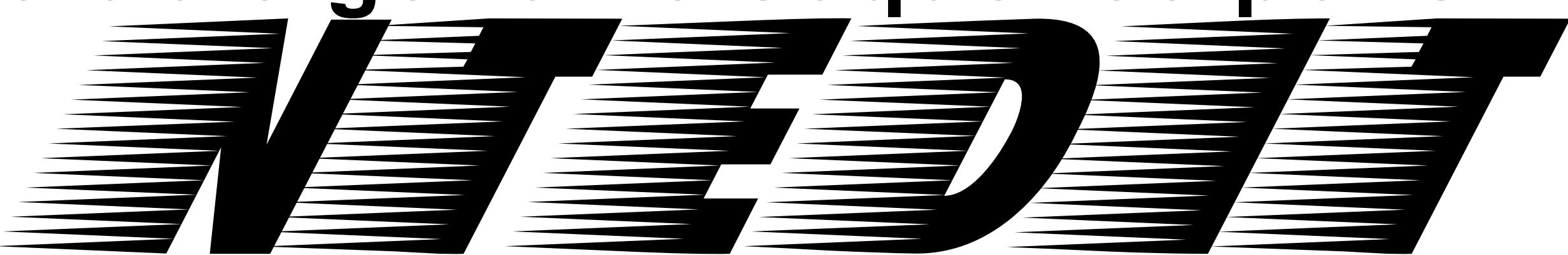
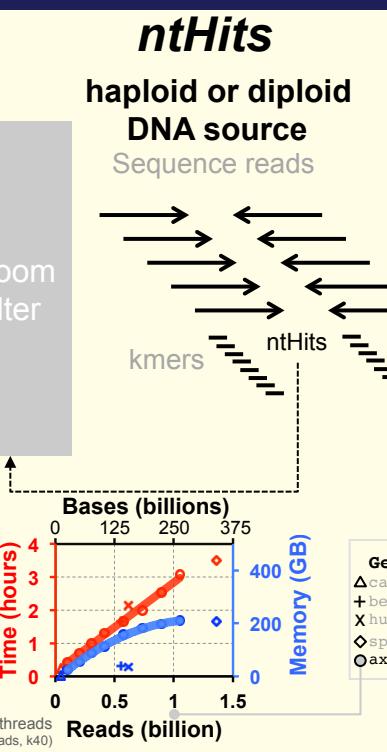
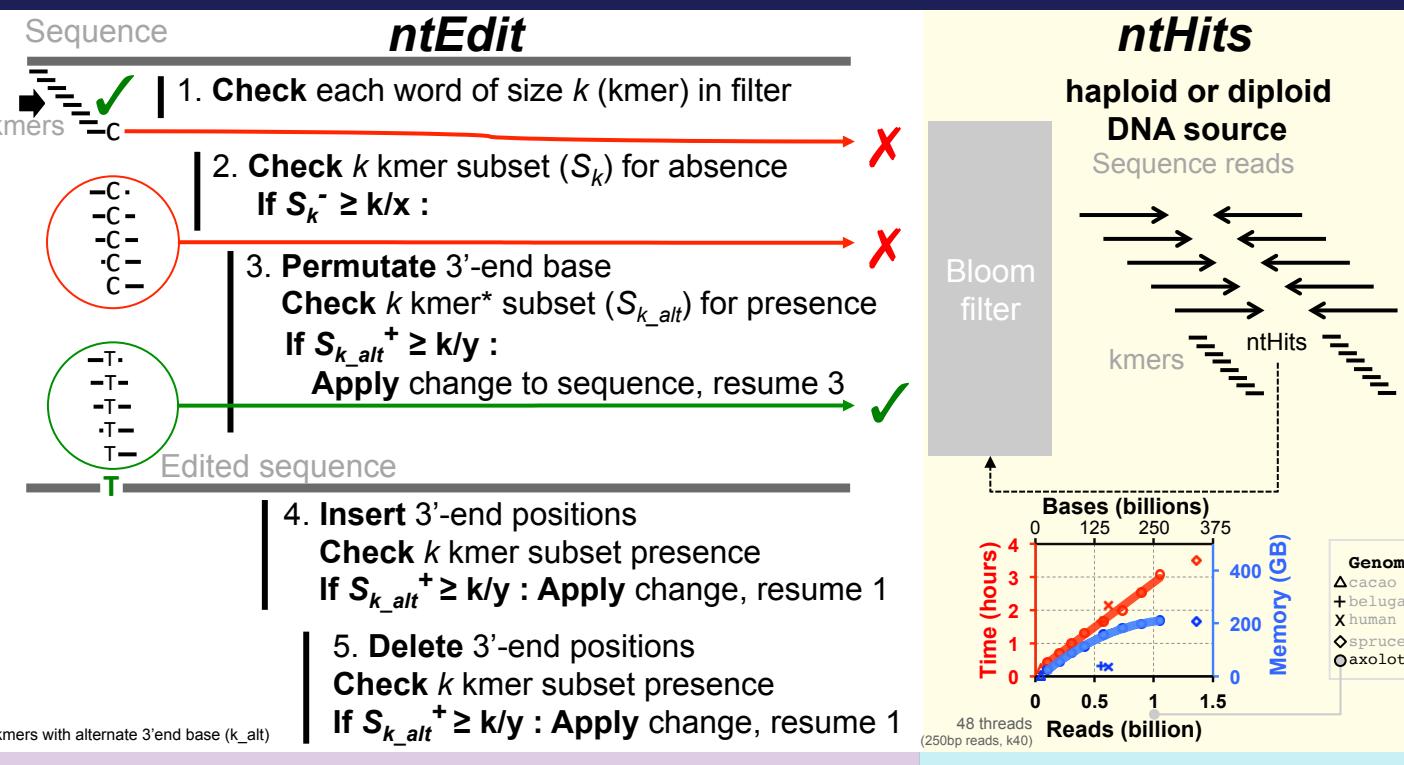


scalable genome sequence polishing



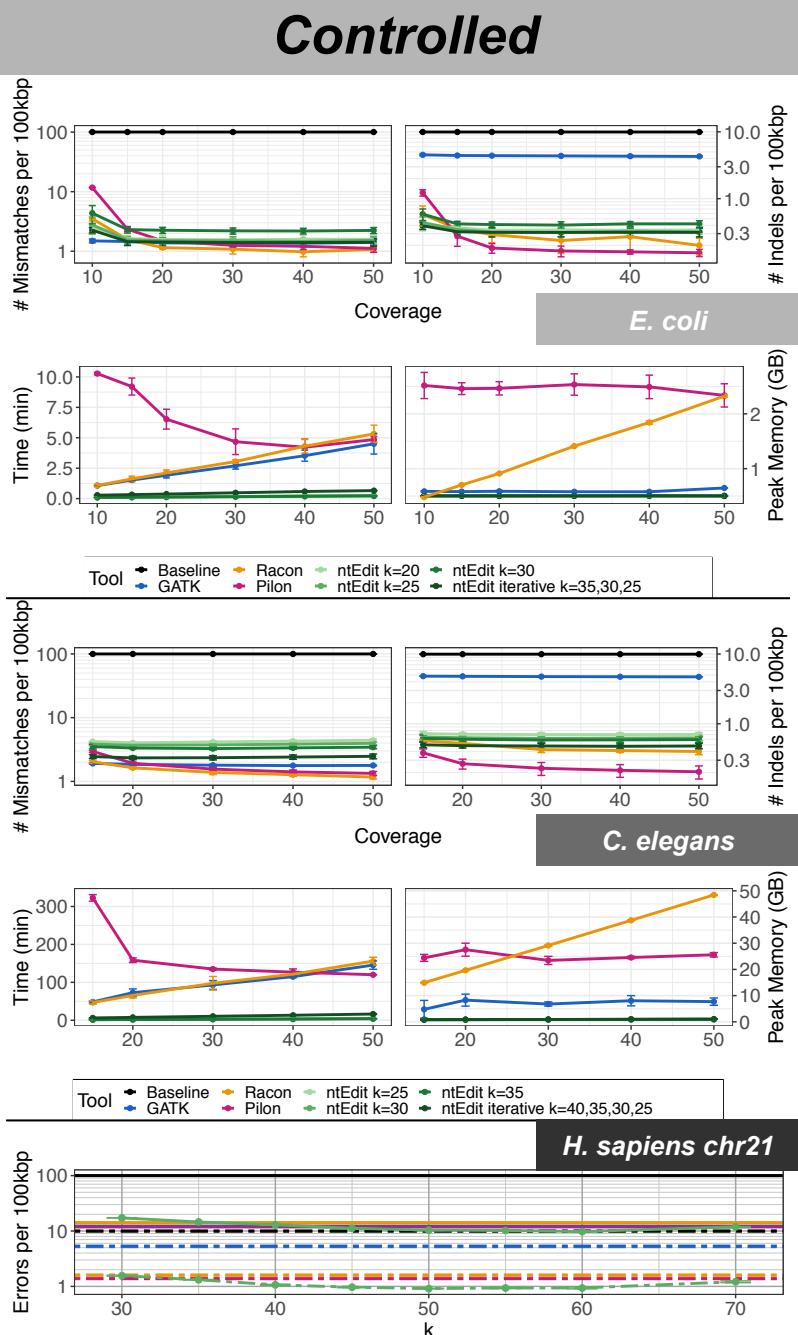
René Warren • Jessica Zhang • Lauren Coombe • Hamid Mohamadi • Inanç Birol

Method



Results

Controlled



Experimental

I. Polishing

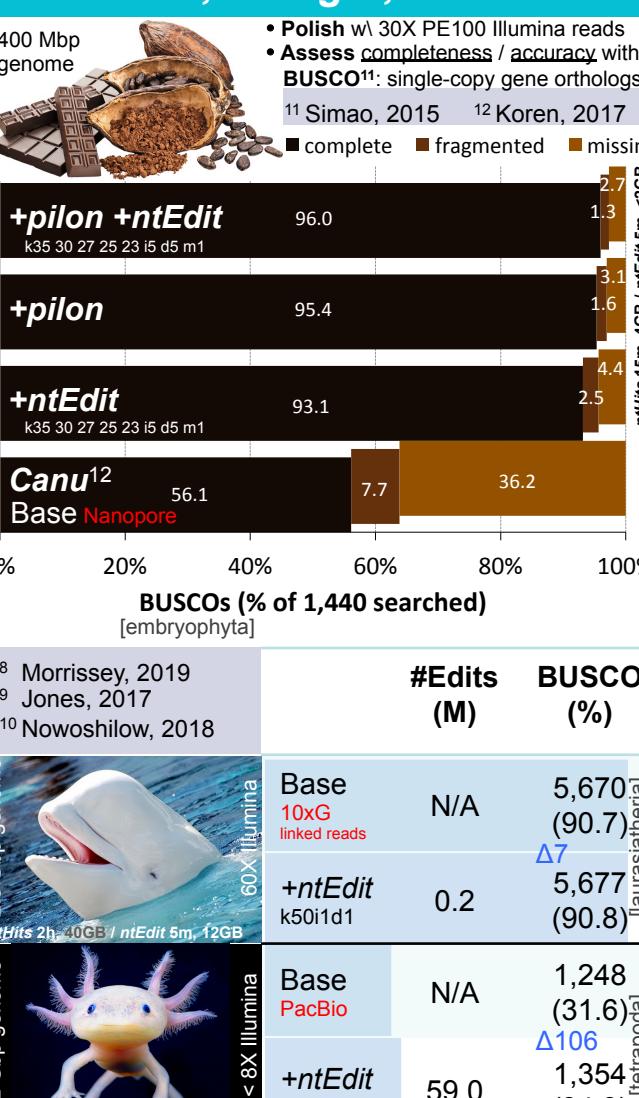
Human*

Polish w/ 54X Illumina	Time**	Edits (M)	BUSCO (%)	Baseline
GATK ⁴	41h45	0.97	5,654 (91.3)	7
ntEdit ^{***} k50i3d3	2h18	0.95	5,670 (91.6)	23
Racon ⁵	45h54	N/A	5,681 (91.7)	34
GATK	42h21	2.66	5,285 (85.4)	0
ntEdit ^{***} k40i3d3	2h10	3.63	5,651 (91.3)	36
Racon	40h55	N/A	5,670 (91.6)	35

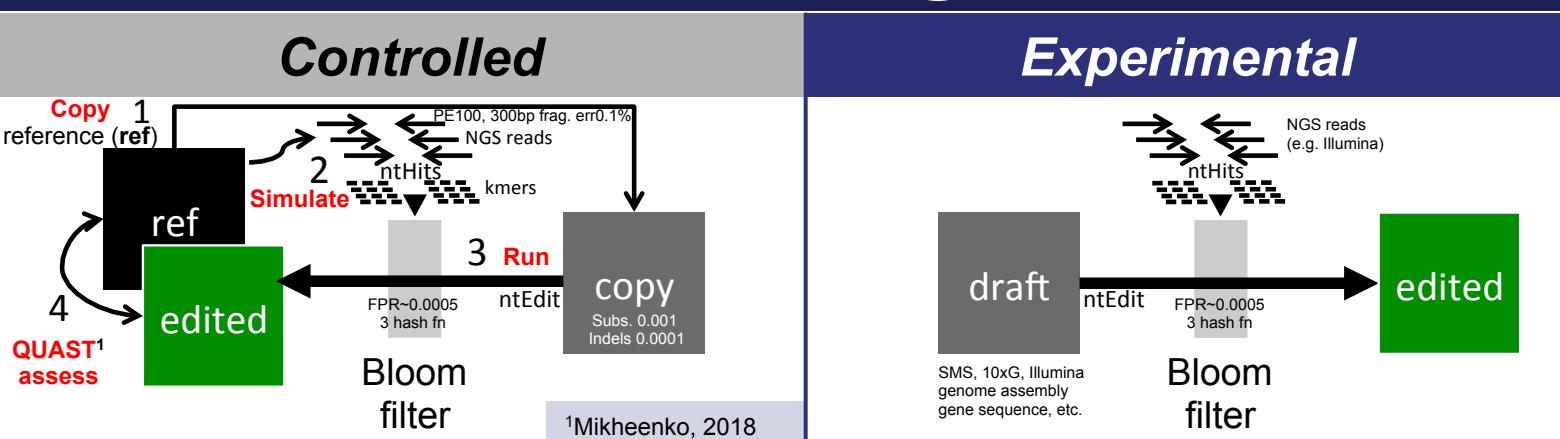
⁴McKenna, 2010 ⁵Vaser, 2017 ⁶Jain, 2018 ⁷Pendleton, 2015

*Single Molecule Sequencing draft genomes **Time for pipeline ***15GB RAM ntEdit 40GB

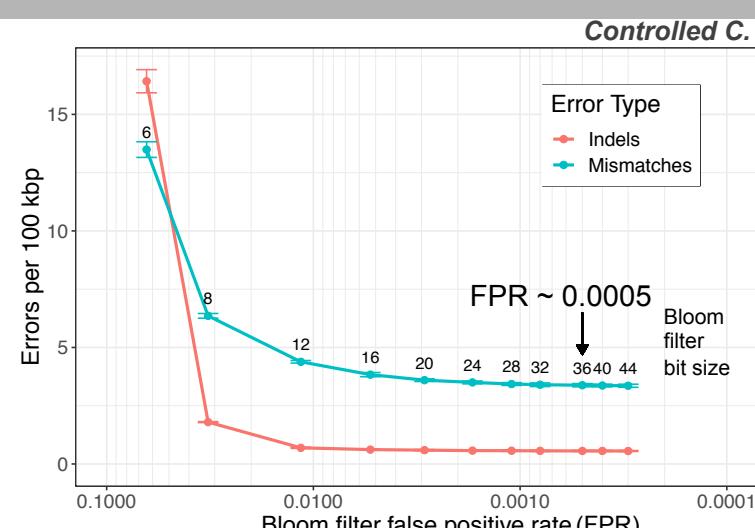
Cacao⁸, Beluga⁹, Axolotl¹⁰



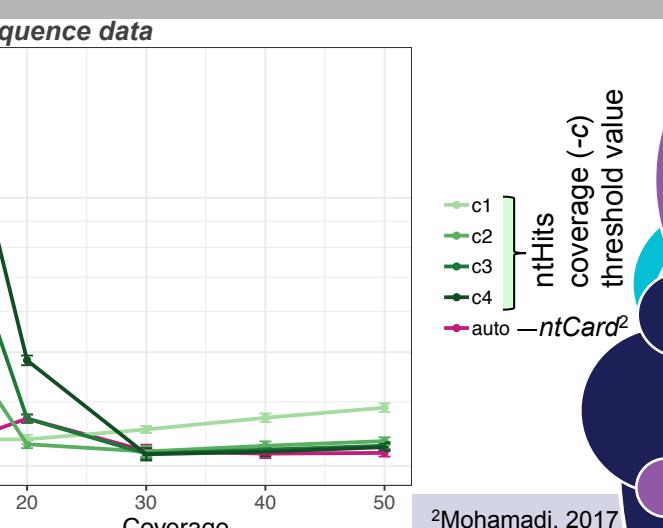
Testing



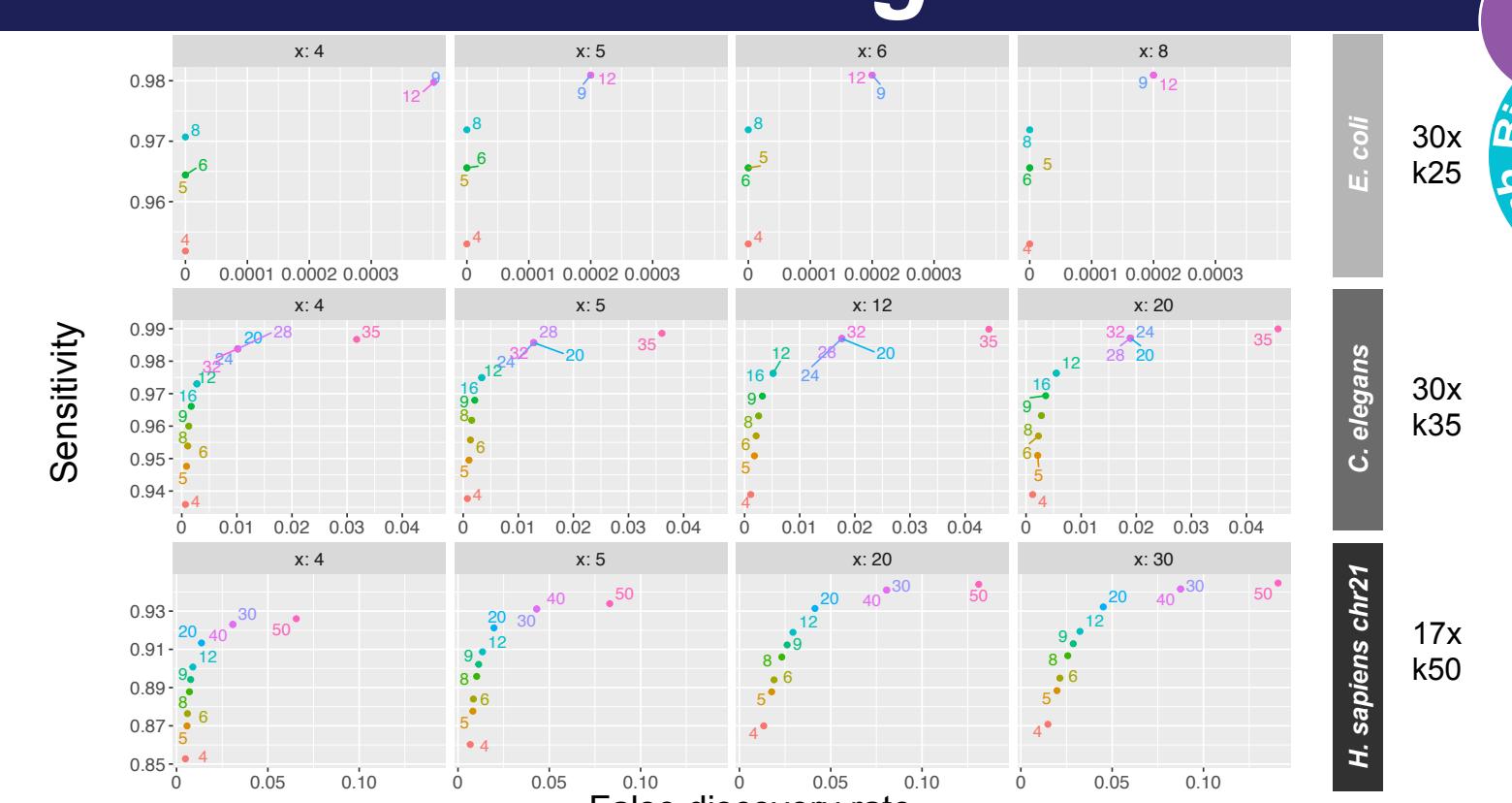
Effect of Bloom filter FPR



Threshold error kmers

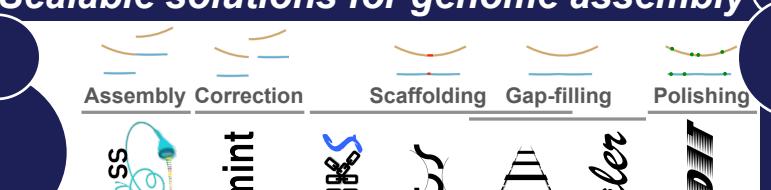


Tuning



Summary

Scalable solutions for genome assembly



Read Technology

Short

Linked

Long

Illumina, SMS drafts (Nanopore/PacBio)

<https://github.com/bcgsc/bioinformatics-lab>

<https://github.com/bcgsc/ntedit>

<https://github.com/bcgsc/nthits>

<https://github.com/bcgsc/ntedit>

[https://github.com/bcgsc/ntedit</](https://github.com/bcgsc/ntedit)