Marine Genomics 22 (2015) 1–9

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Review

Do echinoderm genomes measure up?

R. Andrew Cameron[a,*], Parul Kudtarkar[a], Susan M. Gordon[a], Kim C. Worley[b,c], Richard A. Gibbs[b,c]

[a] Division of Biology 139-74, California Institute of Technology, Pasadena, CA, USA
[b] Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
[c] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Article history:
Received 9 December 2014
Received in revised form 5 February 2015
Accepted 6 February 2015
Available online 17 February 2015

Keywords:
Echinoderms
Echinoids
Genome
Development
Sequence assembly

ABSTRACT

Echinoderm genome sequences are a corpus of useful information about a clade of animals that serve as research models in fields ranging from marine ecology to cell and developmental biology. Genomic information from echinoids has contributed to insights into the gene interactions that drive the developmental process at the molecular level. Such insights often rely heavily on genomic information and the kinds of questions that can be asked thus depend on the quality of the sequence information. Here we describe the history of echinoderm genomic sequence assembly and present details about the quality of the data obtained. All of the sequence information discussed here is posted on the echinoderm information web system, Echinobase.org.

Contents

1. Introduction

Sea urchin gametes and embryos occupied front row seats for many of the innovations that propelled cell and developmental biology over the last 175 years. Using the low resolution microscopes of his day,

* Corresponding author at: Beckman Institute 139-74, California Institute of Technology, 1200 East California Blvd., Pasadena, CA 91125, USA.
E-mail address: acameron@caltech.edu (R.A. Cameron).

http://dx.doi.org/10.1016/j.margen.2015.02.004
1874-7787/

Derbes (1847) demonstrated the necessity of sperm for development to ensue but he couldn't see sperm–egg fusion. By the 1880s the phase contrast microscope was used to observe pronuclear fusion in sea urchin zygotes (Hertwig, 1876). The requirement of a complete set of chromosomes for development emerged from experiments on sea urchins in the early 1900s (Boveri, 1901). As developmental biologists began to examine cell lineages in embryos, the importance of intercellular communication in development grew out of blastomere recombination experiments in a Mediterranean sea urchin (Horstadius, 1939). The term chemical biology was coined mid-20th century to describe the innovations stemming from the use of cell fractionation by ultracentrifugation and allied techniques which again took advantage of the copious amounts of sperm, eggs and embryos available from sea urchins (Brachet, 1950). The advent of biological radionuclides afforded an opportunity to dissect the mechanisms of DNA replication, transcription and translation during this period and soon thereafter (reviewed in Davidson, 1968). Then, solution hybridization using DNA from sea urchin and other easily available sources became a favorite technique to explore genome structure and the mechanisms of gene expression (Britten and Davidson, 1969). The establishment of recombinant DNA technology that followed launched efforts to understand the mechanisms of gene regulation in development (Davidson, 1968). As molecular biology studies expanded, the sea urchin became a favored system for gene transfer (McMahon et al., 1984; Colin, 1986). It seemed remarkable that naked DNA constructs could be injected into zygotes where they were amplified along with nuclear DNA and were expressed in a manner identical to the exogenous sequences (Flytzanis et al., 1985).

By the end of the 20th century the catalog of expressed genes was extensive and the focus of gene expression studies had come to lie on the interactions between genes by means of the cis-regulatory modules that control them. In parallel, a community enterprise arose to support the sequencing of the purple sea urchin genome. It was realized that genome assemblies would be ultimately required to fully describe the intricate gene regulatory networks that drive development (http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/SeaUrchin_Genome.pdf).

It is the purpose of this essay to detail the series of sequencing activities that bring us to the assemblies of multiple echinoderm genomes available today. It relates the history over about 10 years of the efforts to construct an accurate draft genome for the purple sea urchin and the rapid expansion in additional species brought about by the disruptive technology of next-generation sequencing. In the process, we hope to give a sense of the experimental nature of the process of genome sequencing and assembly as well as the intellectual expansions and technical limitations that the quality and extent of the genomic information provide.

> *"Because of the small number of people producing this resource relative to the large number using it, the nature of the data is, unfortunately, not commonly appreciated ….."* (*Mardis et al., 2002*).

As Elaine Mardis says, the relatively solitary nature of genome sequencing efforts impedes a general appreciation for the quality of the data. Perhaps this essay will remedy this for echinoderm genomes.

2. The echinoderm phylum

2.1. Phylogeny

Echinoderms are bilaterian animals even though their adult body plans exhibit pentameral symmetry. The larval stages are definitely bilateral. Based on embryonic feature and recent molecular data, echinoderms occupy the same branch of the bilaterian tree as the chordates. Together with the hemichordates they form the Ambulacraria which is the sister group to the chordates. Of the five classes of echinoderms, four are the free-living eleutherozoans: echinoids (sea urchins), holothuroids (sea cucumbers), asteroids (sea stars), and ophiuroids (brittle stars). The mouth faces the substrate in these forms while the fifth class, the crinoids, has the mouth on the top surface. There have been two competing hypotheses about the relationships among the eleutherozoan classes. Two recent reports utilizing transcriptome data favor the Asterozoa topology where the asteroids and ophiuroids are a sister group to the holothuroids and echinoids (Telford et al., 2014; Reich and Wessel, unpublished data) (Fig. 1). The lack of resolution of these relationships until recently is probably due to a paucity of molecular data for some classes and to the rapid divergence of the groups (Pisani et al., 2012). The interval over which they are estimated to diverge is only about 35 million years in the Cambrian period about 500 million years ago.

That the phylogenetic relationships of echinoderm groups extend into deep evolutionary time offers an opportunity to examine histories of changes at a level available in few other places among the bilaterians. Comparisons of genomic structure among these animals have the capacity to reveal the milestones of genomic change that accompany the divergence of echinoderm classes. A common feature of echinoderms is a particular form of skeleton, the stereome, which is found in all of the adult forms. The development of this unique structure thus extends backward 540 million years (Bottjer et al., 2006). The way in which the structural gene batteries and developmental gene regulatory networks may have changed is intriguing. Only sea urchins and brittle stars have prominent skeletal elements in embryonic stages. (Sea cucumbers have small spicules in the developmental stages. These are likely homologous to the sea urchin ones.) Considering the asterozoan topology these structures are either a result of convergent evolution or existed in the common ancestor of the four eleutherozoan classes and were lost in asteroids.

The data are still scarce but one study found no skeletal matrix proteins shared between the well-studied sea urchins and an ophiuroid (Vaughn et al., 2012). This observation leans the inference toward convergent evolution of larval skeletons.

2.2. Echinoderm sequencing candidates

Representative members of the echinoderm classes were chosen for genome sequencing to complement ongoing research and address some of the evolutionary topics detailed above (Table 1). Due to the extensive body of work on molecular mechanisms of cell and developmental biology, the purple sea urchin, *Strongylocentrotus purpuratus* (Sp) was chosen as the first subject for sequencing. There already existed a suite of resources for genomic studies in this species in the form of arrayed cDNA and genomic DNA libraries (Cameron et al., 2000). An informal network of investigators supported this first project. The cidaroid sea urchin *Eucidaris tribuloides* (Et), is diverged from the reference species by 255 MY and exhibits interesting differences in the mode of skeletal formation. The variegated sea urchin, *Lytechinus variegatus* (Lv) from the east coast of North America is diverged from the common ancestor of the purple sea urchin by about 50 MY. It has been used as a research model for many years and has recently been shown to provide genomic comparisons that reveal conserved non-coding sequences likely to be sites of transcriptional control of protein coding genes.

Based on comparison between five functionally characterized cis-regulatory modules (CRMs) from the *S. purpuratus* genome and orthologous regulatory and flanking sequences obtained from a bacterial artificial chromosome genome library of a congener, *Strongylocentrotus franciscanus* (Sf), it was observed that large indels are statistically almost absent from cis-regulatory modules at this evolutionary distance of about 20 MY (Cameron et al., 2005a). This metric though probabilistic could be used to help characterize CRMs and it was decided to sequence the genomes of two species at this close evolutionary distance. Therefore, *S. franciscanus* and *Allocentrotus fragilis* (Af), were selected for limited sequencing.