

## The 'slog through it together' bioinformatics reading group

**Purpose:** to create a welcoming space to read and critique current papers in bioinformatics. Some of these papers are challenging!

**Goals:** to understand: 1) what these programs are designed to do, 2) how they're implemented, 3) their limitations/biases, 4) how to parameterize them for different types of data with various considerations.

### TOPICS:

**Week 1- de Bruijn graph theory & implementation in short-read aligners**

(Compeau et al 2011, *Comp. Bio.*; Iqbal et al 2001 doi:10.1038/ng.1028)

**Week 2- K-mer selection—how does kmer size affect transcriptome assembly?**

(Durai & Schulz 2016; doi: 10.1093/bioinformatics/btw217)

**Week 3- Khmer RNAseq normalization and kmer analysis—when/how/why?**

(Crusoe et al 2015, doi: 10.12688/f1000research.6924.1)

**Week 4- Algorithmic/performance differences—de novo genome assembly (SOAPdenovo & AbySS)**

(Bradnam et al 2013, <http://www.gigasciencejournal.com/content/2/1/10>, Earl et al 2011 <http://www.genome.org/cgi/doi/10.1101/gr.126599.111>)

**Week 5- Algorithmic/performance differences—transcriptome assembly (SOAPdenovo-Trans, Trans Abyss, Trinity)—pt. 1**

(Grabherr et al 2013 doi:10.1038/nbt.1883; Nakasugi et al 2014 *PLOS one*)

**Week 6- Hidden Markov Models—what are they and why are they used everywhere?**

(Rabiner 1989: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=18626>; Wang et al 2007 [www.genome.org/cgi/doi/10.1101/gr.6861907](http://www.genome.org/cgi/doi/10.1101/gr.6861907))

**Week 7- Orthofinder—how it determines orthology, how to use it, limitations for non-model species**

(Emms and Kelly 2015, doi:10.1186/s13059-015-0721-2)

**Week 8- De novo transcriptome assembly for non-model organisms—comparisons pt. 2**

(Huang et al 2016, DOI 10.1186/s12864-016-2923-8)

**Week 9- Error correction—when and why to correct, and when not to—pt. 1**

(MacManes and Eisen 2013, DOI 10.7717/peerj.113; Yang et al 2012, doi:10.1093/bib/bbs015)

**Week 10- Joint Frequency spectra—theory and usage for inferring species/population history**

(Chen et al 2007, DOI: 10.1534/genetics.107.070730; Gutenkunst et al 2009; doi:10.1371/journal.pgen.1000695)

**Week 11- Batch effects—how to deal with (and hopefully avoid) them**

(Gilad and Mizrahi-Man study of mouse ENCODE data: <http://f1000research.com/articles/4-121/v1>; Leek et al. review: <http://www.nature.com/nrg/journal/v11/n10/abs/nrg2825.html>)

**Week 12- F statistics and demographic inference in the genomic era**

(Peter 2016, doi: 10.1534/genetics.115.183913)

The 'slog through it together' bioinformatics reading group

**Week 13- How to compare genes sets across species—XGSA**

(Djordjevic et al 2016, doi: 10.1093/bioinformatics/btw428)

**Week 14- Genome selection scans for adaptation—PCAdapt**

(Luu et al 2016, doi: 10.1111/1755-0998.12592)

**Week 15- Error correction across software packages and data platforms—pt. 2**

(Yang et al 2012, doi:10.1093/bib/bbs015)