Genomic studies using the invertebrate chordate *Ciona intestinalis*

Jerry S. Chen and Robert W. Zeller

August 2011

Publication Number: CSRCR2011-01

COMPUTATIONAL SCIENCE & ENGINEERING

Genomic studies using the invertebrate chordate
Ciona intestinalis

Jerry S. Chen[1,2*], Robert W. Zeller[1,2]

ABSTRACT— Phylogenetically, the chordate *Ciona intestinalis* is positioned at the bridge between invertebrates and vertebrates, and as such facilitates comparative genomic studies relating the developmental regulatory mechanisms operating among non-chordate deuterostomes, invertebrate chordates and vertebrates. Here, with the recent publication of a new gene annotation model, we provide an updated report of developmental genetic studies using *Ciona intestinalis*, focusing on insights gained from the new gene model.

Index Terms—Ciona, genome, chordate

I. Ideal phylogenetic position of Ciona

The invertebrate chordate *Ciona intestinalis* belongs to the tunicate (urochordate) subphylum (**Figure 1**). It is one of a group of closely related marine tunicates known as ascidians or sea squirts. Ascidians were initially classified as mollusks based on their adult morphology but were later correctly classified as chordates based on the discovery of a prominent notochord and dorsal nerve cord in the larval tadpole (Cone, 2005; Dehal et al., 2002; Passamaneck and Di Gregorio, 2005; Satoh et al., 2003). In his publication on The Descent of Man in 1871, Charles Darwin claimed that the ascidian tadpole was a legitimate relative of vertebrates (Corbo et al., 2001; Dehal et al., 2002; Satoh et al., 2003). Before genome sequencing of was available it was thought based on morphologic comparisons that cephalochordates, which include the commonly used amphioxus Branchiostoma Floridae, were the closest invertebrate relative of vertebrates with tunicates being more distantly related. However, when the sequences of many representative chordate genomes became available, it was rigorously proven that tunicates are actually the sister group to vertebrates (Bourlat et al., 2006; Delsuc et al., 2006). This discovery has motivated the recent use of ascidians as a model organism for understanding basic developmental features common to chordates. Moreover, the ideal phylogenetic position of ascidians near the base of the chordate tree facilitates comparative studies relating the developmental regulatory mechanisms operating among

protostomes, non-chordate deuterostomes, invertebrate chordates and vertebrates (Cone, 2005).

II. Embryonic development of Ciona

The ascidian larva possesses a representative chordate body plan that includes an axial notochord flanked by muscle tissue, a dorsal nerve cord and a ventral endodermal strand (Corbo et al., 2001). As such, it is an ideal animal model of early chordate development: it possesses the basic morphologic and developmental characteristics of vertebrates but has the genomic and cellular simplicity of invertebrates (Corbo et al., 2001). The tadpole larva consists of ~2600 cells that comprise a limited set of tissues including notochord, muscle, central and peripheral nervous systems, epidermis, mesoderm and endoderm (Cone, 2005; Passamaneck and Di Gregorio, 2005; Satoh et al., 2003). Importantly, the lineage of embryonic cells is invariant and completely defined up until the early gastrula stage (Cone, 2005; Satoh et al., 2003). Moreover, each lineage leading to the formation of central and peripheral nervous systems, epidermis, endoderm, mesenchyme, trunk lateral cells, muscle or notochord is already reasonably well-characterized (Corbo et al., 2001; Pasini et al., 2006; Satoh et al., 2003).

Furthermore, unlike complex vertebrates such as humans and mice, ascidian development is extremely rapid – *Ciona intestinalis* grows into a fully-developed tadpole in only ~12 cell divisions and hatches at 18 hr post-fertilization when reared at 18 C. The entire life cycle takes less than three months, facilitating mutagenesis and genetic screens (Satoh et al., 2003).
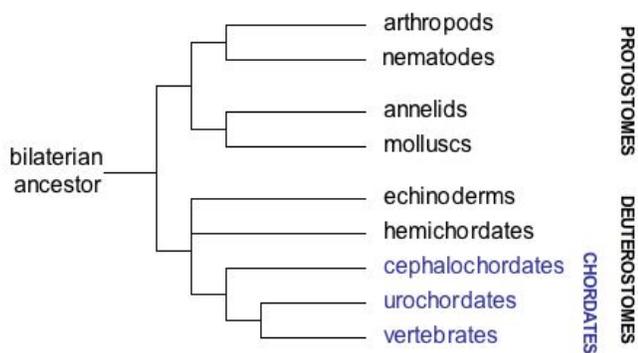
III. Genome annotation of Ciona

One of the other distinctive advantages of using Ciona intestinalis as a model organism is the thoroughness and accuracy with which the genome has been assembled and annotated. The *C.intestinalis* genome was sequenced in 2002 using a whole-genome shotgun approach at 8x coverage (Dehal et al., 2002). Approximately 115 of the ~155 Mbp (~75%) genome was accurately sequenced and assembled. The remaining ~25% could not be accurately assembled, presumably because these regions contain highly repetitive sequence or high allelic polymorphism (Dehal et al., 2002). Indeed, it was discovered that at least 1.2% of the genome exhibits allelic polymorphism, with even a greater percentage

Figure 1. **Bilaterian animal phylogeny.** Bilaterians are classified into two major superphylums based on the presence (deuterostomes) or absence (protostomes) of an anus. Among deuterostomes, chordates are classified by the presence of an axial notochord and a dorsal nerve cord. Ciona intestinalis belong to the urochordate subphylum, the sister group to vertebrates; and as such, possess many of the basic developmental mechanisms found in vertebrates.

of the genome likely to exhibit intraspecies polymorphism (Dehal et al., 2002). The genome is relatively compact and small, approximately 1/20 the size of the human genome, simplifying genome-wide studies using high-throughput technologies such as microarrays and deep sequencing. Finally, unlike the human genome, the *C.intestinalis* genome is particularly AT-rich (65%),

In 2008, using extensive 5'-full-length EST sequences, bacterial artificial chromosome-based end-sequence and chromosomal in situ hybridization data, an updated and much improved genome assembly was published (Satou et al., 2008). Approximately 68% of the total assembly was mapped to the 14 chromosomes of *C.intestinalis*. The remaining 32% was also accurately assembled but could not be confidently mapped.

One of the biggest improvements in the 2008 KH gene model set was the use of over a million *C.intestinalis* ESTs, which allowed for accurate prediction and annotation of the *C.intestinalis* genome. 96% of these ESTs were confidently mapped to the genome. Altogether, the 2008 KH gene model set consists of 24,025 transcripts representing 15,254 distinct gene loci (Satou et al., 2008). This number is comparable to other commonly used invertebrate model organisms and is about half the number of genes found in humans (Dehal et al., 2002). Transcribed genes account for ~15% of the entire genome, with only ~23% of the transcriptome being translated into protein (**Figure 2**). The remaining ~77% of the transcriptome consists primarily of introns (68%) and a smaller percentage of 3'UTR (~7%) and 5'UTR (2%). This data suggests that, as in other organisms, the majority of the genome consists of regulatory regions, highlighting the importance of understanding gene regulatory mechanisms. However, given the small size of the genome, gene density is relatively high (one gene per 7.5 kb) compared with human (one gene per 100 kb), indicating that gene regulation in *C.intestinalis* is greatly simplified compared to vertebrates.

Two of the more interesting features discovered through accurate mapping of gene loci were (1) the high prevalence of alternatively-spliced transcripts, particularly SL-trans-splicing, and (2) an unusually high percentage of operons (Matsumoto et al., 2010; Satou et al., 2008). On average, there are ~1.6 alternative transcripts per gene, with as many as 29 alternative transcripts for a single gene (gene ID: KH.C10.46). These include transcripts with alternate exons as well as alternative UTRs. An interesting class of alternative transcripts found in Ciona are those that arise from spliced-leader (SL) trans-splicing (Matsumoto et al., 2010; Satou et al., 2008). SL trans-spliced transcripts are formed by splicing at an acceptor site of a pre-mRNA molecule followed by 5'-end joining of a specialized donor exon known as an SL RNA (Matsumoto et al., 2010). In this way, the original 5'segment of the pre-mRNA is replaced by a unique SL RNA oligonucleotide. Remarkably, 58% of Ciona genes are SL trans-spliced in the tadpole larva, indicating its importance in early development (Matsumoto et al., 2010). Although one of the known auxiliary purposes of SL trans-splicing in Ciona is to resolve polycistronic operon transcripts into individual 5'-capped mRNAs, the primary function of this mechanism during development remains entirely unknown (Matsumoto et al., 2010).

Operons are typically associated with prokaryotes and are infrequently found in eukaryote genomes (Zeller, 2010). The most well-studied operon-present eukaryote is the nemaotde C.elegans, for which about 15% of the total genes are organized into operson (Zeller, 2010). Interestingly, the C.intestinalis genome is actually more operon-dense (~20% of total genes) than C.elegans. C.intestinalis operons usually contain two genes, but can contain as many as six genes (Satou et al., 2008; Zeller, 2010). Interestingly, the genes for only about half of the operons are co-expressed, suggesting complex mechanisms regulating operon transcription (Zeller, 2010). Finally, *C.intestinalis* operons contain a significantly high proportion of single-exon genes (38% versus 15% in non-operon genes) and are enriched for RNA transcription and splicing genes, suggesting a role of operons in fast and efficient expression of transcriptional processing machinery (Chen, 2011; Satou et al., 2008).

Aside from its unique characteristics, the *C.intesitnalis* genome provides an ideal platform for understanding genetic mechanisms in vertebrates. Approximately 80% of the *C.intestinalis* genome is homologous with the human genome. It has been proposed that large-scale gene duplications and addition of cis-regulatory machinery occurred in vertebrates after their divergence from tunicates and cephalochordates (Dehal et al., 2002). The resultant increase in gene number and regulatory complexity appears to underlie the complex developmental processes seen in vertebrates. This suggests that ascidians possess a greatly simplified, non-duplicated set of chordate genes. Indeed, recent studies have found that in many cases the *C.intestinalis* genome contains only a single homolog of the multiple paralogous genes found in vertebrate genomes. For example, there are only six
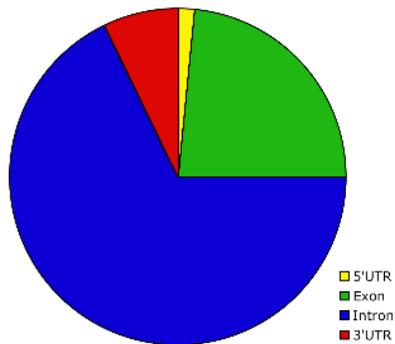
Figure 2. Genomic organization of Ciona intestinalis. An estimated 13 percent of the Ciona genome is transcribed into RNA. Among transcribed RNA, the majority (67.7%) codes for intronic regions, with the remainder coding for exon (23.3%), 3'UTR (7.3%) and 5'UTR (1.8%).

fibroblast growth factor genes in *C.intestinalis*, compared with twenty-two found in vertebrates (Passamaneck and Di Gregorio, 2005). Many other examples include zinc-finger DNA binding transcription factors, immunoglobulins and G protein-coupled receptors. The marked expansion of immunoglobins and sensory transmembrane proteins reflects two of the more important advancements in vertebrates: adaptive immunity and the nervous system (Dehal et al., 2002).

IV. GENETIC AND BIOINFORMATIC STUDIES USING CIONA

A wealth of experimental techniques and data are available for genetic studies in *Ciona intestinalis*. One of the greatest advantages of *C.intestinalis* as a model organism is the ability to generate viable transgenic animals by DNA electroporation into fertilized eggs.

By injecting DNA into fertilized eggs using a simple electroporation device, one can efficiently generate hundreds of transgenic embryos (Chen, 2009; Zeller et al., 2006). Although these embryos often express transgenes mosaically, the mosaic expression is predictable, and thus the technique has proven invaluable for examining the expression of tissue-specific regulatory modules that function during embryogenesis. Efficient DNA electroporation allows rapid, large-scale screening of genomic DNA fragments linked to reporter genes for the identification of cis-regulatory elements controlling gene expression, as well as misexpression of tissue-specific genes to analyze their function in the developing embryo. Recently, Chen, et.al. have also utilized this technique to develop a lineage-specific microRNA sensor assay (Chen, San Pedro and Zeller, in review).

Other techniques that have been well-developed in Ciona include the functional suppression of genes with morpholino oligonucleotides (Satou et al., 2001); generation of developmental mutants by chemical mutagenesis or insertional mutagenesis with a Minos transposable element (Passamaneck and Di Gregorio, 2005; Sasakura et al., 2003); the ability to generate stable, transgenic lines of ascidians (Deschet et al.,

2003); and more recently, the ability to generate knockouts using zinc-finger nucleases (Day and Zeller, manuscript in progress). With these tools available, Ciona researchers can perform gain- and loss-of-function studies on any gene of interest, further motivating the use of Ciona as a model organism for understanding chordate genomics.

A plethora of high-throughput data is publicly available for computational bioinformatics studies with Ciona. Global microarray gene expression data is available for 13 embryonic stages, tadpole and juvenile stages, and four adult stages (Azumi et al., 2007). This temporal data can be coupled with spatial in situ hybridization data to analyze the expression of most genes in the Ciona genome (Imai et al., 2004). In addition, ~2 million ESTs are available spanning the entire Ciona life cycle and covering several adult tissues (Satou et al., 2003; Satou et al., 2005; Satou et al., 2008). For analysis of small RNAs such as microRNAs, ~20 million small RNA sequences are available spanning six developmental stages (Hendrix et al., 2010; Shi et al., 2009).

Finally, many online resources are available for visualization and analysis of high-throughput data. Several genome browsers currently exist for visualizing various versions of the Ciona genome (Chen, 2010; Satou et al., 2008; Tassy et al., 2010). The KH and ANISEED genome browsers in particular provide links to a wealth of related resources such as EST and in situ expression data and GO annotations (KH: http://ghost.zool.kyoto-u.ac.jp/cgi-bin/gbrowse/kh/, ANISEED: http://www.aniseed. cnrs.fr/). For proteomics analysis, an integrated protein database has become recently available providing information on protein expression and localization as well as proteomics analysis tools (Endo et al., 2011). Finally, a very useful web-based resource called Four-Dimensional Body Atlas (FABA) has been developed for cell-resolution visualization of every stage of Ciona intestinalis development from fertilized egg until juvenile stage (FABA: http://chordate.bpni.bio.keio.ac.jp/ faba, FABA2: http://chordate. bpni.bio.keio.ac.jp/faba2).

V. CONCLUSION

Ciona intestinalis research is maturing at a rapid pace thanks to a robust genome assembly and new gene model annotation, as well as the development of experimental technologies for efficient manipulation of ascidian embryos. Coupled with the many web-based bioinformatics resources and high-throughput expression data available for Ciona, researchers can now begin to focus on more functional studies exploring the fundamental genetic mechanisms governing Ciona development; and by comparison with vertebrates, better understand the core components of developmental gene mechanisms common to all chordates.

ACKNOWLEDGMENTS

REFERENCES

[1] Azumi, K., Sabau, S. V., Fujie, M., Usami, T., Koyanagi, R., Kawashima, T., Fujiwara, S., Ogasawara, M., Satake, M., Nonaka, M., et al. (2007). Gene expression profile during the life cycle of the urochordate Ciona intestinalis. Dev Biol 308, 572-582.

[2] Bourlat, S. J., Juliusdottir, T., Lowe, C. J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E. S., Thorndyke, M., Nakano, H., Kohn, A. B., et al. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444, 85-88.

[3] Chen, J., RW Zeller (2009). Regulation of gene expression by the microRNA miR-124 in the developing nervous system of C.intestinalis. ACSESS Proceedings.

[4] Chen, J. S., RW Zeller (2010). Visualization of the invertebrate chordate Ciona intestinalis genome using a custom genome browser. CSRC Research Report.

[5] Chen, J. S., RW Zeller (2011). Operon analysis using an updated annotation of the Ciona intestinalis genome reveals a possible role of operons in transcription and RNA processing. ACSESS Proceedings.

[6] Cone, A. C. a. R. W. Z. (2005). Using ascidian embryos to study the evolution of developmental gene regulatory networks. Canadian Journal of Zoology 83, 75-89.

[7] Corbo, J. C., Di Gregorio, A., and Levine, M. (2001). The ascidian as a model organism in developmental and evolutionary biology. Cell 106, 535-538.

[8] Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., et al. (2002). The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science 298, 2157-2167.

[9] Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965-968.

[10] Deschet, K., Nakatani, Y., and Smith, W. C. (2003). Generation of Ci-Brachyury-GFP stable transgenic lines in the ascidian Ciona savignyi. Genesis 35, 248-259.

[11] Endo, T., Ueno, K., Yonezawa, K., Mineta, K., Hotta, K., Satou, Y., Yamada, L., Ogasawara, M., Takahashi, H., Nakajima, A., et al. (2011). CIPRO 2.5: Ciona intestinalis protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. Nucleic Acids Res 39, D807-814.

[12] Hendrix, D., Levine, M., and Shi, W. (2010). miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol 11, R39.

[13] Imai, K. S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131, 4047-4058.

[14] Matsumoto, J., Dewar, K., Wasserscheid, J., Wiley, G. B., Macmil, S. L., Roe, B. A., Zeller, R. W., Satou, Y., and Hastings, K. E. (2010). High-throughput sequence analysis of Ciona intestinalis SL trans-spliced mRNAs: alternative expression modes and gene function correlates. Genome Res 20, 636-645.

[15] Pasini, A., Amiel, A., Rothbacher, U., Roure, A., Lemaire, P., and Darras, S. (2006). Formation of the ascidian epidermal sensory neurons: insights into the origin of the chordate peripheral nervous system. PLoS Biol 4, e225.

[16] Passamaneck, Y. J., and Di Gregorio, A. (2005). Ciona intestinalis: chordate development made simple. Dev Dyn 233, 1-19.

[17] Sasakura, Y., Awazu, S., Chiba, S., and Satoh, N. (2003). Germ-line transgenesis of the Tc1/mariner superfamily transposon Minos in Ciona intestinalis. Proc Natl Acad Sci U S A 100, 7726-7730.

[18] Satoh, N., Satou, Y., Davidson, B., and Levine, M. (2003). Ciona intestinalis: an emerging model for whole-genome analyses. Trends Genet 19, 376-381.

[19] Satou, Y., Imai, K. S., and Satoh, N. (2001). Action of morpholinos in Ciona embryos. Genesis 30, 103-106.

[20] Satou, Y., Kawashima, T., Kohara, Y., and Satoh, N. (2003). Large scale EST analyses in Ciona intestinalis: its application as Northern blot analyses. Dev Genes Evol 213, 314-318.

[21] Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., and Satoh, N. (2005). An integrated database of the ascidian, Ciona intestinalis: towards functional genomics. Zoolog Sci 22, 837-843.

[22] Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L., Matsumoto, J., Wasserscheid, J., Dewar, K., et al. (2008). Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. Genome Biol 9, R152.

[23] Shi, W., Hendrix, D., Levine, M., and Haley, B. (2009). A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol 16, 183-189.

[24] Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R., et al. (2010). The ANISEED database: Digital representation, formalization, and elucidation of a chordate developmental program. Genome Res.

[25] Zeller, R. W. (2010). Computational analysis of Ciona intestinalis operons. Integr Comp Biol 50, 75-85.

[26] Zeller, R. W., Virata, M. J., and Cone, A. C. (2006). Predictable mosaic transgene expression in ascidian embryos produced with a simple electroporation device. Dev Dyn 235, 1921-1932.