Current Biology 19, 2014–2019, December 15, 2009 ©2009 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.10.036

Report

Highly Divergent Gene Expression Programs Can Lead to Similar Chordate Larval Body Plans

Daniel Sobral,[1,2,*] Olivier Tassy,[1,3] and Patrick Lemaire[1,*]
[1]Institut de Biologie du Développement de Marseille-Luminy (IBDML), UMR6216, CNRS, Université de la Méditerranée, Campus de Luminy, Case 907, F-13288 Marseille Cedex 09, France

Summary

The diversity of animal morphologies is thought to result largely from spatial or temporal variations in gene expression. Conversely, we explored here the extent of divergence in transcriptional expression patterns compatible with a common morphological output, the chordate larva. We compared two organisms that share a prototypical tadpole larval body plan but are separated by over half a billion years of divergent evolution: the zebrafish (*Danio rerio*) and the ascidian *Ciona intestinalis*, an invertebrate chordate belonging to the sister group of vertebrates [1]. The large databases of whole-mount in situ hybridization expression patterns available for these two species allowed us to carry out a systematic large-scale comparison of spatiotemporal expression patterns of 1103 groups of orthologous genes. We found an extensive overall divergence in gene expression profiles between the two species that was similar at all developmental stages and did not discriminate developmental regulators from their targets. The level of conservation in individual tissues, however, varied. Conservation of tissue-specific expression patterns was highest in tissues involved in locomotion, including muscle, notochord, and the central nervous system. Thus, a broad divergence in gene expression profiles is compatible with the conservation of similar body plans across large evolutionary distances.

Results and Discussion

A Pipeline to Compare Spatiotemporal Gene Expression Profiles across Evolution

Expression patterns in *Ciona* and zebrafish (*Danio rerio*) are classically described for each developmental stage of their normal development table, with species-specific anatomical vocabularies. Because the developmental stages and the anatomical ontologies differ between species, it is difficult to systematically compare expression patterns of orthologous genes. We thus first developed a pipeline to homogenize stages and annotations (Figure 1; see also Figure S1 available online). We defined a temporal correspondence between ascidian and fish developmental stages by extending to ascidians the core bilaterian stages concept [2]. Each bilaterian stage extends between major, evolutionarily conserved, embryonic transitions and includes several species-specific stages (Figure 1A). This approach alleviates the difficulty of precisely mapping individual stages, often arbitrarily defined, in each organism. For each of the three core stages that we used (gastrulation, neurulation, and organogenesis), we defined a generic chordate ontology (Figure 1B) onto which we projected species-specific anatomical terms. Terms describing structures without homologs in both species (e.g., paired fins, which are not present in *Ciona*) were omitted.

This mapping strategy allowed us to describe in a unified format the expression profiles of 561 (4213), 580 (4133), and 2047 (6315) genes in *Ciona* (zebrafish) at the gastrula, neurula, and organogenesis stages, respectively. Transcription factors and signaling ligands constituted 37% of the genes for which we could compare expression patterns between the two species (5% in the total *Ciona* gene complement), reflecting the fact that these important regulators of development have been studied extensively in both organisms.

Global Overview of the Transcriptional Programs of *Ciona* and Zebrafish

We observed in both organisms a progressive spatial restriction of expression patterns with time (compare the "Ubiquitous" columns in "Gastrula" and "Organogenesis" in Figure 2A). This spatial restriction occurred earlier in *Ciona* than in zebrafish, with 50% of genes with tissue-restricted expression at gastrulation in *Ciona* versus 10% in zebrafish. This result fits the proposition that fate restriction occurs earlier in *Ciona* [3] than in zebrafish [4].

We next clustered tissues in each species on the basis of the relatedness of their transcriptional programs. For this, we considered the expression of each gene as a discrete character for tissues (a gene is expressed in the tissue or not) and used a parsimony argument to cluster together, with the PARS program [5], tissues that expressed similar genes (see Supplemental Experimental Procedures). Reassuringly, this clustering recapitulated the known separation between ectodermal and endomesodermal germ layers during *Ciona* organogenesis (1662 tissue-restricted genes, Figure 2B). Similar results were obtained via neighbor joining or UPGMA algorithms, except that the position of the epidermis was more variable (data not shown). For *Ciona* gastrula and neurula stages, the relationships between tissues were less clear, possibly owing to a smaller number of tissue-restricted genes at these stages (284 genes at gastrula, 388 genes at neurula; data not shown). In zebrafish, expected relationships between tissues could be found at all stages (Figure 2B and data not shown), with the exception of epidermis at the organogenesis stage, which appeared surprisingly different from other ectodermal tissues (Figure 2B). The unexpected position of fish epidermis and variable position of *Ciona* epidermis as obtained via neighbor joining or UPGMA algorithms may reflect a biological reality, tree generation artifacts, or the difficulty to distinguish ubiquitous and epidermal labeling upon whole-embryo inspection: at stages where the epidermis does not cover the zebrafish embryo entirely (e.g., gastrula), this tissue grouped with other ectodermal tissues (data not shown).

We conclude that, with the possible exception of epidermal expression, the in situ hybridization data sets that we used

*Correspondence: sobral@ebi.ac.uk (D.S.), lemaire@ibdml.univ-mrs.fr (P.L.)
[2]Present address: European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[3]Present address: Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA

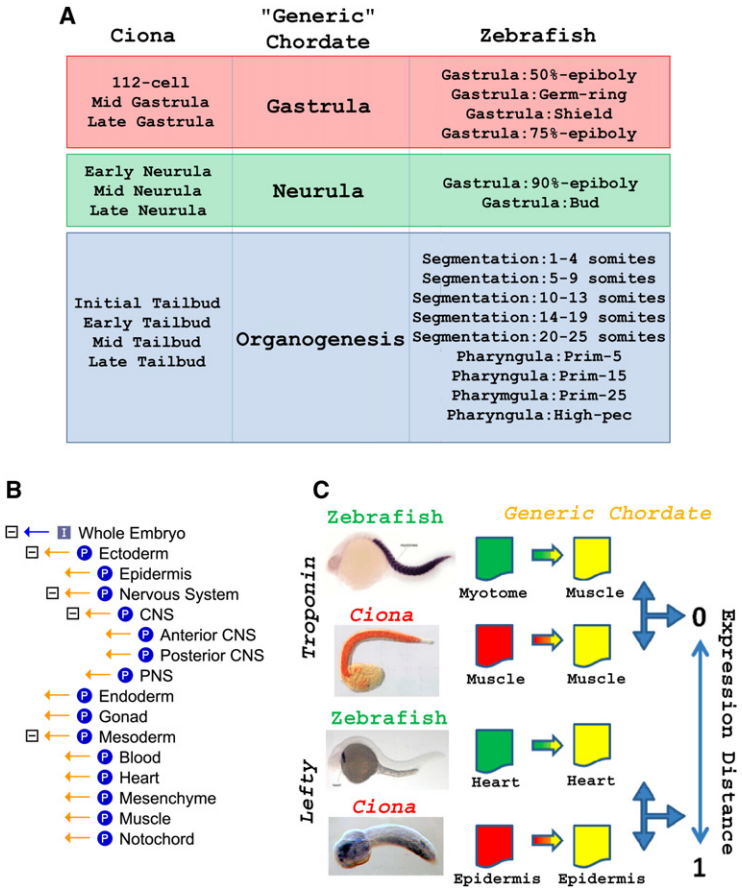

Figure 1. A Pipeline to Compare Spatiotemporal Gene Expression Profiles across Evolution

(A) Correspondence of developmental stages between *Ciona* and zebrafish through an intermediate generic chordate. The stages for the generic chordate are the same as the bilaterian stages from [2].

(B) Generic chordate anatomical ontology. All relationships are "part-of" relationships.

(C) General overview of the comparison process. Species-specific annotations were translated to the "General Chordate" ontology. Unified in situ hybridization (ISH) annotations of orthologous genes were then compared by using an expression distance (see Supplemental Experimental Procedures).

and our reannotation pipeline are consistent with previous knowledge of the transcriptional programs in *Ciona* and zebrafish.

The Expression Patterns of *Ciona* and Zebrafish Orthologs Are Broadly Divergent

We next compared the expression profiles of orthologous genes between *Ciona* and zebrafish. We could obtain one or more zebrafish orthologs for 1570 (76%) of the 2079 *Ciona* genes with expression data for at least one bilaterian stage. Of these *Ciona* genes, 1103 had at least one zebrafish ortholog (655 had one, 228 had two, 220 had three or more) with expression data available for a matching stage. Of note, there was no relation between the efficiency of ortholog detection and the tissues in which genes were expressed. Overall, we could compare the expression profiles in *Ciona* and zebrafish of 265, 259, and 1082 groups of orthologous genes at the gastrula, neurula, and organogenesis stages, respectively (totaling 1103 distinct orthologous groups).

Differences in the expression of orthologs were quantified (Figure 1C) by using an expression distance that considers the hierarchical relationships between terms in the anatomical ontology: two genes expressed in distinct tissues of the same germ layer are considered to have more similar expression patterns than two genes expressed in distinct germ layers (see Supplemental Experimental Procedures for more details). With this metric, we plotted for each stage the distribution of all distances of expression between any in situ hybridization (ISH) from a *Ciona* gene and any ISH from any of its fish orthologs (explained in Figure S1).

ISH expression patterns of orthologs are more similar than those of pairs of randomly associated genes (Figure 3A). The distribution of the expression distances between ISHs of orthologs, however, has a very high variance and widely overlaps with the distribution of the expression distance between ISHs of randomly associated gene pairs, indicating that most genes have diverging expression patterns between species. This result did not depend on the particular expression distance chosen (see Supplemental Experimental Procedures). Figure 3B shows two examples of orthologous genes with strongly diverging expression profiles in *Ciona* and zebrafish. The level of divergence observed was not altered when epidermal expression, possibly misannotated during zebrafish organogenesis, was ignored (data not shown). A broad divergence was observed even if we considered only the best matching ISH patterns between orthologs, instead of the whole distribution (Figure S3). Finally, we tested the relationship between the tissues of both species according to their overall transcriptional program by using the same discrete character parsimony argument that we used for the species-specific trees. Tissues tended to group together by species rather than according to their homology. (See Figure 3C for organogenesis; data not shown for other stages. For comparison, Figure S4 shows a tree obtained with *Ciona* genes and shuffled zebrafish orthologs.) Muscle was the only tissue that showed a strong support for the conservation of its transcriptional program between the two species. Additionally, the phylum-defining notochord showed weak support (50% of the parsimony trees).

This surprising observed divergence in the transcriptional programs of two distant chordates could be due to our a priori mapping of homologous developmental stages, which does not take into account possible heterochronies in the development of ascidians and fish. We thus attempted to map stages between species according to the similarity of their transcriptional programs (see Figure S5). This strategy did not detect major global or tissue-specific heterochronies between the two species up to the larval stages, nor did it point to clear homologies between developmental stages. Finally, although some tissues (heart, gut) only differentiate in *Ciona* after metamorphosis, the juvenile program of *Ciona* and the organogenesis program of fish showed no detectable similarity (Figure S6).

The teleost genome underwent three rounds of duplication since the split with ascidians. The divergence between orthologous gene expression patterns in fish and *Ciona* may thus