Assessing the Utility of the Oxford Nanopore MinION for Snake Venom Gland cDNA Sequencing

Nicholas Perry

April 29, 2019

Professor Harbert

Background

- Hargeaves and Mulley sought to sequence cDNA from venom-encoding genes of *Echis coloratus*, the painted, saw-scaled viper

- Develop antivenom as a result

- Previous Illumina and Sanger not accurate

- Highly variable!

Background

- 2 Samples
  - Eco6 and Eco8
- Both run for 48 hours
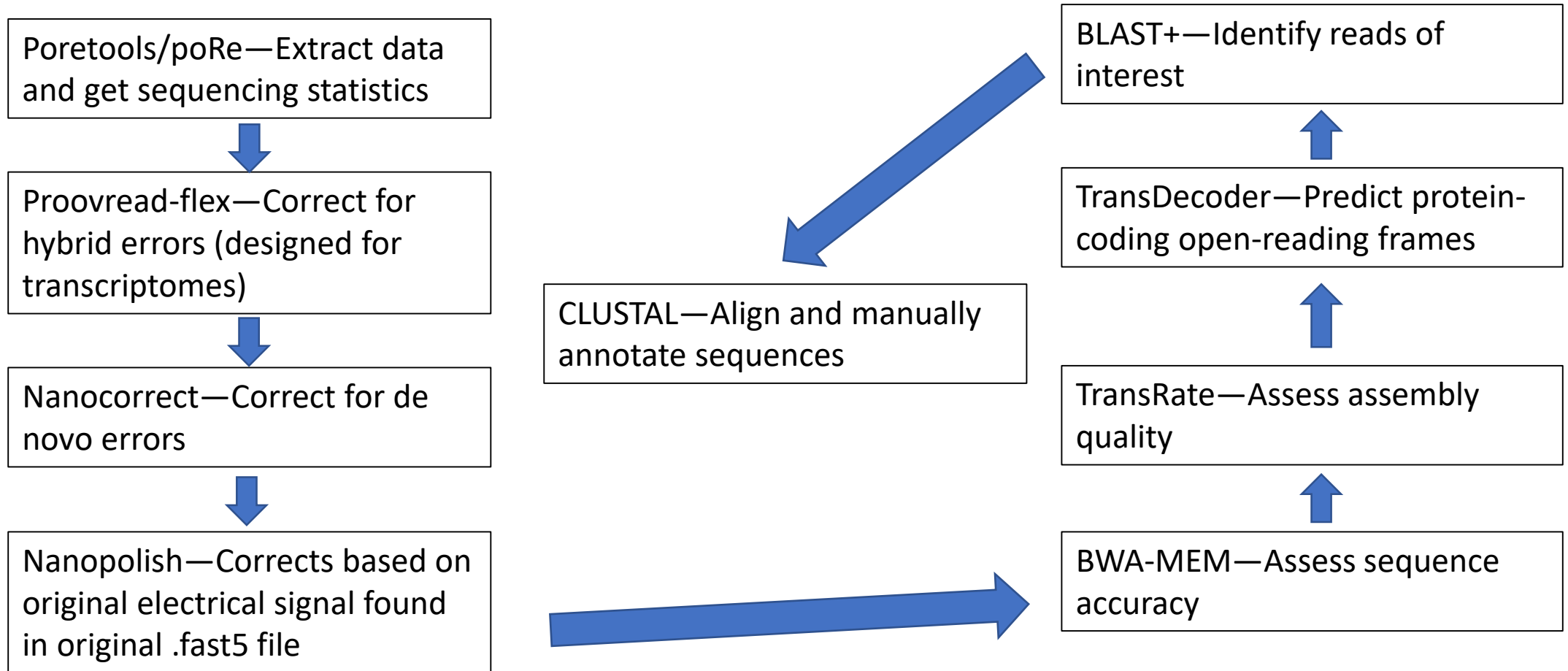- Both also run for 4x8 hours
  - Remux

https://nanoporetech.com/sites/default/files/s3/inline-images/minion-flowcell.png
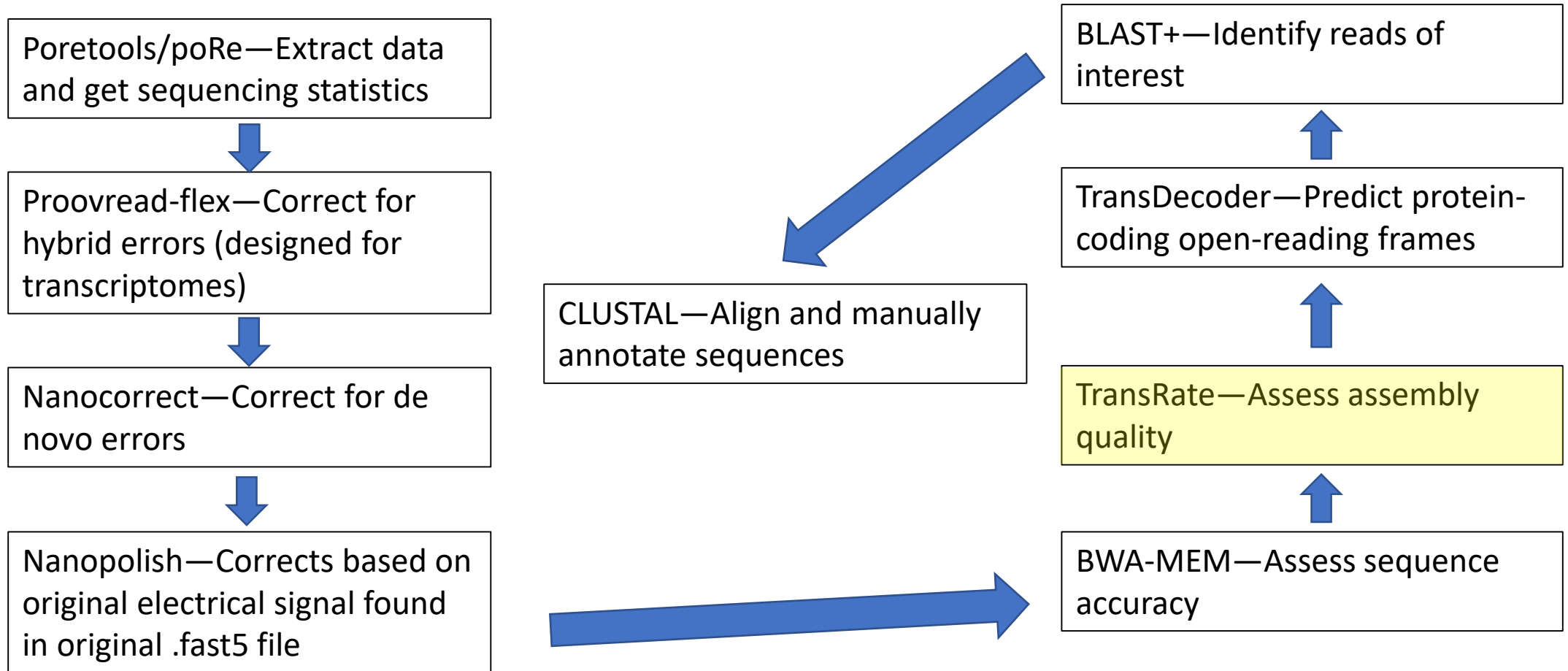
https://farm5.staticflickr.com/4597/24195096637_90455b9ed2_z.jpg

What I Hoped to Do

Poretools/poRe—Extract data and get sequencing statistics

Proovread-flex—Correct for hybrid errors (designed for transcriptomes)

Nanocorrect—Correct for de novo errors

Nanopolish—Corrects based on original electrical signal found in original .fast5 file

BWA-MEM—Assess sequence accuracy

TransRate—Assess assembly quality

TransDecoder—Predict protein-coding open-reading frames

BLAST+—Identify reads of interest

CLUSTAL—Align and manually annotate sequences

What I Hoped to Do

Poretools/poRe—Extract data and get sequencing statistics

Proovread-flex—Correct for hybrid errors (designed for transcriptomes)

Nanocorrect—Correct for de novo errors

Nanopolish—Corrects based on original electrical signal found in original .fast5 file

CLUSTAL—Align and manually annotate sequences

BLAST+—Identify reads of interest

TransDecoder—Predict protein-coding open-reading frames

TransRate—Assess assembly quality

BWA-MEM—Assess sequence accuracy

Why?

- Software Outdated!
  - Outdated basecaller, software discontinued
- No current method to accurately basecall old files!
- Had to use different software
  - Software from class

Ha'il, Saudi Arabia
© Ahmad Alshammary

http://www.reptarium.cz/content/photo_rd_04/Echis-coloratus-03000032067_01.jpg

Results

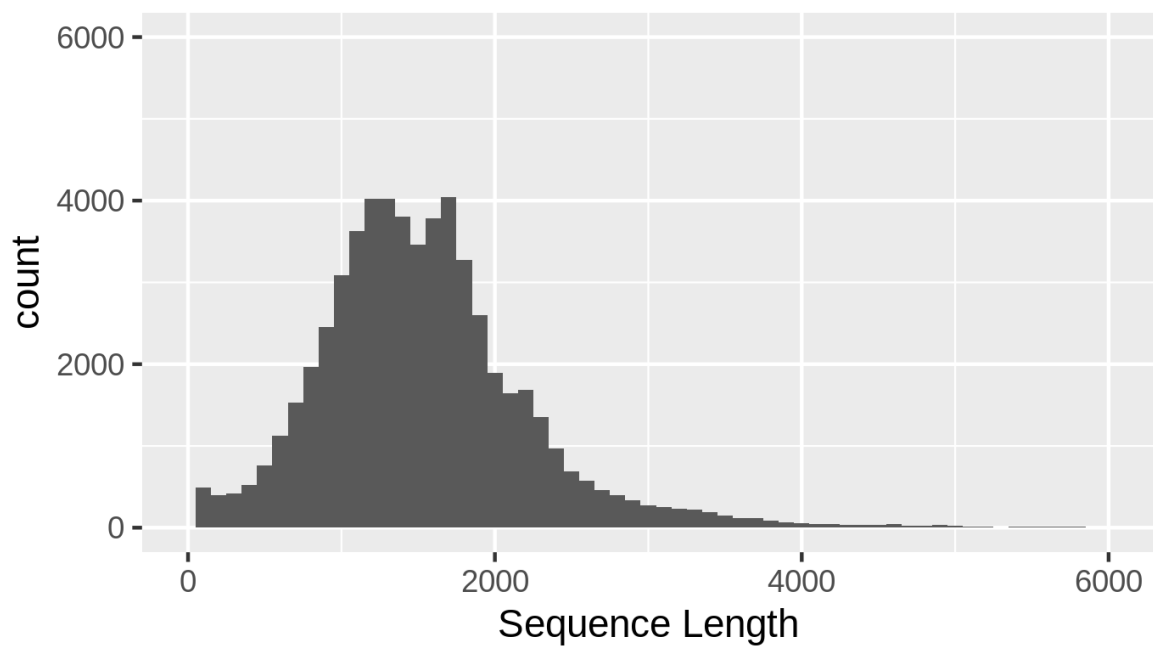|	Table 1	
	Eco6 (4 x 8)	Eco8
Total Reads	57,884	92,133
Total Bases (Mb)	89779650	154361646
Max Length (bp)	940362	5225
Min Length (bp)	1	1
Mean Length (bp)	1554	1,675.40
N50 (bp)	1,664	1,785

Results

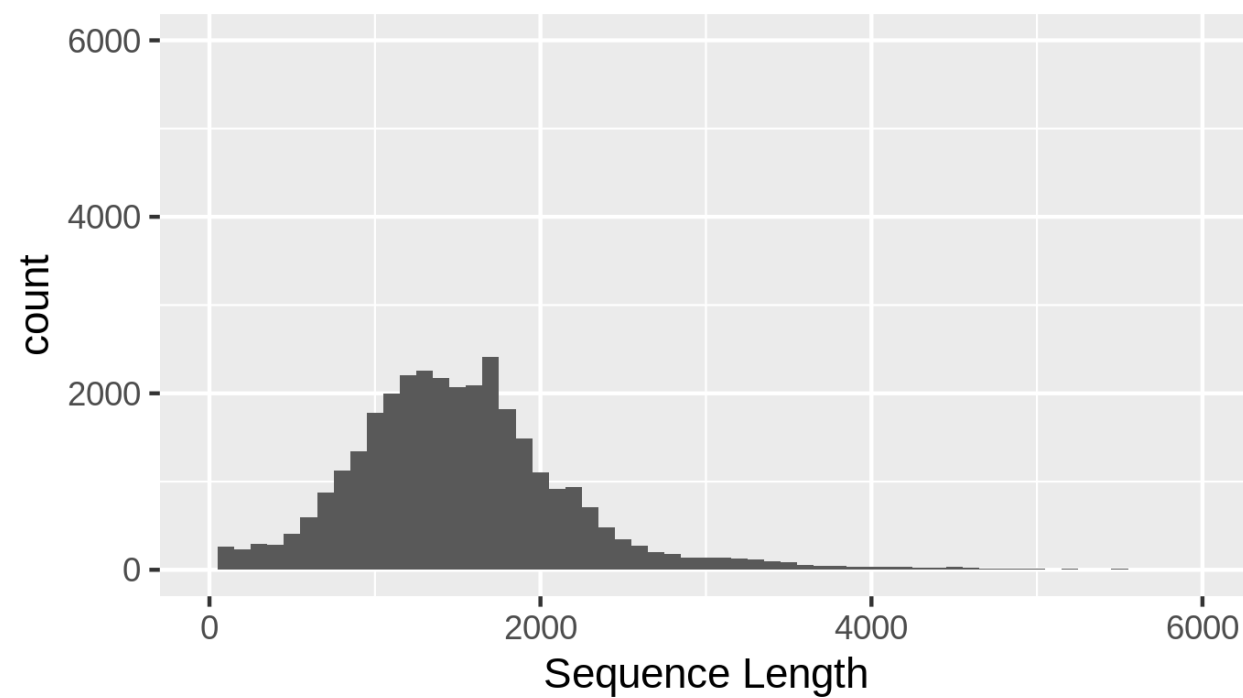Eco6_4x8 Read Length Distribution
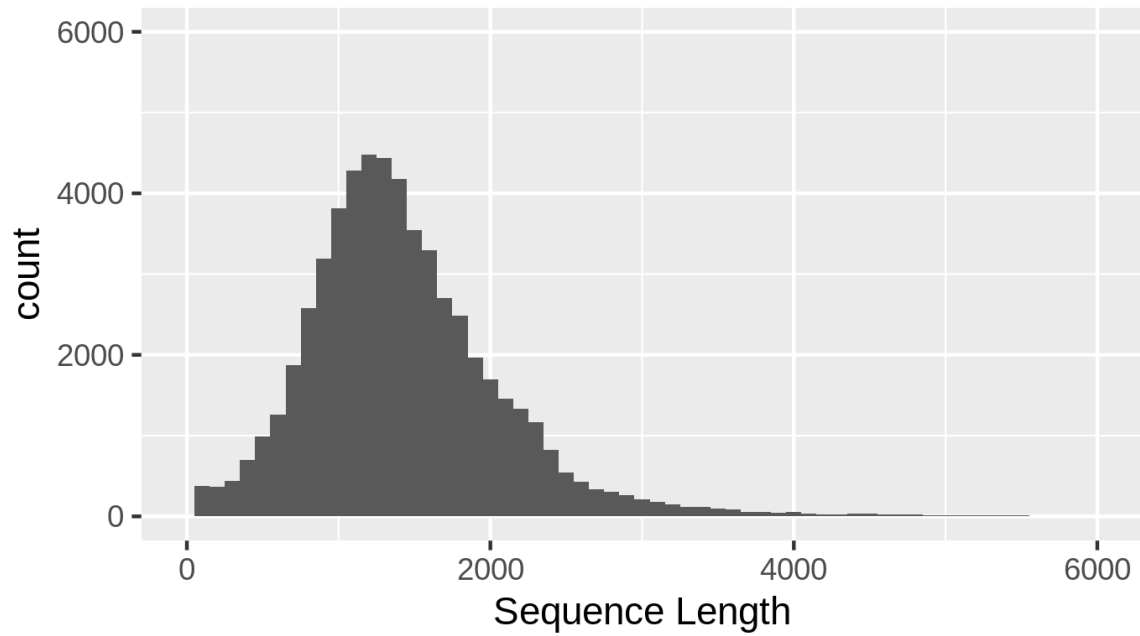
Eco6_48hr Read Length Distribution
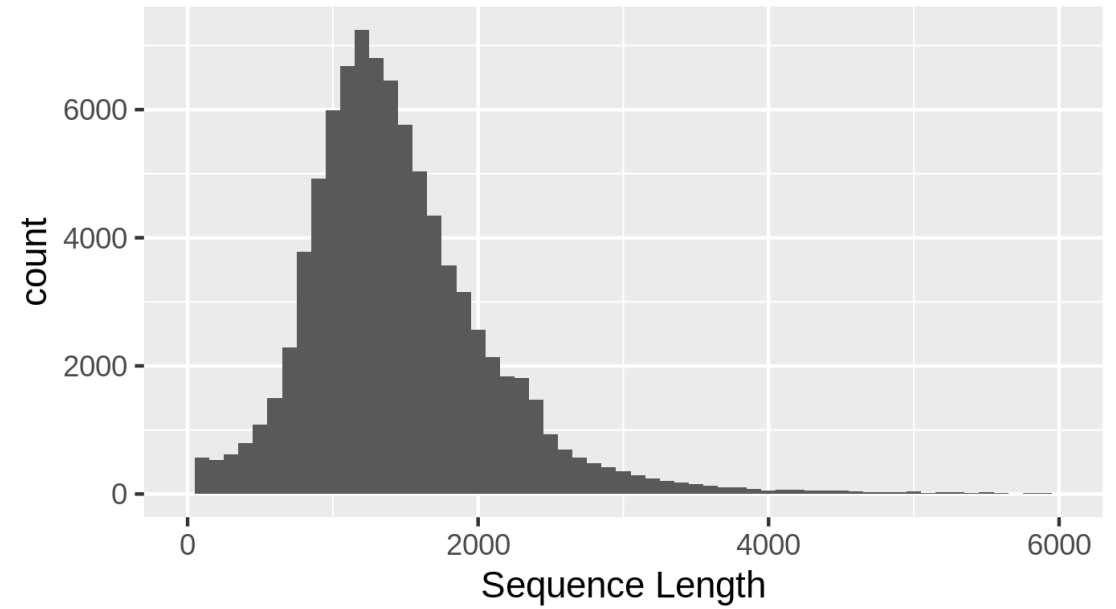
Results

Eco8_4x8 Read Length Distribution

Eco8_48hr Read Length Distribution

Results

	Table 2			
	Illumina Corrected	Illumina Corrected	Pilon Corrected	Pilon Corrected
	Eco6 (4 x 8)	Eco8	Eco6 (4 x 8)	Eco8
Total reads/Contigs	249	526	249	526
Read/Contig N50 (bp)	1,882	1899	1515	1862
Total Bases	449,307	971910	437856	948478
Mean Read Length	1,804.40	1848	1758.4	1803
Max Length (bp)	4434	4956	4326	4699
Min Length (bp)	1040	1046	1031	1018

Results

Protein Counts
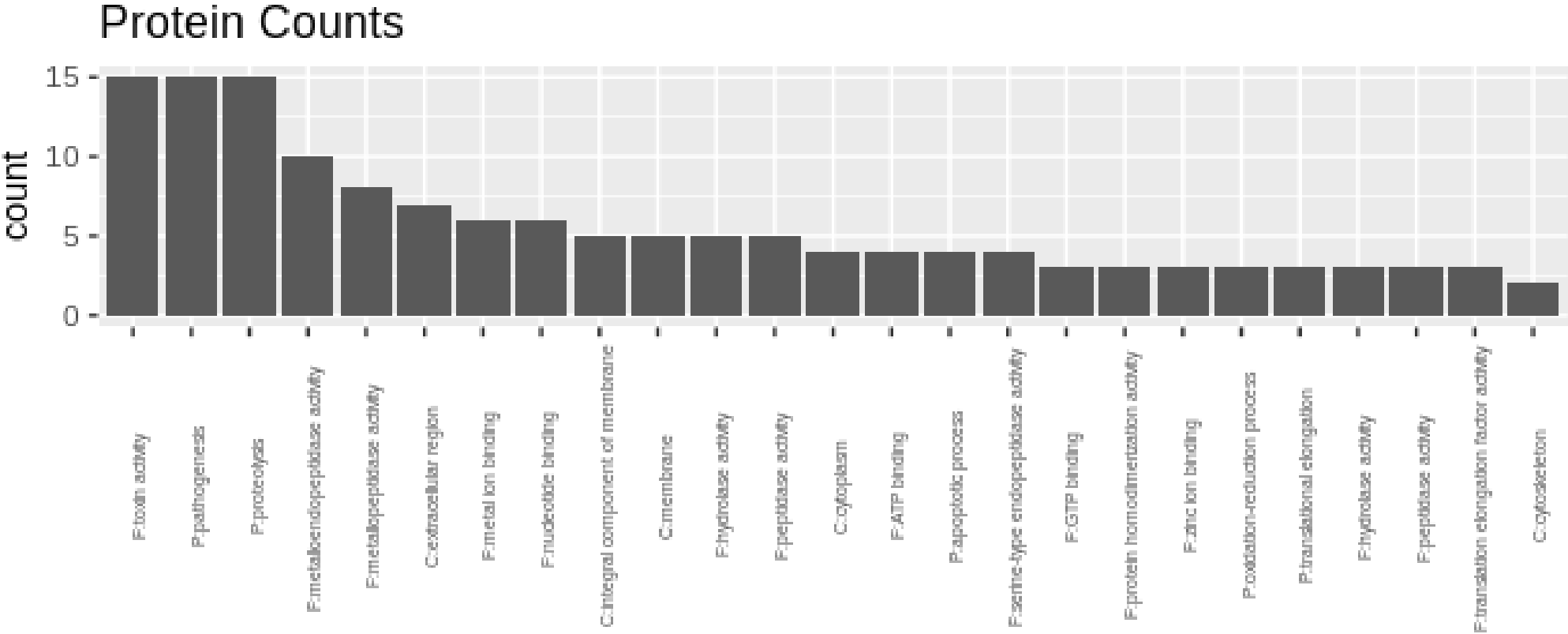
Conclusion

- How accurate was this data?
  - Pretty accurate given circumstances!
  - Metalloendopeptidase breaks alpha peptide bonds
- Problems
  - Poor read quality for some of the sequences
  - No fastq files!
  - Different read counts from original study
- Future
  - Get fastq files from original research

References

- Cao, M.D., Nguyen, S.H., Ganesamoorthy, D., Elliott, A.G., Cooper, M.A. and Coin, L.J., 2017. Scaffolding and completing genome assemblies in real-time with nanopore sequencing. Nature Communications, 8, p.14515.

- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), pp.3674-3676.

- De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M. and Van Broeckhoven, C., 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), pp.2666-2669.

- Hargreaves, A.D. and Mulley, J.F., 2015. Assessing the utility of the Oxford Nanopore MinION for snake venom gland cDNA sequencing. *PeerJ*, *3*, p.e1441.

- Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, *11*(1), p.119.

- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, *27*(5), pp.722-736.

- Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M. and Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research*, *26*(8), pp.1134-1144.

- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K. and Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, *9*(11), p.e112963.