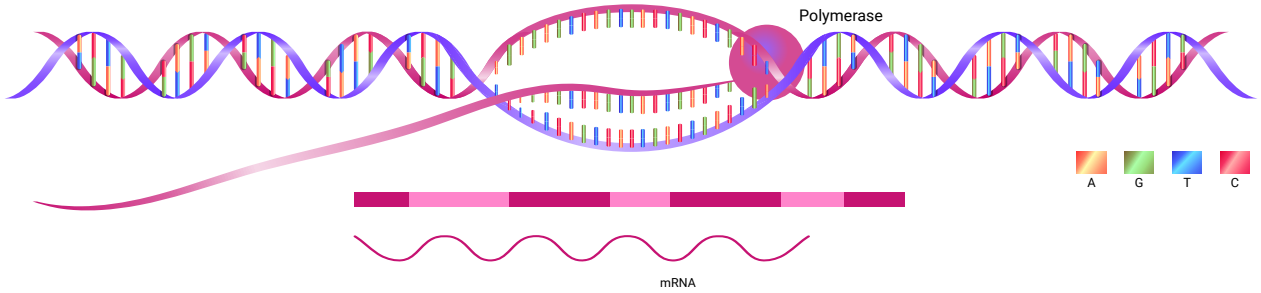
Transcriptome analysis using RNA-seq

Life may be possible without DNA, but without RNA molecules, it is definitely not. They perform life-essential functions like carrying the genetic information out of the nucleus (mRNA), translating the information encoded into amino acid information (tRNA) and making the complex structure that synthesizes the proteins encoded by mRNA (rRNA). Understanding the types and abundance of RNA molecules present in each cell or a cell type under specific conditions is the goal of RNA research. The advent of next-generation sequencing technologies has enabled massive and deep studies on RNA. The information contained in the genome (DNA) is transcribed into RNA molecules by a process called transcription. The transcribed regions of the genome encode different types of RNA molecules, like mRNA, tRNA, rRNA, miRNA, lincRNA, snoRNA and others. The full complement of transcribed RNA products constitutes the transcriptome. The use of high throughput sequencing enables the study of transcriptomes and is termed as RNA-Seq.

RNA-Seq typically has been mainly employed to selectively study messenger RNA (mRNA), although other RNA types have also been studied. The mRNA, a single-stranded RNA derived by transcription of gene coding segments of the genome, is translated by the protein synthesis machinery in the cells to produce proteins. Proteins ultimately drive almost all the biological processes within a cell. RNA-Seq provides a comprehensive and quantitative survey of the cellular transcriptome. It allows the identification of novel transcripts, splicing isoforms, and gene-fusion transcripts. The deep sequencing technology can identify sparsely expressed transcripts with low background noise. Apart from profiling, RNA-Seq data also provides quantitative and differential gene expression information.



Applications

RNA-seq has multiple applications in agriculture, animal husbandry, and health. For research that involves monitoring the cellular response to stimuli, treatment or environment change, understanding the differential gene expression patterns can provide insights into their physiological responses. Profiling the various RNA molecules provides insights into factors defining specific phenotypes for e.g. the color of tomato, medicinal properties of herbs, secondary metabolites in prokaryotes, a bioactive compound in a marine organism etc.

RNA-Seq at AgriGenome

RNA-Seq data collection at AgriGenome involves the following

1. Illumina Paired-end libraries

- RNA Quality Control – High-quality RNA quantified and qualified using Qubit Fluorometer and Agilent Tapestation.
- Library generation – Stranded, non-stranded, ribosomal RNA-depleted, mRNA enriched, etc
- Sequencing of Libraries – HiSeq 4000 and HiSeq X platforms to generate read lengths of 2x100 bp or 2x150 bp
- Libraries are produced by generating cDNA from RNA and then adding adapters to the cDNAs. Fig.1 provides details of RNA-Seq library construction, and sequencing.

2. PacBio Iso-Seq

- Full-length cDNA preparation from RNA and fractionation into small and large transcripts
- Library preparation using PacBio protocols
- Sequencing on PacBio sequel platform to obtain full-length circular consensus sequence (CCS) reads

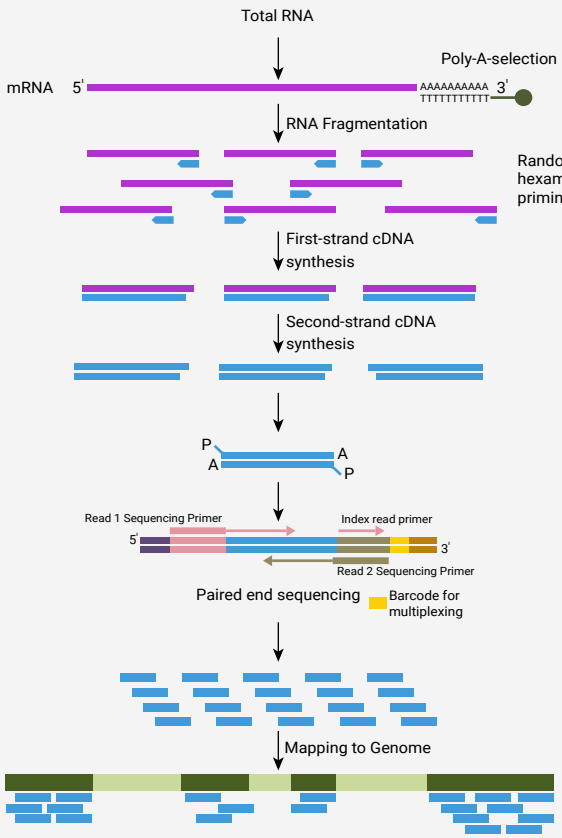


Fig. 1: RNA-Seq workflow at AgriGenome

Bioinformatics analysis at AgriGenome

The major application of RNA-Seq and Iso-Seq are

- Novel transcripts discovery
- Transcript abundance and gene expression profiles
- Coding regions in a genome
- Splice junctions, splice variants and splicing patterns
- Comparison of transcript abundance between samples
- Fusion transcripts

Types of analysis - Reference-based or *de novo*?

The most suitable method can be chosen from the large number of available methods based on the required analysis. Typically, RNA-Seq reads of an organism can be mapped back to, and expression estimates calculated based on, its reference genome. For organisms that do not have a reference genome, the reads can be assembled *de novo*. The reads are mapped back to this draft assembly for estimating expression levels. Expression estimates are normalized for gene length and read abundance, so it can be compared across datasets. AgriGenome has developed custom pipelines that can carry out accurate assembly and analysis of the RNAseq data generated. The general scheme of analysis can be seen in Figure.2

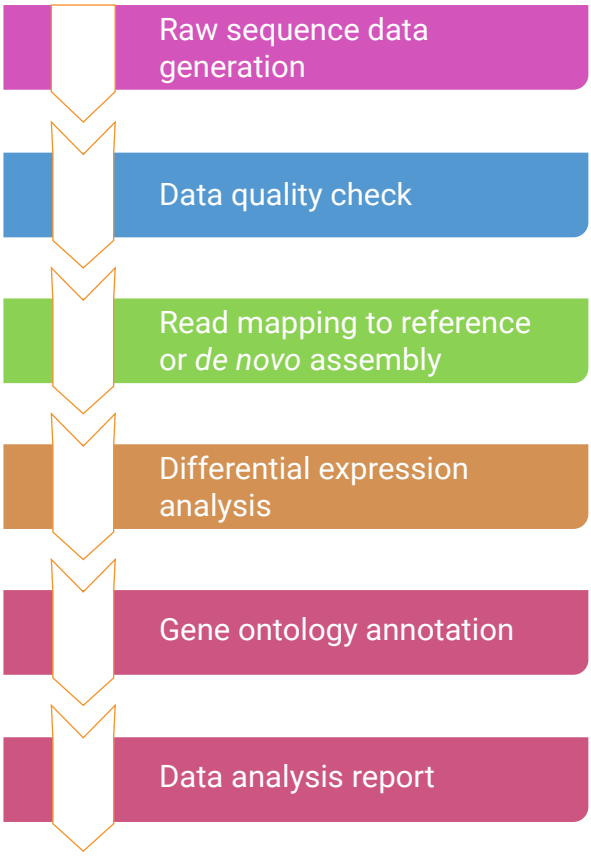


Fig. 2: Bioinformatics Workflow

Recent publications from projects executed at AgriGenome

Suryamohan, K., Krishnankutty, S.P., Guillory, J., Jevit, M., Schröder, M.S., Wu, M., Kuriakose, B., Mathew, O.K., Perumal, R.C., Koludarov, I., Goldstein, L.D., Senger, K., Dixon, M.D., Velayutham, D., Vargas, D., Chaudhuri, S., Muraleedharan, M., Goel, R., Chen, Y.-J.J., Ratan, A., Liu, P., Faherty, B., de la Rosa, G., Shibata, H., Baca, M., Sagolla, M., Ziai, J., Wright, G.A., Vucic, D., Mohan, S., Antony, A., Stinson, J., Kirkpatrick, D.S., Hannoush, R.N., Durinck, S., Modrusan, Z., Stawiski, E.W., Wiley, K., Raudsepp, T., Kini, R.M., Zachariah, A., Seshagiri, S., 2020. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nature Genetics 52, 106–117. https://doi.org/10.1038/s41588-019-0559-8

Arvind, K., Rajesh, M.K., Josephrajkumar, A., Grace, T., 2020. Dataset of de novo assembly and functional annotation of the transcriptome of certain developmental stages of coconut rhinoceros beetle, *Oryctes* rhinoceros L. Data in Brief 28, 105036. https://doi.org/10.1016/j.dib.2019.105036

Ranjani, T.S., Pitchika, G.K., Yedukondalu, K., Gunavathi, Y., Daveedu, T., Sainath, S.B., Philip, G.H., Pradeepkiran, J.A., 2020. Phenotypic and transcriptomic changes in zebrafish (*Danio rerio*) embryos/larvae following cypermethrin exposure. Chemosphere 249, 126148. https://doi.org/10.1016/j.chemosphere.2020.126148

Thakur, V., Bains, S., Pathania, S., Sharma, S., Kaur, R., Singh, K., 2020. Comparative transcriptomics reveals candidate transcription factors involved in costunolide biosynthesis in medicinal plant- *Saussurea lappa*. International Journal of Biological Macromolecules 150, 52–67. https://doi.org/10.1016/j.ijbiomac.2020.01.312

Puvvala, S.S., Muddanuru, T., Thangella, P.A.V., Kumar, O.A., Chakravartty, N., Vettath, V.K., Katta, A.V.S.K.M., Lekkala, S.P., Kuriakose, B., Gupta, S., Mulpuri, S., Lachagari, V.B.R., 2019. Deciphering the transcriptomic insight during organogenesis in Castor (*Ricinus communis L.*), Jatropha (*Jatropha curcas L.*) and Sunflower (*Helianthus annuus L.*). 3 Biotech 9, 434. https://doi.org/10.1007/s13205-019-1960-9

Nguyen, T.T., Suryamohan, K., Kuriakose, B., Janakiraman, V., Reichelt, M., Chaudhuri, S., Guillory, J., Divakaran, N., Rabins, P.E., Goel, R., Deka, B., Sarkar, S., Ekka, P., Tsai, Y.-C., Vargas, D., Santhosh, S., Mohan, S., Chin, C.-S., Korlach, J., Thomas, G., Babu, A., Seshagiri, S., 2018. Comprehensive analysis of single molecule sequencing-derived complete genome and whole transcriptome of *Hyposidra talaca* nuclear polyhedrosis virus. Sci Rep 8, 1–11. https://doi.org/10.1038/s41598-018-27084-y

FOR MORE INFORMATION

Kochi
5th Floor, SCK 01 Building "SmartCity Kochi",
Infopark Road, Kakkanad,
Kerala, 682 042, India

Hyderabad
DS-10, The Sustainability Innovation Center (SINC),
IKP Knowledge Park, Genome Valley, Koltur, Hyderabad,
Telengana, 500 078, India

Delhi
27 Ground Floor, TDI Centre,
Jasola District Centre,
New Delhi, 110 025, India

+91-484-4234-234 sales@aggenome.com