

Dicyemida and Orthonectida: Two stories of body plan simplification

O.A. Zverkov, K.V. Mikhailov, S.V. Isaev, L.Y. Rusin, M.D. Logacheva, A.A. Penin,
Y.V. Panchin, V.A. Lyubetsky, V.V. Aleoshin

*Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute),
Bolshoy Karetny per. 19, Moscow, 127051, Russia, zverkov@iitp.ru*

L.L. Moroz

*Whitney Laboratory for Marine Bioscience, University of Florida,
St Augustine, 32080, USA, moroz@whitney.ufl.edu*

Abstract. Mesozoa is an enigmatic animal phylum proposed in the 19th century as a missing link between protists and multicellular animals. Mesozoa encompasses two groups of morphologically simple parasites of invertebrates: Dicyemida and Orthonectida. We obtained transcriptomic data for two *Dicyema* spp. and genomic data for one species. Genomic data confirms the existence of heterogenous extrachromosomal circular DNAs, which includes coding regions of their mitochondrial genome. The dicyemids display reduction of genes linked to the formation of basement membrane and function of muscle and nervous systems. Homeodomain (homeobox) transcription factors, G-protein-coupled receptors and many other families have undergone massive reduction in dicyemids compared to other animals. We constructed and analyzed a large dataset of predicted orthologous proteins from three species of *Dicyema* and a set of spiralian animals including the orthonectid *Intoshia linei*. Bayesian analyses recovered the orthonectid lineage within the Annelida. In contrast, dicyemids form a separate clade with weak affinity to the Rousphozoa (Platyhelminthes plus Gastrotricha). Mesozoa is a polyphyletic taxon. Dramatic simplification of the body plan in dicyemids and orthonectids as well as their metagenetic life cycle evolved independently in these two lineages.

“Circular” contigs and mitochondrial DNA in *Dicyema* sp. Using NGS data we have identified 24,065 “circular” contigs. The search for mtDNA genes in the entire genome concluded that all the sequences found are encoded in distinct circles according to the rule which can be formulated as “one gene, one circle”. The signal presence in circles both with mtDNA genes and without them seems to be surprising. It can be interpreted as a

consequence of a similar mechanism of generation and maintaining of circles irrespective of their origin. Therefore, we confirm the localization of the mitochondrial genes of dicyemids on mini-circle DNAs. This is a fairly rare feature for the animal's mitochondrial genome.

Homeobox transcription factors. Phylogenetic inference with PRD class homeoboxes outlines six dicyemid clusters, including the orthologs of the Pax6 and Otx families. The dicyemid genes suggests that overall they possess fewer homeoboxes than the orthonectid *I. linei* and their sequences are also markedly more diverged. Dicyemids also experience several lineage-specific expansions of homeoboxes, including the duplication of central Hox gene DoxC. Expansions in their Hox-like gene complement might be an indication of major restructuring of their developmental program following the loss of organs and classical tissue segmentation.

Basal lamina. The BLAST and Pfam searches showed that these domains are absent from the genome of *Dicyema* sp., so there is no evidence that dicyemids have a basement membrane at all. The absence of the basement membrane is consistent with a reported loss of muscular and nervous systems in these animals.

Membrane receptor proteins. We found only two PF00001 7 transmembrane receptor (rhodopsin family) proteins in *Dicyema* sp. Four proteins from another GPCR 7 transmembrane receptor family (PF00002 Secretin family) were predicted in our *Dicyema* sp. data. Another big group of ionotropic receptors is the Cys-loop ligand-gated ion channel superfamily that is composed of nicotinic acetylcholine, GABAA, GABAA- ρ , glycine, 5-HT₃, and zinc-activated (ZAC) receptors. In *Dicyema* sp. we found 8 genes for this superfamily, identified by the specific transmembrane region domain (PF02932) and the ligand binding domain (PF02931).

Proteins for neuro-muscular system. Muscle cells are absent in dicyemids and detection of the muscle-specific genes in these animals is intriguing. In summary, we can conclude that genes related to the basement membrane, neuronal and muscular systems data confirms extreme simplicity of dicyemids. At the same time the spectra of loss and presence of such genes in *Dicyema* sp. and Orthonectida are different, supporting the conclusion that these animal groups are not close relatives and have simplified independently. Another intriguing feature is the presence of troponin complex that is specific for striated muscles in dicyemids

that lack muscle cells. Together with a relatively big set of nicotinic acetylcholine receptors, that are often associated with neuromuscular signaling it appears important to re-evaluate the present view that neuro-muscular system was totally lost in dicyemids.

Gap junctions and adhesion molecules. BLAST and Pfam searches with our dicyemid data detected 21 hits with the innexin/pannexin-specific domain (PF00876) and no connexins. The universal metazoan proteins Integrin alpha and Integrin beta are detected in dicyemids in single copies, immunoglobulin domain is present in 6 sequences and Cadherin in 18 copies.

Axon guidance molecules and their receptors. Both Orthonectida and Dicyemida, lack semaphorins, important neuronal pathfinding signaling molecules, and their receptors (plexins). Other players potentially involved in the nervous system development such as Netrin, Ephrins, Ephrin receptors are present in Orthonectida but not in Dicyemida.

Peroxisome. Analyzing conserved genes lost in Dicyemida we found that proteins and PFAM domains specific to peroxisome organelles found in virtually most eukaryotic cells are absent from the dicyemid data. In this respect dicyemids are similar to Orthonectida and parasitic flatworms.

Phylogenetic analyses. To clarify the relationships of the two mesozoan groups, Orthonectida and Dicyemida, we sequenced transcriptomes of two unidentified species of *Dicyema*. The consensus Bayesian tree was obtained from four independent chains. Importantly, in all topologies, the orthonectid is nested within the Annelida, and the dicyemid lineage is inferred sister to the assemblage of Platyhelminthes, Gastrotricha, Entoprocta, and Cycliophora (to the inclusion of Bryozoa in some topologies). Maximum likelihood (ML) analyses of the same dataset with RAxML and IQ-TREE produce a different view on the phylogeny of Mesozoa. The dicyemids and the orthonectid form a monophyletic group in ML trees with maximal support. Although ML analyses show disagreement with the result of Bayesian inference, modeling of site-heterogeneity by the IQ-TREE profile mixture model does shed light on some spurious cases in spiralian relationships. The divergent annelid *Myzostoma* is correctly grouped with other annelids in the IQ-TREE analysis, in contrast with the RAxML tree where it forms a clade with long branches of the Rousphozoa, Gnathifera, and Mesozoa. To test for expected LBA effects, particularly to exclude the possibility of the orthonectid being attracted to annelids by the divergent *Myzostoma*, we conducted additional

analyses excluding each of the long branched lineages. Bayesian analyses of additional datasets recovered placement of *Intoshia* within annelids in the absence of *Myzostoma*. The position of dicyemids is also unaffected by exclusion of other long-branched taxa – the dicyemids occupy a basal position within the Rousphozoa+Trochozoa subtree in all analyses of the additional datasets. We also tested our dataset for effects of compositional heterogeneity by discarding highly heterogeneous partitions and utilizing the data recoding approach. Bayesian inference with 150 partitions retained after discarding the highly compositionally heterogeneous part of the dataset recovers the same groupings of the mesozoan taxa compared to the intact matrix. The orthonectid is nested within the annelid clade and the dicyemids branch off at the base of the Rousphozoa + Entoprocta + Cycliophora clade with weak support. Similarly, inference with the Dayhoff-recoded alignment groups the orthonectid with annelids, while leaving the position of the dicyemids uncertain within the Lophotrochozoa. Remarkably, the PhyloBayes run with recoded data shows adequate convergence between chains and infers the monophyletic Gastrotricha. Several conventional groupings, such as the Rousphozoa, are not recovered. Consistent with the proposed artefactual nature of the grouping of Bryozoa and Entoprocta due to compositional heterogeneity, both test datasets support the monophyletic Lophophorata, whereas the alternative Polyzoa was frequently observed for the complete and non-recoded data. We do not observe monophyly of the Mesozoa in any of the chains, nor do we observe their direct relations with Platyhelminthes. The orthonectid *Intoshia linei* occupies a stable position within the annelid part of the tree.

Conclusions. Orthonectids are extremely simplified annelids and do not form a monophyletic group with dicyemids. Mesozoa is a polyphyletic taxon. Dramatic simplification of the body plan in dicyemids and orthonectids as well as their metagenetic life cycle evolved independently in these two lineages. Many, but not the same Bilateria's conservative genes are absent in the genomes of Dicyemida and Orthonecida, which indicates the independence of their loss. Thousands of minicircle molecules are present in the total dicyemid DNA. All protein-encoding genes, rRNA and tRNA genes typical of metazoan mitochondrial DNA are located on minicircle molecules. There are short nucleotide motifs specific to minicircle DNA, which are contained in 5-90% of these molecules.