

De-novo transcriptome assembly of starfish, *Asterias rubens*

Asiya Sadritdinova

Engelhardt Institute of Molecular Biology, RAS, 119991, Moscow, Russia

Maxim Belenikin

Research Institute of Physico-Chemical Medicine, 119435 Moscow, Russia, camd@niifhm.ru

The Common Starfish or *Asterias rubens* is the most common and familiar starfish found in Northern Atlantic region. Sea stars are using as a model organism for histological and cytological studies. In this work we perform de novo assembly of transcriptome data of digestive caeca of *Asterias rubens* in normal and stress (30°C during 30 min) conditions. All data were produced by next-generation sequencing using Illumina GAIIX (Institute of Molecular Biology RAS). The raw sequenced data were 64.3M and 55.8M paired-end 76 bp reads for samples at normal (1) and stress (2) conditions, respectively. We have performed de novo assembly using Trinity [1] and SOAPdenovo [2] de novo transcriptome assemblers. Using Trinity, the total length of transcripts varied depending on read preprocessing strategies from 38.50M bp to 51.61M bp for (1) and from 52.25M bp to 57.27M bp for (2). The total number of transcripts varied from 70444 to 83053 for (1) and from 84460 to 88577 for (2). The max length of transcripts varied from 13943 bp to 14100 bp for (1) and from 15284 bp to 14385 bp for (2). SOAPdenovo assembler generated 96210 scaffolds and contigs. To discriminate differences between assembled transcriptomes for samples at normal and stress conditions and to compare the quality of both assemblers the differential expression analyses were performed. Tblastx was used for transcriptome annotation. *Strongylocentrotus purpuratus* genome was used to identify a few reference genes which were confirmed experimentally. We thank to A. Kudryavtseva, N. Melnikova, and A. Speranskaya from Engelhardt Institute of Molecular Biology for sequencing and providing the data of next-generation sequencing (Illumina GAIIX).

1. M.G.Grabherr et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nature Biotechnology*, **29**:644-652.
2. Luo et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, *GigaScience*, **1**:18.