

SD EPSCoR Bioinformatics workshop

August 10th, 2017 from 10am to 4pm, Solberg 103, SDSU main campus.

Topics: Tool discovery, Meta-genomics, Gene regulation, and RNA-Seq

Note: You can also participate over the internet via Webinar. The 30 desktops are available on a first come first server basis. Otherwise bring your own laptop. Pre-install required software.

It's free. Spots are limited, please register here: <https://GOO.GL/qsZyqR>

Contact: Xijin Ge, Xijin.Ge@sdstate.edu 605-215-9254

1. Bio-TDS (Bioscience Query Tool Discovery Systems)

Dr. Carol Lushbough, USD (30 minutes)

The Bio-TDS (Bioscience Query Tool Discovery Systems, <http://biotds.org/>) has been developed to assist researchers in retrieving the most applicable analytic tools by allowing them to formulate their questions as free text. The Bio-TDS is a flexible retrieval system that affords users from multiple bioscience domains (e.g. genomic, proteomic, bio-imaging) the ability to query over 12 000 analytic tool descriptions integrated from well-established, community repositories. One of the primary components of the Bio-TDS is the ontology and natural language processing workflow for annotation, curation, query processing, and evaluation. <https://academic.oup.com/nar/article/45/D1/D1117/2290903/Bio-TDS-bioscience-query-tool-discovery-system>

2. Introduction to metagenomics data analysis and mining

Dr. Erling Zeng & graduate student, USD (2 hours)

We are going to introduce metagenomics data analysis including:

1. from raw data to OTU table,
2. from OTU table to pathway information, and
3. from OTU to ROC curves.

For task 1, QIIME should be installed on a Linux system. If it is on Windows PC, a virtual box can be used (http://qiime.org/install/virtual_box.html). For task 2, PICRUST will be used. PICRUST is now available on Galaxy. Students should have Galaxy installed (<http://galaxy.morganlangille.com/>). For task 3, an in-house tool available on <http://biocomps.org/tools/> will be used. Right now, this tool is not interactive, by workshop time, we should have it ready to use.

3. Integrated and systematic views of regulatory DNA motif identification and analyses

Dr. Qin Ma, SDSU (2 hours)

I am going to introduce regulatory DNA motifs analysis and modeling. I will cover the background, rationale, computational challenges, genomics data analysis (ChIP-seq data), and mathematical modeling of the motif finding problem. And I will show the students how to proceed through my recent web sever, DMINDA. We do not need to install any software if we have the access to internet. The DMINDA link is: <http://bmbi.sdstate.edu/DMINDA2/index.php>.

4. RNA-Seq analysis made easy with iDEP

Dr. Xijin Ge, SDSU (1.5 hours)

We will first demonstrate how to use alignment-free mappers like Salifish or Kallisto to convert raw reads into read counts per gene. This can be done either through command line or on the web using Galaxy (<https://usegalaxy.org/>). Then we will use iDEP (integrated Differential Expression and Pathway analysis), a web application that can turn read-counts into pathways in minutes (<http://ge-lab.org/idep/>). Developed by Dr. Ge based on Shiny, iDEP makes many powerful R/Bioconductor packages easily accessible to biologists by wrapping them under a graphical inter-face, alongside annotation databases for 100+ species. See preprint: <http://www.biorxiv.org/content/early/2017/06/09/148411> We will also demonstrate ShinyGO, a tool for Gene Ontology enrichment analysis. <http://ge-lab.org:3838/go/>

Supported by SD EPSCoR, BioSNTR, NSF, with assistance from SDSU Dept. of Math & Stat.

