of development, fluorescence was restricted to pigment cells (Fig. 8I), one of several cell types that are derived from SMCs (Gibson and Burke, 1985). Ectopic fluorescence was strictly confined to the pigment cells, with no fluorescence detected in the ectoderm, PMCs, or other SMC derivatives. It is important to note that microinjection of the *Endo16* promoter into eggs of the same species did not produce ectopic fluorescence in the SMCs or any other cell type.

Microinjection of *LvEndo16*-GFP into *S. purpuratus* eggs resulted in a pattern of GFP expression similar to that observed in the reciprocal experiment. Fluorescence was observed in the vegetal plate of the hatched blastula, and later in the invaginating archenteron (Fig. 8D,E). In addition, fluorescence was observed in the midgut of the pluteus larva until at least the four-arm stage (Fig. 8F). Fluorescence was not observed in the hindgut, consistent with the endogenous pattern of

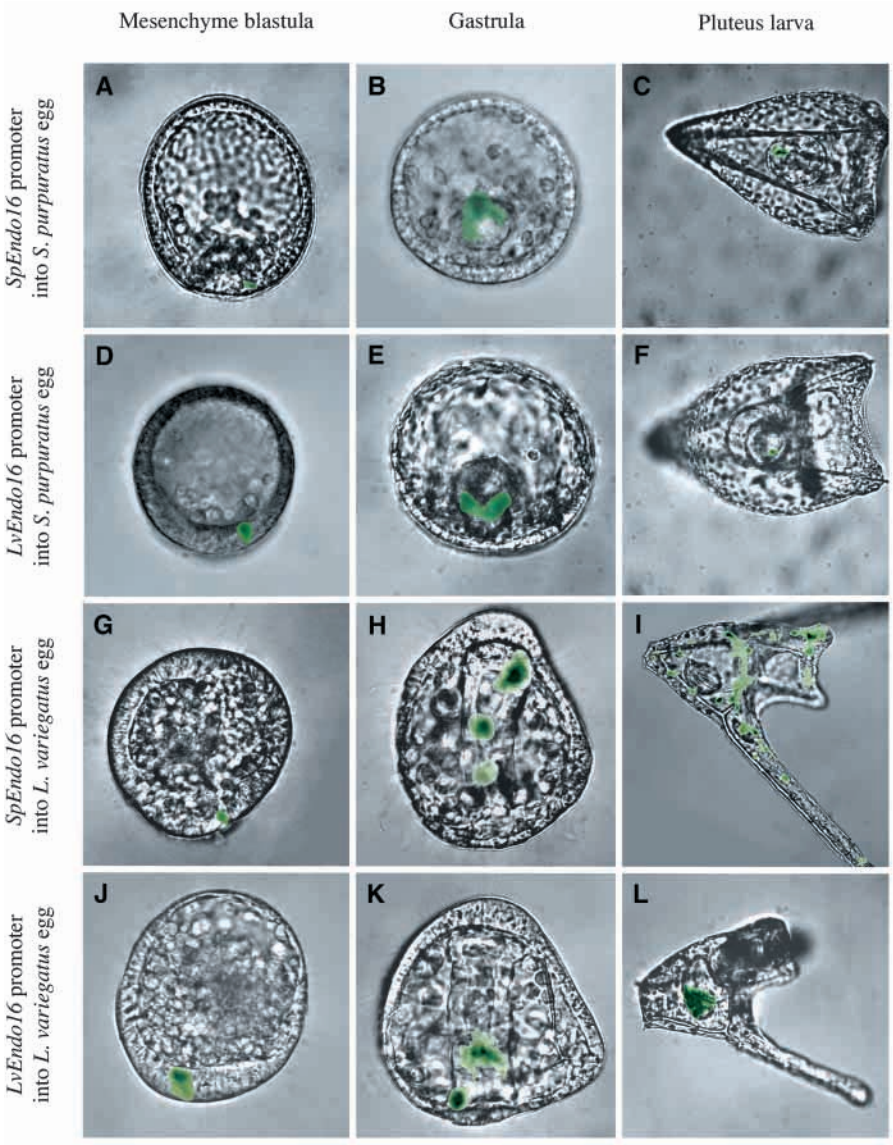

Fig. 8. Reciprocal cross-species transient expression assays using the *Endo16* promoter. GFP reporter constructs were microinjected in a reciprocal cross-species experimental design. Images were captured at three stages of development: mesenchyme blastula (A,D,G,J), gastrula (B,E,H,K), and pluteus larva (C,F,I,L). Microinjection of *SpEndo16*-GFP into *S. purpuratus* eggs results in a pattern of GFP expression that recapitulates the results of in situ hybridization of the endogenous gene (Nocente-McGrath et al., 1989; Ransick et al., 1993), and as observed by Yuh et al. (Yuh et al., 1994) in transient expression assays (A-C). Microinjection of *LvEndo16*-GFP into *S. purpuratus* eggs results in the same pattern of GFP expression (D-F). Note that it does not drive GFP expression in the hindgut of the pluteus larva (F). Microinjection of *SpEndo16*-GFP into *L. variegatus* eggs produces ectopic fluorescence in the SMCs as well their pigment cell derivatives (G-I). As in the reciprocal experiment, no fluorescence is detected in the hindgut of the pluteus larva (I). Microinjection of *LvEndo16*-GFP into *L. variegatus* eggs results in a pattern of GFP expression (J-L) that recapitulates the results of in situ hybridization of the endogenous gene as shown in Fig. 2. Fluorescence persists in both the midgut and hindgut of the pluteus larva (L).

SpEndo16 expression. Unlike the reciprocal experiment, ectopic fluorescence was not observed in the SMCs or any other cell type. These data are summarized in Fig. 9.

DISCUSSION

Our analysis of the *Endo16* promoter reveals an unexpectedly complex evolutionary dynamic. Capitalizing on detailed biochemical and functional analyses of the *Endo16* promoter in the purple sea urchin, *S. purpuratus* (Yuh et al., 1994; Yuh et al., 1996; Yuh and Davidson, 1996; Yuh et al., 1998; Yuh et al., 2001a), we have analyzed the structure and function of this promoter in a second sea urchin species, *L. variegatus*. The *LvEndo16* cDNA sequence encodes a large 4.6 kb protein with several motifs, suggesting a role in cell adhesion (Soltysik-Espanola et al., 1994). Indeed, experiments using antisense morpholinos indicate that Endo16 may be required for the dynamic changes in cell adhesion that occur during gut morphogenesis (L.A.R. and G.A.W., unpublished). Remarkably, the *Endo16* promoter displays a mosaic pattern of evolution, with only module A being conserved between the two species. Reciprocal cross-species transient expression assays indicate that the set of transcription factors that bind to the *Endo16* promoter has also diverged to some extent. Nonetheless, *LvEndo16* is expressed in a pattern similar to that observed in *S. purpuratus*, suggesting that stabilizing selection has acted on the transcriptional output of the *Endo16* promoter throughout the past 35 million years.

Evolutionary changes in the *Endo16* promoter

Yuh et al. (Yuh et al., 1994) have demonstrated that *Endo16* expression is regulated by 2.2 kb of sequence immediately upstream of the transcriptional start site. This sequence contains at least 56 transcription factor binding sites that are clustered into six functionally distinct modules that regulate the level, timing and spatial transcription of *Endo16* in *S. purpuratus*. We have shown that 2.2 kb of sequence immediately upstream of the transcriptional start site is sufficient to drive *Endo16* expression throughout embryonic and larval development in *L. variegatus* as well. Although the pattern of *Endo16* expression is similar between *S. purpuratus* and *L. variegatus* (Fig. 3), our data demonstrate that drastic changes have evolved in the *Endo16* promoter since these two species diverged. Of the entire *Endo16* promoter, only the most proximal region, module A, is conserved between the two species (Fig. 7).

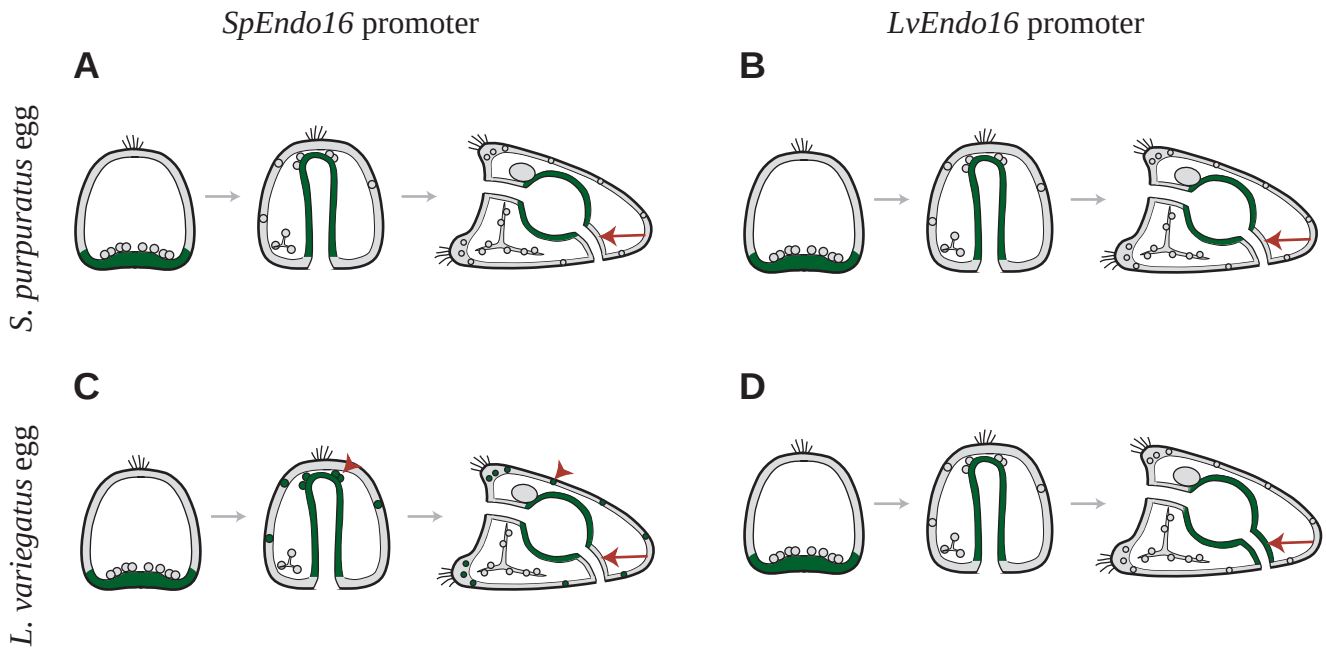

Fig. 9. Summary of reciprocal cross-species transient expression assays using the *Endo16* promoter. Microinjection of the *Endo16* promoter into eggs of the same species results in a pattern of GFP expression (green) that recapitulates the results of in situ hybridization (A,D). Microinjection of *LvEndo16*-GFP into *S. purpuratus* eggs results in a host-specific pattern of GFP expression (B), while microinjection of *SpEndo16*-GFP into *L. variegatus* eggs results in a donor-specific pattern of GFP expression with ectopic fluorescence in the SMCs and their pigment cell derivatives (C). These data indicate that evolutionary changes have arisen both cis and trans to the *Endo16* gene. (Arrows indicate hindgut. Arrowheads indicate ectopic fluorescence in the SMCs and pigment cells.)

These results indicate that different regions within the *Endo16* promoter are under different levels of functional constraint. Specifically, module A appears to be under a much higher level of functional constraint than the rest of the promoter. It is not surprising that certain modules of the *Endo16* promoter are more conserved than others because they perform different functions. Modularity in cis-regulatory sequences allows changes in gene expression to evolve in one tissue independently of another, and has been proposed to facilitate the evolution of morphological diversity (Kitchhamer et al., 1996; Gerhart and Kirschner, 1998; Carroll et al., 2001). Within the *Endo16* promoter, the conservation of module A makes functional sense given its essential roles in relaying the integrated output of all modules to the basal promoter and serving as the primary activator of *Endo16* expression during embryogenesis (Yuh et al., 1998). Nucleotides within binding sites are more conserved than those not in binding sites presumably because they are directly responsible for activating *Endo16* expression. This pattern of functional constraint on binding sites versus non-binding sites has been noted for a few genes (e.g. Core et al., 1997). It is likely that negative selection has maintained functionally important binding sites within module A of the *Endo16* promoter since *S. purpuratus* and *L. variegatus* last shared a common ancestor.

Functional conservation of the *Endo16* promoter

The pattern of *Endo16* expression is similar in *S. purpuratus* and *L. variegatus* despite the fact that only module A of the *Endo16* promoter is conserved. It has been postulated that selection for compensatory mutations is a primary mechanism by which patterns of gene expression are conserved for long periods of evolutionary time (Ludwig et al., 2000). Several studies provide support for this idea (e.g. Ludwig and Kreitman, 1995; Maduro and Pilgrim, 1996; Tamarina et al., 1997; Ludwig et al., 1998; Piano et al., 1999; Takahashi et al., 1999; Ludwig et al., 2000; Tumpel et al., 2002). Functional compensation appears to have also evolved within the *Endo16* promoter, although the changes are more extensive than in any of these previously known cases.

Several pieces of evidence are relevant to understanding the genetic basis for conservation of function despite such divergence in sequences. Yuh and Davidson (Yuh and Davidson, 1996) demonstrated that microinjection of a GFP reporter construct containing only module A drives GFP expression in the vegetal plate and archenteron, but is not sufficient to maintain expression in the midgut of the pluteus larva in *S. purpuratus* (Yuh and Davidson, 1996). Despite the fact that only module A is conserved, the 2.2 kb region immediately upstream of the transcriptional start site of the *LvEndo16* gene is sufficient to drive later phases of *LvEndo16* expression. It is possible that module A is entirely responsible for the pattern of *LvEndo16* expression, although this seems unlikely given its inability to drive larval expression in *S. purpuratus*. It is also possible that binding sites could not be identified upstream of module A within the *LvEndo16* promoter because of unrecognized variation in their consensus sequences. Alternatively, the remaining region of the 2.2 kb region of the *LvEndo16* promoter may contain binding sites for a different set of transcription factors that are functionally equivalent to those in modules B-G of the *SpEndo16* promoter. That is, during the evolution of the *Endo16* promoter, some

binding sites may have been replaced by others that generate a similar pattern of *Endo16* expression. The transcription factors that interact with the *Endo16* promoter may have co-evolved to maintain this pattern of *Endo16* expression, as has been documented for the *bicoid* promoter in insects (Shaw et al., 2002). In any case, the *SpEndo16* and *LvEndo16* promoter sequences are very different, yet generate a similar pattern of *Endo16* expression. Although this situation suggests the operation of stabilizing selection, we cannot rule out the possibility that drift or directional selection have been important contributors until data are obtained for additional species.

Divergence in the pattern of *Endo16* expression

Although the pattern of *Endo16* expression is generally conserved, transcription persists only in the midgut of the pluteus larva in *S. purpuratus* (Nocente-McGrath et al., 1989; Ransick et al., 1993), but in both the midgut and hindgut of the pluteus larva in *L. variegatus*. This difference in transcriptional regulation may have evolved in several different ways. The *SpEndo16* and *LvEndo16* promoters may contain binding sites for different transcription factors involved in segmentation of the tripartite gut. Alternatively, the expression and/or activity of these transcription factors may be different between the two species. For example, the transcription factor UI binds within module B of the *SpEndo16* promoter, and is directly responsible for maintaining *SpEndo16* expression in the midgut of the pluteus larva (Yuh et al., 1998). Although a binding site for the transcription factor UI could not be identified within the *LvEndo16* promoter, it is possible that *LvEndo16* expression persists in the hindgut due to expansion of the spatial domain of UI expression in *L. variegatus*. Another possibility is the existence of a transcription factor that represses *Endo16* expression, and is expressed in the hindgut of *S. purpuratus* but not *L. variegatus*.

Evolutionary changes in transcription factors that bind to the *Endo16* promoter

Binding sites within modules B-G of the *SpEndo16* promoter do not appear to be present in any region of the *LvEndo16* locus including the 2.2 kb region that was shown to drive the correct pattern of GFP expression (Fig. 7). This result suggests that *Endo16* expression is regulated, at least in part, by a different set of transcription factors in *S. purpuratus* and *L. variegatus*. Indeed, reciprocal injection of the *Endo16* promoter between the two species revealed differences in the expression and/or activity of transcription factors that bind to the *Endo16* promoter.

Microinjection of *SpEndo16*-GFP into *L. variegatus* eggs, as well as microinjection of *LvEndo16*-GFP into *S. purpuratus* eggs, produced fluorescence in the vegetal plate and archenteron (Fig. 9B,D). This result is consistent with the fact that module A is responsible for activating *Endo16* expression in these regions (Yuh et al., 1996; Yuh and Davidson, 1996). Moreover, this most proximal region of the *Endo16* promoter is conserved between *S. purpuratus* and *L. variegatus*. A few nucleotide substitutions and indels occur within known transcription factor binding sites of module A (Fig. 6). Some of these changes occur within multiply represented binding sites for the 'structural' protein GCF1, which stabilizes DNA looping (Zeller et al., 1995). However, a few changes occur within binding sites for proteins with a regulatory function. These changes may have been tolerated because they have little or no effect on DNA/protein interactions, a possibility that can be tested with mobility shift assays.

Reciprocal injection also produced fluorescence in the midgut of the pluteus larva (Fig. 9B,D). Yet, module B, which was shown to maintain *SpEndo16* expression in this region of endoderm (Yuh et al., 1998), is not present in *L. variegatus*. Thus, it appears as if changes have evolved within the *Endo16* promoter to maintain the regulatory output of module B even in the absence of any obvious sequence similarity. Interestingly, the fact that the *SpEndo16* promoter correctly drives GFP expression in the midgut of *L. variegatus* indicates that the appropriate transcription factors are expressed in both species in a conserved manner. If this were not the case, GFP reporter expression would not mimic the expression of the endogenous gene in reciprocal cross-species microinjection experiments. For example, microinjection of the CyIIIa promoter from *S. purpuratus* into *L. variegatus* eggs resulted in ectopic CAT activity in several cell types (Franks et al., 1988).

Fluorescence was not detected in the hindgut upon microinjection of *SpEndo16*-GFP into *L. variegatus* eggs (Fig. 9C). Microinjection of *LvEndo16*-GFP into *S. purpuratus* eggs also failed to produce fluorescence in the hindgut, despite the fact that *LvEndo16* is expressed in this region of endoderm (Fig. 9B). Either the appropriate transcription factors are not present in this region of *S. purpuratus*, or there has been a change in the activity of co-factors that are required for these transcription factors to bind to the *LvEndo16* promoter.

Interestingly, microinjection of *SpEndo16*-GFP into *L. variegatus* consistently produced ectopic fluorescence in the SMCs and their descendents, the pigment cells (Fig. 9C). By contrast, microinjection of *LvEndo16*-GFP into *S. purpuratus* did not produce ectopic fluorescence (Fig. 9B). These data suggest that *L. variegatus* and *S. purpuratus* use different mechanisms to repress *Endo16* expression in the SMCs. The transcription factors that normally repress *SpEndo16* expression in the SMCs may not be present in *L. variegatus*. However, any transcription factors that normally repress *LvEndo16* expression in the SMCs must be present in *S. purpuratus*. Alternatively, it is possible that there are no binding sites within the *LvEndo16* promoter capable of activating *LvEndo16* expression in the SMCs and other non-endodermal cell types.

Thus, it appears as though compensatory changes have evolved that lie both cis and trans to the *Endo16* gene. Only a few studies have analyzed promoter sequences in the context of another species to determine the extent to which the corresponding transcription factors have co-evolved (Klueg et al., 1997; Takahashi et al., 1999; Shaw et al., 2002). For example, Takahashi et al. (Takahashi et al., 1999) performed reciprocal injections of the *brachyury* promoter in two species of ascidians, *Ciona intestinalis* and *Halocynthia roretzi*. Extensive changes have evolved in the *brachyury* promoter, although it activates notochord-specific expression in both species (Corbo et al., 1997; Takahashi et al., 1999). Microinjection of the *C. intestinalis* brachyury promoter into *H. roretzi* eggs produced ectopic *lacZ* expression in other mesodermally derived tissues, suggesting that there have also been alterations in the set of transcription factors that bind to the *brachyury* promoter. Most other studies carried out

unidirectional analysis of promoter sequences in the context of another species (e.g. Franks et al., 1988; Ludwig et al., 1998; Ludwig et al., 2000; Shashikant et al., 1998), and may therefore have missed finding evidence for trans components to changes in transcriptional regulation.

In summary, this study combines expression, sequence and functional data to analyze changes in cis-regulatory sequences that influence transcription. Data from additional species of sea urchins will help provide a more complete understanding of how changes in transcriptional regulation relate to the evolution of morphological diversity. In addition, site-directed mutagenesis and biochemical assays will allow us to test the functional consequences of specific nucleotide substitutions and indels on *Endo16* expression both within and between closely related species.

We thank Cyndi Bradham (Duke University) and members of the Wray laboratory (Jim Balhoff, Chisato Kitazawa, Ann Klatt, Margaret Pizer and Matt Rockman) for their insightful comments on a draft of this manuscript. We are very grateful to Eric Davidson, Andy Cameron and Cathy Yuh (CalTech) for their helpful advice, and for providing us with unpublished data. Finally, we are very grateful to Titus Brown (CalTech) for assisting us with the seqcomp analysis, and Hyla Sweet (Carnegie Mellon University) for advice regarding in situ hybridization. This work was supported by NASA grant NAG-2-1377 and NSF grant IBN-96346 awarded to Gregory Wray.

REFERENCES

Arnone, M. I., Bogarad, L. D., Collazo, A., Kirchhamer, C. V., Cameron, R. A., Rast, J. P., Gregorians, A. and Davidson, E. H. (1997). Green fluorescent protein in the sea urchin: new experimental approaches to transcriptional regulatory analysis in embryos and larvae. *Development* **124**, 4649-4659.

Averof, M. and Patel, N. H. (1997). Crustacean appendage evolution associated with changes in *Hox* gene expression. *Nature* **388**, 682-686.

Brown, C. T., Rust, A. G., Clarke, P. J., Pan, Z., Schilstra, M. J., de Buysscher, T., Griffin, G., Wold, B. J., Cameron, R. A., Davidson, E. H. and Bolouri, H. (2002). New computational approaches for analysis of *cis*-regulatory networks. *Dev. Biol.* **246**, 86-102.

Calzone, F. J., Theze, N., Thiebaud, P., Hill, R. L., Britten, R. J. and Davidson, E. H. (1988). Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes Dev.* **2**, 1074-1088.

Calzone, F. J., Grainger, J., Coffman, J. A. and Davidson, E. H. (1997). Extensive maternal representation of DNA-binding proteins that interact with regulatory target sites of the *Strongylocentrotus purpuratus CyIIIa* gene. *Mol. Mar. Biol. Biotechnol.* **6**, 79-83.

Carroll, S. B., Grenier, J. K. and Weatherbee, S. D. (2001). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. London: Blackwell Science.

Coffman, J. A., Kirchhamer, C. V., Harrington, M. G. and Davidson, E. H. (1996). SpRunt-1, a new member of the runt domain family of transcription factors, is a positive regulator of the aboral ectoderm-specific *CyIIIa* gene in sea urchin embryos. *Dev. Biol.* **174**, 43-54.

Coffman, J. A., Kirchhamer, C. V., Harrington, M. G. and Davidson, E. H. (1997). SpMyb functions as an intramodular repressor to regulate spatial expression of *CyIIIa* in sea urchin embryos. *Development* **124**, 4717-4727.

Corbo, J. C., Levine, M. and Zeller, R. W. (1997). Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589-602.

Core, N., Charroux, B., McCormick, A., Vola, C., Fasano, L., Scott, M. P. and Kerridge, S. (1997). Transcriptional regulation of the *Drosophila* homeotic gene *teashirt* by the homeodomain protein Fushi tarazu. *Mech. Dev.* **68**, 157-172.

Crawford, D. L., Segal, J. A. and Barnett, J. L. (1999). Evolutionary analysis of TATA-less proximal promoter function. *Mol. Biol. Evol.* **16**, 194-207.

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

Doebley, J. and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**, 1075-1082.

Dudareva, N., Cseke, L., Blanc, V. M. and Pichersky, E. (1996). Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell* **8**, 1137-1148.

Franks, R. R., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1988). Spatially deranged though temporally correct expression of *Strongylocentrotus purpuratus* actin gene fusion in transgenic embryos of a different sea urchin family. *Genes Dev.* **2**, 1-12.

Gerhart, J. and Kirschner, M. (1997). *Cells, Embryos, and Evolution*. Malden, MA: Blackwell Science.

Gibson, A. W. and Burke, R. D. (1985). The origin of pigment cells in embryos of the sea urchin Strongylocentrotus purpuratus. *Dev. Biol.* **107**, 414-419.

Godin, R. E., Klinzing, D. C., Porcaro, W. A. and Ernst, S. G. (1997) Specification of endoderm in the sea urchin embryo. *Mech. Dev.* **67**, 35-47.

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M. F. and Weinberger, D. R. (2002) Serotonin transporter genetic variation and the response of the human amygdala. *Science* **297**, 400-403.

Kirchhamer, C. V. and Davidson, E. H. (1996) Spatial and temporal information processing in the sea urchin embryo: modular and intramodular organization of the *CyIIIa* gene *cis*-regulatory system. *Development* **122**, 333-348.

Kirchhamer, C. V., Yuh, C. H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc. Natl. Acad. Sci. USA* **93**, 9322-9328.

Klueg, K. M., Harkey, M. A. and Raff, R. A. (1997). Mechanisms of evolutionary changes in timing, spatial expression, and mRNA processing in the *msp130* gene in a direct-developing sea urchin, *Heliocidaris erythrogramma*. *Dev. Biol.* **182**, 121-133.

Kragh-Hansen, U. (1990). Structure and ligand binding properties of human serum albumin. *Danish Med. Bull.* **37**, 57-84.

Li, X. and Noll, M. (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. *Nature* **367**, 83-87.

Littlewood, D. T. and 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. and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila*. *Mol. Biol. Evol.* **12**, 1002-1011.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of *eve* stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* **125**, 949-958.

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

Maduro, M. and Pilgrim, D. (1996). Conservation of function and expression of unc-119 from two *Caenorhabditis* species despite divergence of non-coding DNA. *Gene* **183**, 77-85.

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

Nocente-McGrath, C., Brenner, C. A. and Ernst, S. G. (1989). Endo16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* **136**, 264-272.

Piano, F., Parisi, M. J., Karess, R. and Kambysellis, M. P. (1999). Evidence for redundancy but not *trans* factor-*cis* element coevolution in the regulation of *Drosophila Yp* genes. *Genetics* **152**, 605-616.

Raff, R. A. (1996). *The Shape of Life*. Chicago: University of Chicago Press.

Ransick, A., Ernst, S., Britten, R. J. and Davidson, E. H. (1993). Whole mount *in situ* hybridization shows *Endo16* to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**, 117-124.

Rockman, M. V. and Wray, G. A. (2002). Abundant raw material for *cis*-regulatory evolution in humans. *Mol. Biol. Evol.* **19**, 1991-2004.

Schulte, P. M., Gomez-Chiarri, M. and Powers, D. A. (1997). Structural and

functional differences in the promoter and 5′ flanking region of *Ldh-B* within and between populations of the teleost *Fundulus heteroclitus*. *Genetics* **145**, 759-769.

Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. and Miller, W. (2000). PipMaker–a web server for aligning two genomic DNA sequences. *Genome Res*. **10**, 577-586.

Shashikant, C. S., Kim, C. B., Borbely, M. A., Wang, W. C. and Ruddle, F. H. (1998). Comparative studies on mammalian *Hoxc8* early enhancer sequence reveal a baleen whale-specific deletion of a *cis*-acting element. *Proc. Natl. Acad. Sci. USA* **95**, 15446-15451.

Shaw, P. J., Wratten, N. S., McGregor, A. P. and Dover, G. A. (2002) Coevolution in *bicoid*-dependent promoters and the inception of regulatory incompatibilities among species of higher Diptera. *Evol. Dev*. **3**, 265-277.

Singh, N., Barbour, K. W. and Berger, F. G. (1998). Evolution of transcriptional regulatory elements within the promoter of a mammalian gene. *Mol. Biol. Evol*. **15**, 312-325.

Sinha, N. and Kellogg, E. (1996). Parallelism and diversity in multiple origins of C4 photsynthesis in the grass family. *Am. J. Bot.* **83**, 1458-1470.

Soltysik-Espanola, M., Klinzing, D. C., Pfarr, K., Burke, R. D. and Ernst, S. G. (1994). Endo16, a large multidomain protein found on the surface and ECM of endodermal cells during sea urchin gastrulation, binds calcium. *Dev. Biol*. **165**, 73-85.

Sonnhammer, E. L. and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **176**, GC1-10.

Spitz, F., Gonzalez, F., Peichel, C., Vogt, T. F., Duboule, D. and Zakany, J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev*. **15**, 2209-2214.

Stern, D. L. (1998). A role of Ultrabithorax in morphological differences between *Drosophila* species. *Nature* **396**, 463-466.

Stockhaus, J., Schlue, U., Koczor, M., Chitty, J. A., Taylor, W. C. and Westhoff, P. (1997). The promoter of the gene encoding the C4 form of phosphoenolpyruvate carboxylase directs mesophyll-specific expression in transgenic C4 *Flaveria* spp. *Plant Cell* **9**, 479-489.

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

Tamarina, N. A., Ludwig, M. Z. and Richmond, R. C. (1997). Divergent and conserved features in the spatial expression of the *Drosophila pseudoobscura esterase-5B* gene and the *esterase-6* gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **94**, 7735-7741.

Theze, N., Calzone, F. J., Thiebaud, P., Hill, R. L., Britten, R. J. and Davidson, E. H. (1990). Sequences of the *CyIIIa* actin gene regulatory domain bound specifically by sea urchin embryo nuclear proteins. *Mol. Reprod. Dev*. **25**, 110-122.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Tumpel, S., Maconochie, M., Wiedemann, L. M. and Krumlauf, R. (2002). Conservation and diversity in the cis-regulatory networks that integrate information controlling expression of *Hoxa2* in hindbrain and cranial neural crest cells in vertebrates. *Dev. Biol*. **246**, 45-56.

Wang, D. G., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995) SpZ12-1, a negative regulator required for spatial control of the territory-specific *CyIIIa* gene in the sea urchin embryo. *Development* **121**, 1111-1122.

Wray, G. A. and Lowe, C. J. (2000). Developmental regulatory genes and echinoderm evolution. *Syst. Biol*. **49**, 151-174.

Wray, G. A., Hahn, M., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V. and Romano, L. A. (2003). Evolution of eukaryotic transcription. *Mol. Biol. Evol*. (in press).

Yuh, C.-H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069-1082.

Yuh, C.-H., Ransick, A., Martinez, P., Britten, R. J. and Davidson, E. H. (1994). Complexity and organization of DNA-protein interactions in the 5′-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech. Dev.* **47**, 165-186.

Yuh, C.-H., Moore, J. G. and Davidson, E. H. (1996). Quantitative functional interrelations within the *cis*-regulatory system of the *S. purpuratus Endo16* gene. *Development* **122**, 4045-4056.

Yuh, C.-H., Bolouri, H. and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**, 1896-1902.

Yuh, C.-H., Bolouri, H. and Davidson, E. H. (2001a). *Cis*-regulatory logic in the *Endo16* gene: switching from a specification to a differentiation mode of control. *Development* **128**, 617-629.

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

Yuh, C. H., Brown, C. T., Livi, C. B., Rowen, L., Clarke, P. J. C. and Davidson, E. H. (2002). Patchy interspecific sequence similarities efficiently identify positive cis-regulatory elements in the sea urchin. *Dev. Biol*. **246**, 148-161.

Zeller, R. W., Griffith, J. D., Moore, J. G., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995). A multimerizing transcription factor of sea urchin embryos capable of looping DNA. *Proc. Natl. Acad. Sci. USA* **92**, 2989-2993.

Zhu, X., Mahairas, G., Illies, M., Cameron, R. A., Davidson, E. H. and Ettensohn, C. A. (2001). A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development* **128**, 2615-2627.