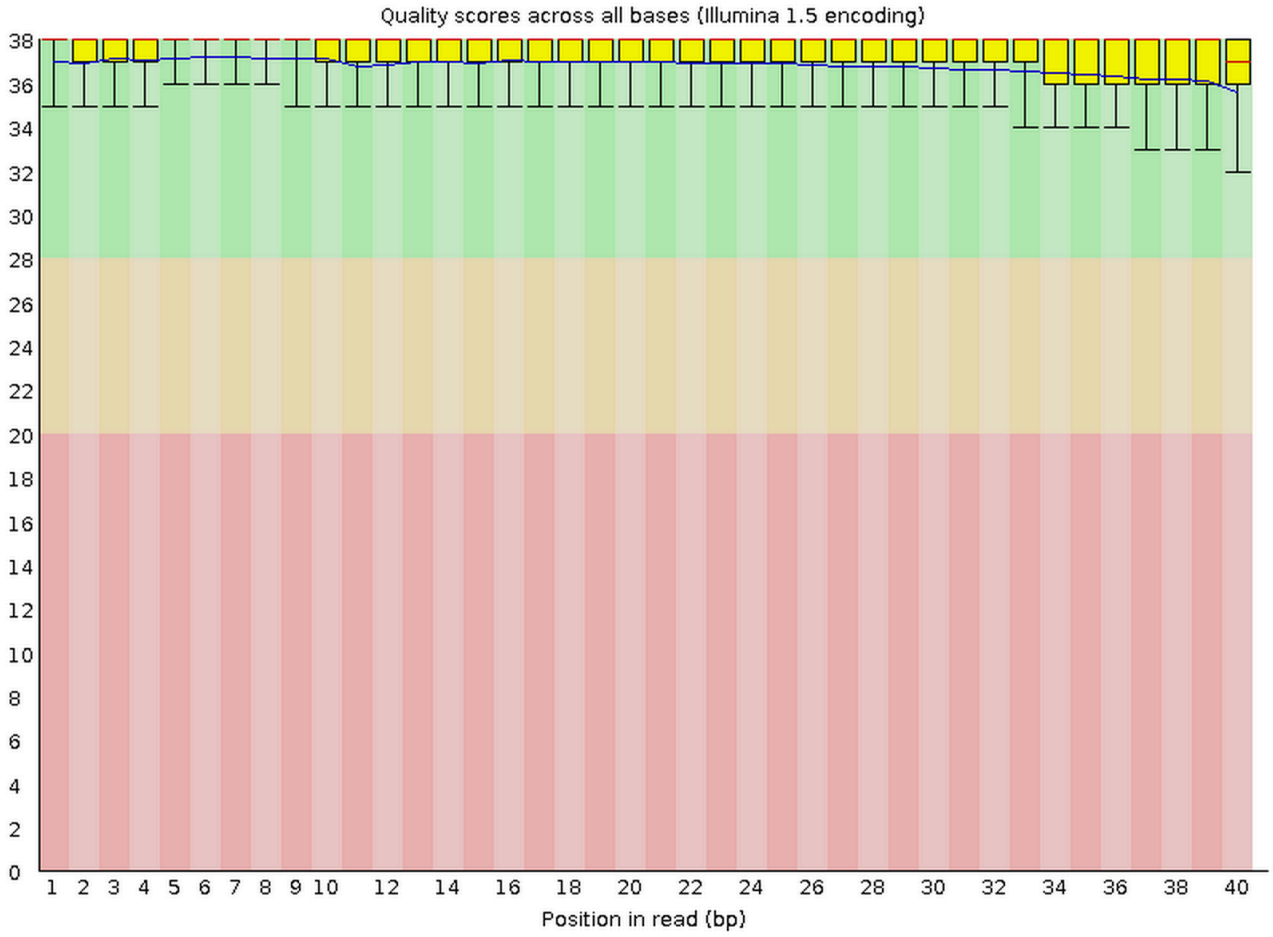
FastQC Report

Wed 25 Mar 2015
good_sequence_short.txt

Summary

Per base sequence quality

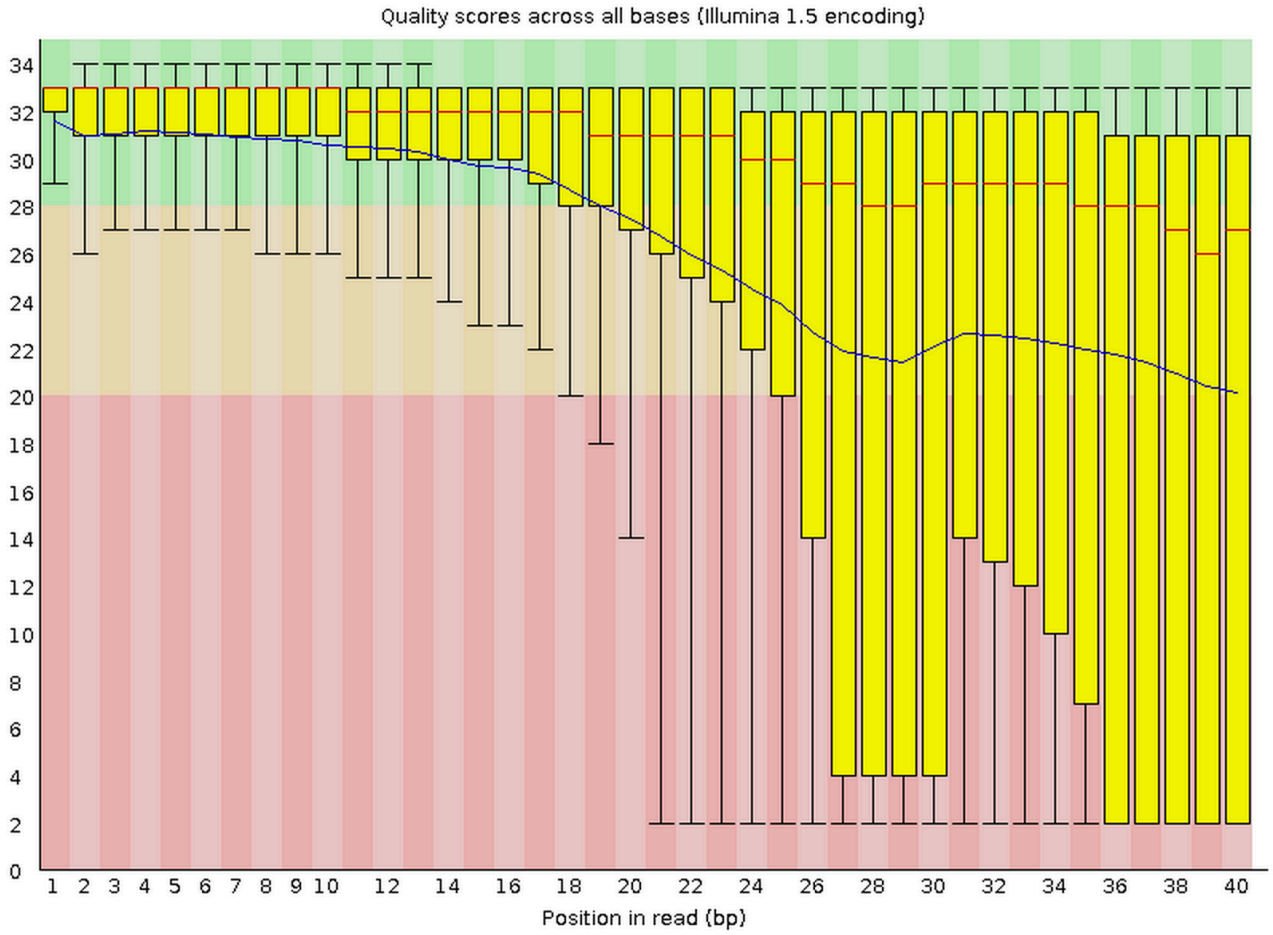


FastQC Report

Wed 25 Mar 2015
bad_sequence.txt

Summary

Per base sequence quality



What to do?

- Trim the reads?
- Start over – try sequencing it again?

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 31 January 2014
doi: 10.3389/fgene.2014.00013

On the optimal trimming of high-throughput mRNA sequence data

Matthew D. MacManes[1,2]*

[1] Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH, USA
[2] Hubbard Center for Genome Studies, Durham, NH, USA

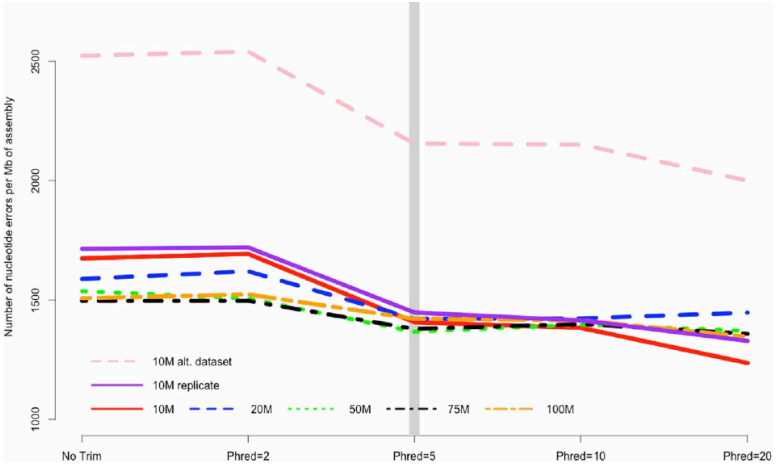
"… researchers interested in assembling transcriptomes de novo should elect for a much gentler quality trimming, or no trimming at all."

"… trimming at PHRED=2 or PHRED=5 optimizes assembly quality."

Aggressive Trimming may be harmful, whereas light trimming could be beneficial

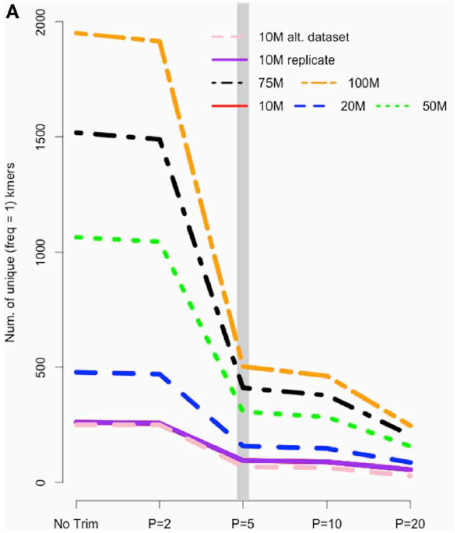
Fewer errors in the assembly

Nucleotide errors / Mb



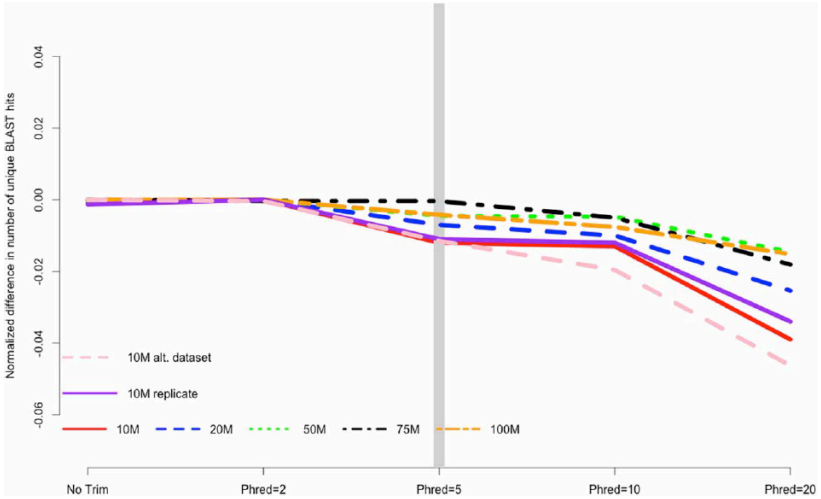
Fewer unique kmers

unique kmers



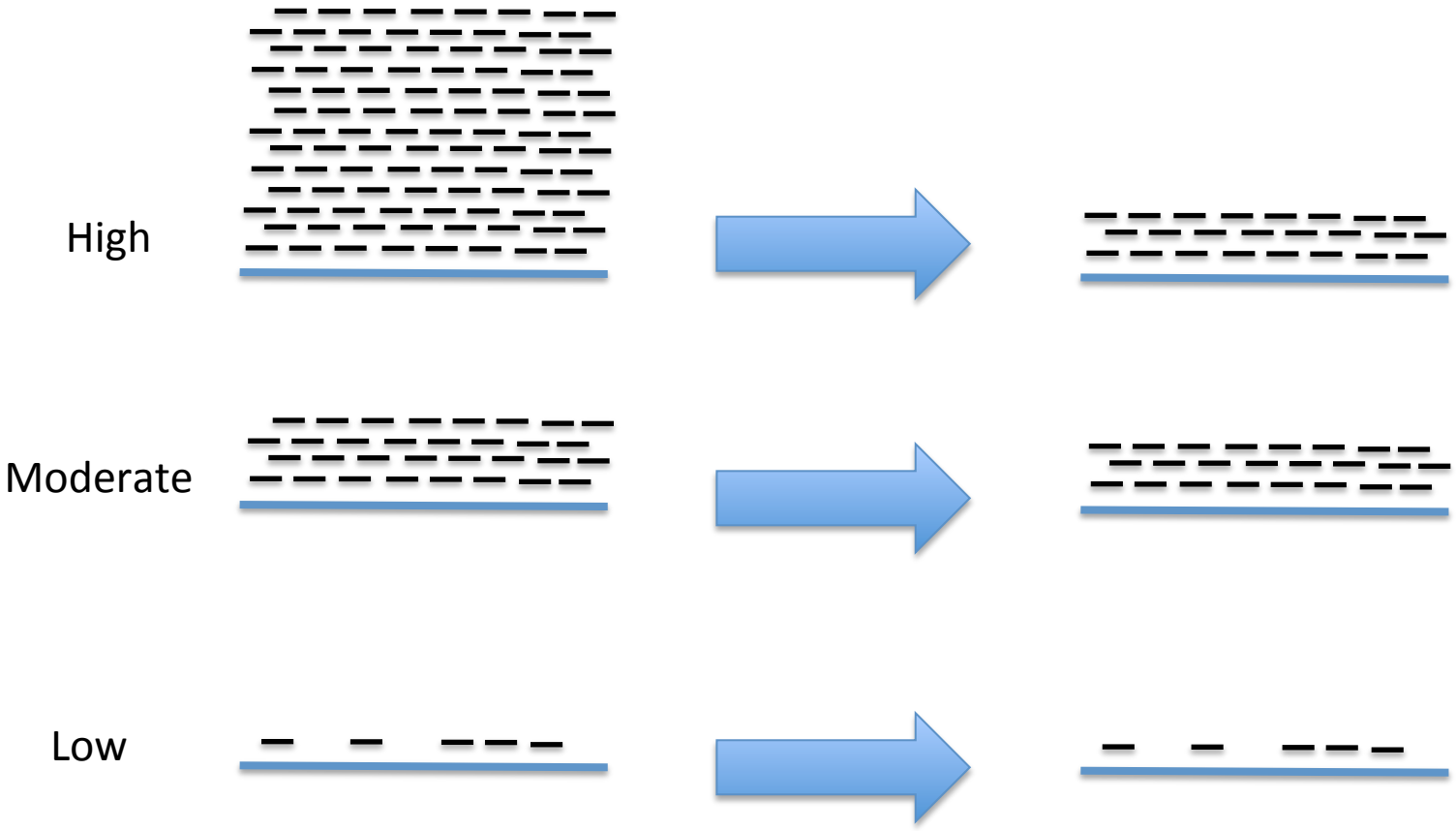
Light trimming doesn't reduce number of blast matches w/ higher sequencing depths.

Normalized # of blast matches

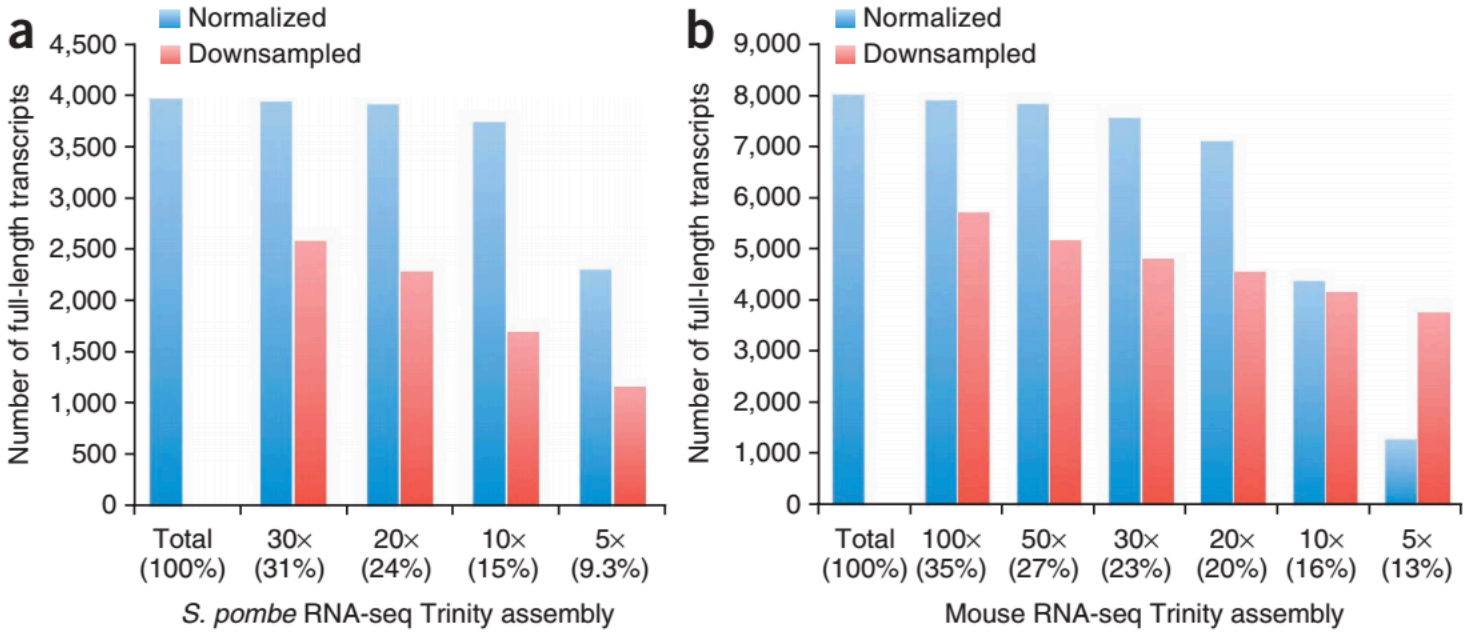


MacManes 2014

In silico normalization of reads



Impact of Normalization on *De novo* Full-length Transcript Reconstruction



Largely retain full-length reconstruction, but use less RAM and assemble much faster.

Haas et al., 2013

Quality Trimming and Normalization via Trinity

- Quality Trimming using Trimmomatic:
  - Trinity --trimmomatic
- Normalization of reads:
  - Trinity --normalize_reads *(now on by default!)*
- You can do both in a single Trinity assembly run:
  - Trinity --trimmomatic --normalize_reads

Fastqc, trimming, and normalization practical