Novo Aligners Benchmarking Document

Summary

Fast and accurate short read aligners for paired/single-end data
Comparable mapping error rates and higher good quality alignment yield
Format is compatible with MAQ tools

Computing Architecture

K=1 , S=2. Consuming at most 16.2 GB RAM for human genome alignments
Run on a Sun server with an Intel ®Xeon® CPU E5410 @ 2.33GHz having a 6,144 KB Cache size.
Binaries are also available for Intel®-based Macintosh OS X 10.5.3 and higher.

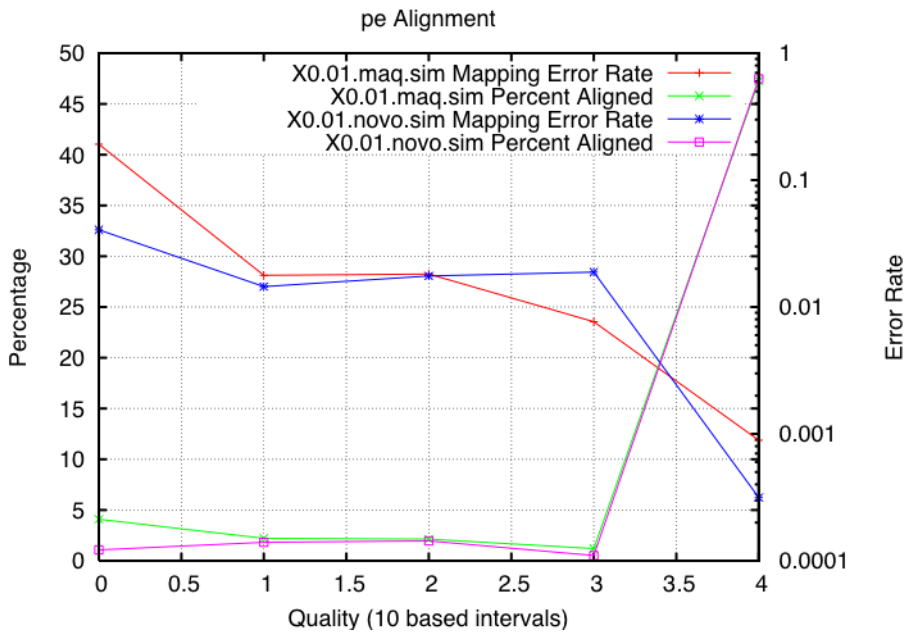
Results from Simulated Data

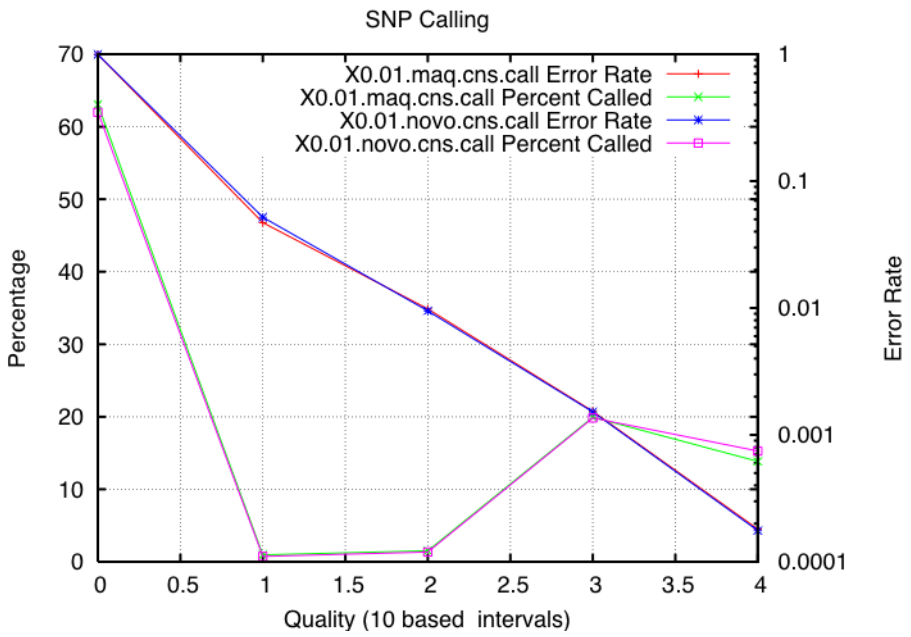
The MAQ tools on simulated human data to check indel and SNP calling.
We simulated 1000,000,000 pairs of reads from human chromosome 22 and mapped it back to a database consisting of chr21 and chr22.
A range of mutation rates were used. The fraction of 1bp indels allowed was 0.5. The data was simulated from quality trained on 200x36x37-071207_EAS51_0064-s_2_1.fastq available at ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/SRA000271/fastq/

On the next pages are some figures for SNP calling (generated using maq simucns) at various mutation rates.

Copyright Novocraft Technologies August 2008 www.novocraft.com

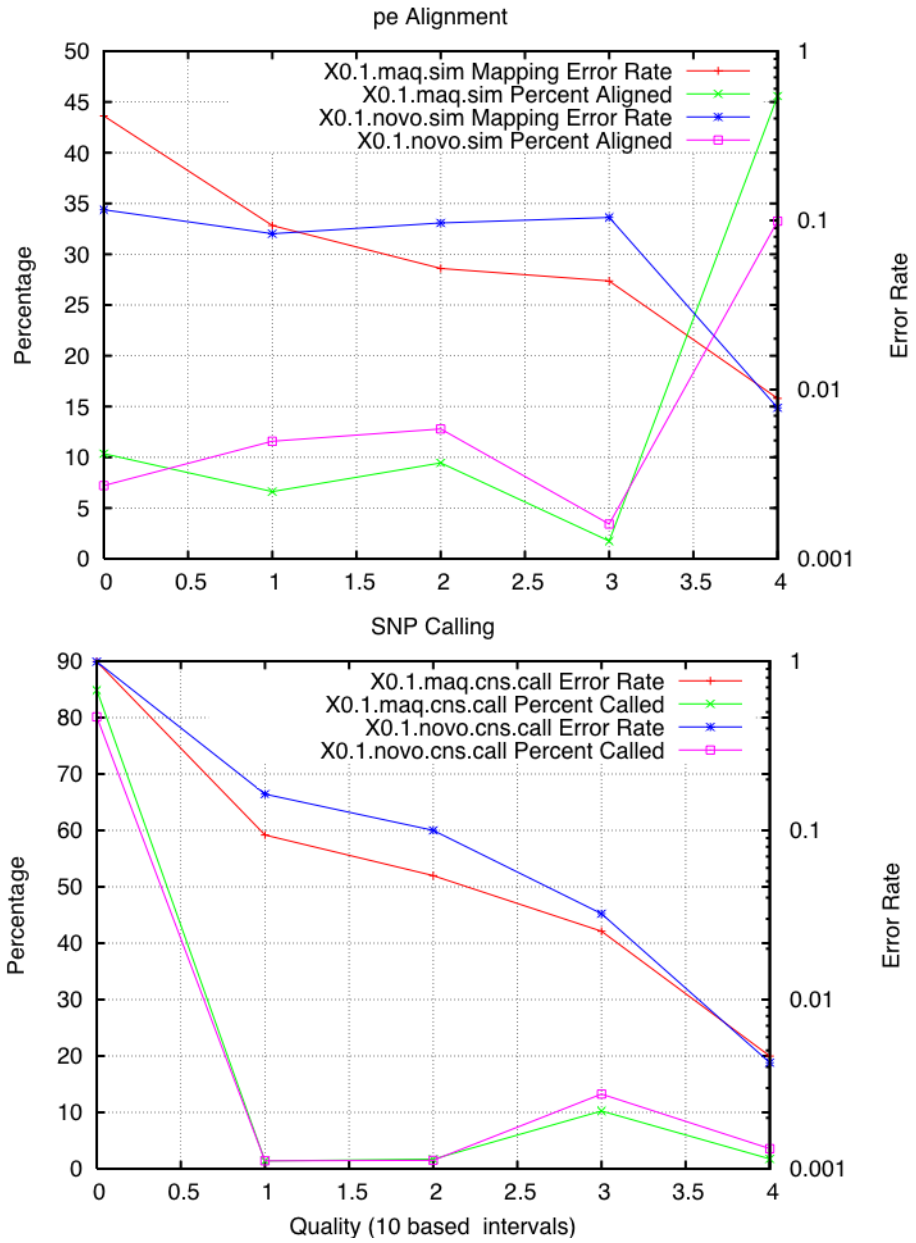
1% Mutation Rate





Copyright Novocraft Technologies August 2008 www.novocraft.com

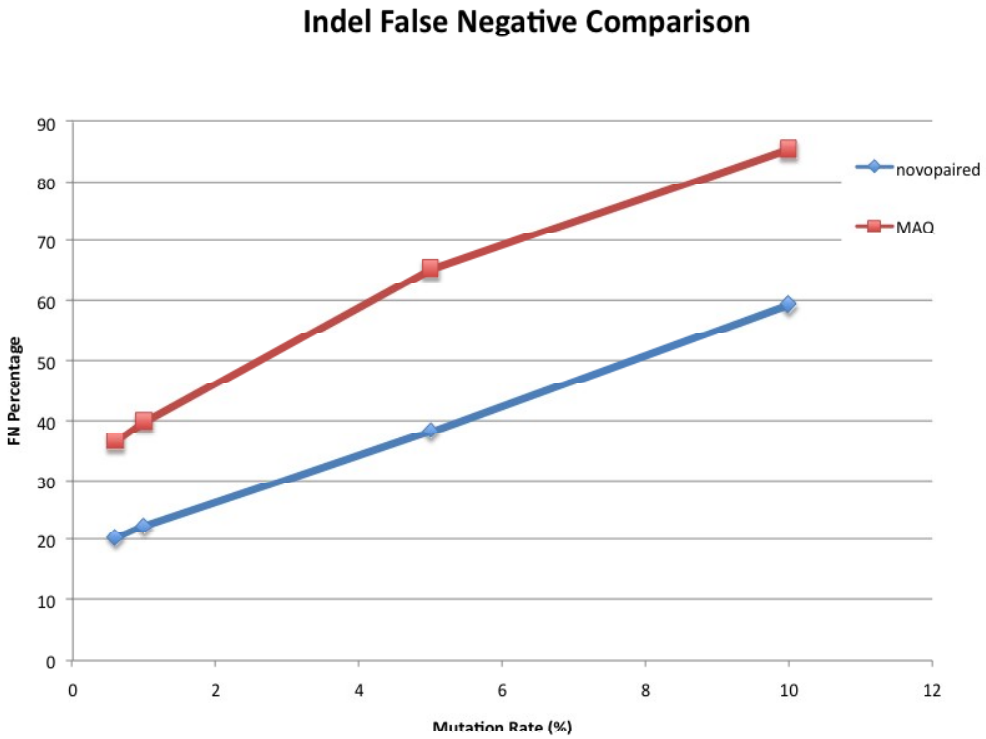
10% Mutation Rate



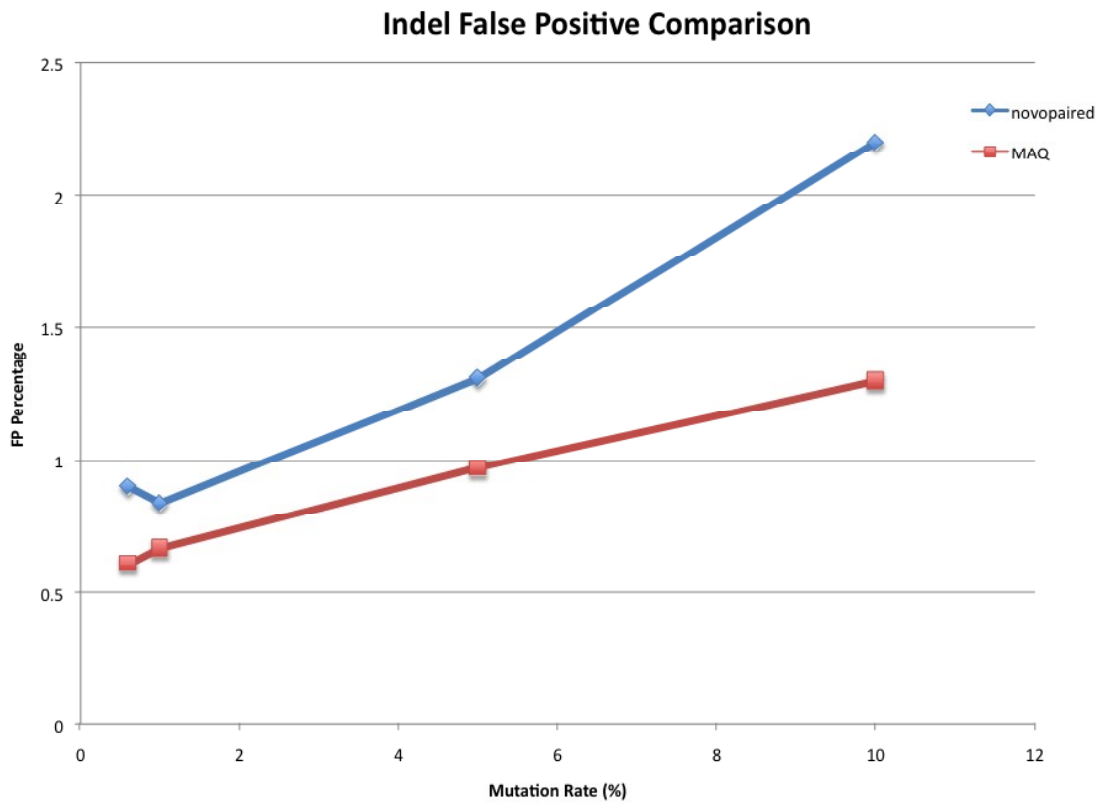
Copyright Novocraft Technologies August 2008 www.novocraft.com

Comparison of Indels

MAQ and Novopaired results were converted to maq binary format and further processed to report indels using the "maq indelpe" command.



Copyright Novocraft Technologies August 2008 www.novocraft.com



Conclusions

Novoalign and novopaired align a higher percentage of reads to unique locations than the other aligners tested. They also identify more true positive indel mutations than maq (the only tool compared for indel calling).

Copyright Novocraft Technologies August 2008 www.novocraft.com