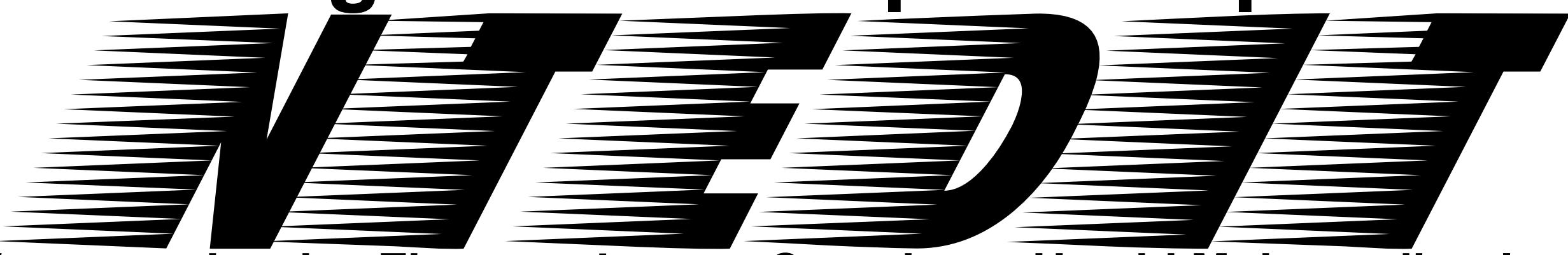
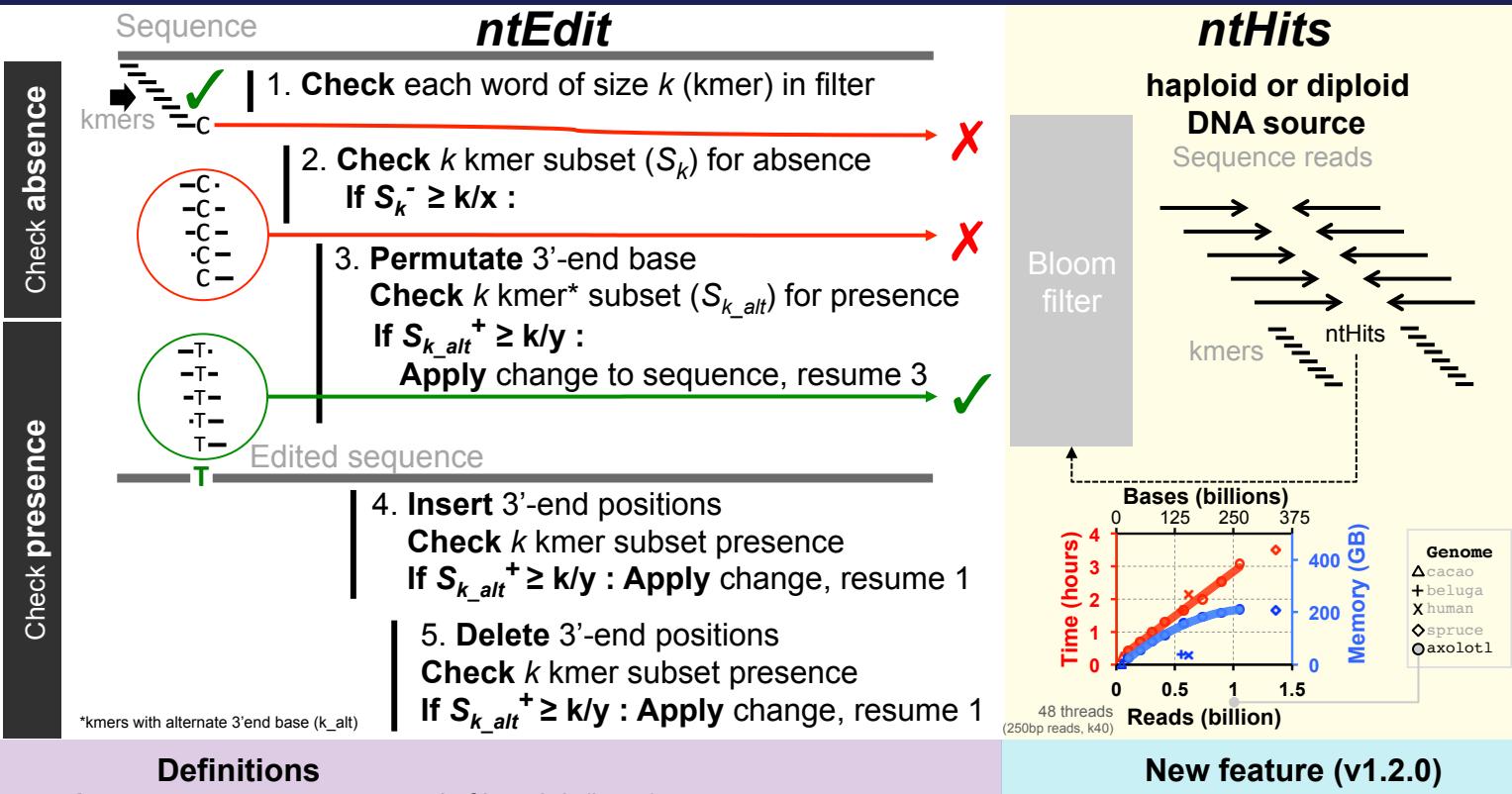


scalable genome sequence polishing



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

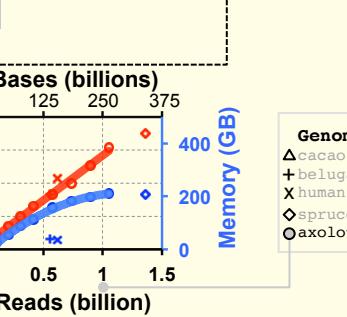
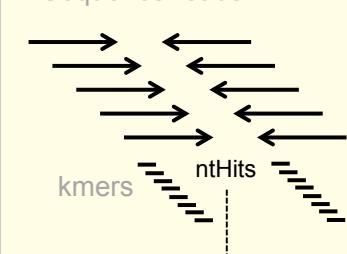
Method



ntHits

haploid or diploid DNA source

Sequence reads



Tool: Baseline, GATK, Pilon, ntEdit (k=20, 30, 35, 40, 30, 25)

Time (min)

Coverage

Mismatches per 100kbp

Indels per 100kbp

Peak Memory (GB)

Reads (billion)

Time (hours)

Bases (billions)

Memory (GB)

Genome: Δ *cacao*, \diamond *beluga*, \times *human*, \diamond *spruce*, \diamond *axolotl*

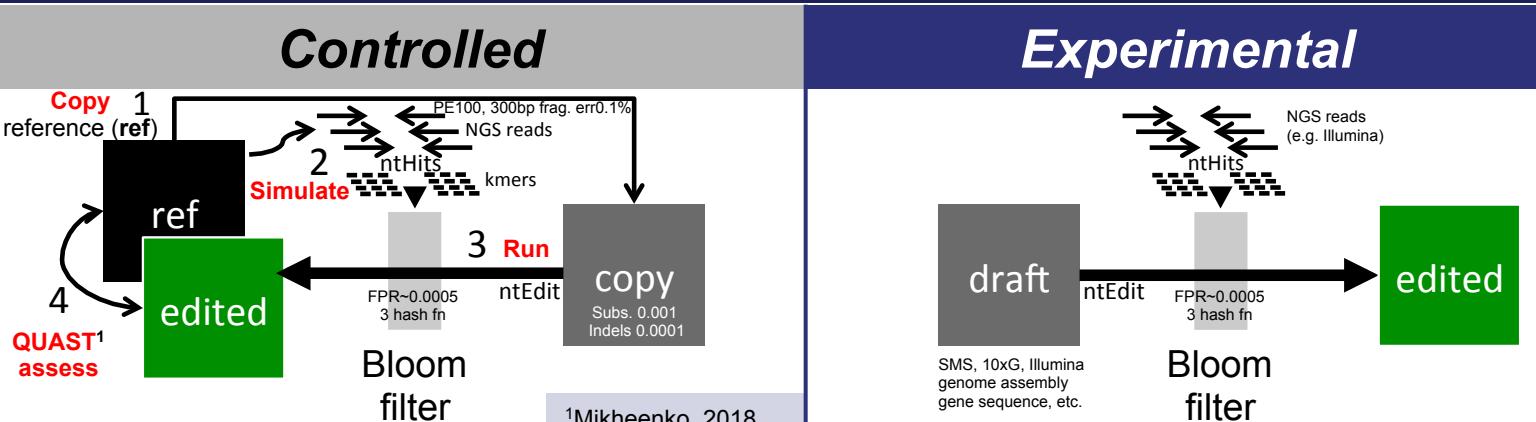
New feature (v1.2.0)

-m option editing mode 0-2 [default=0]

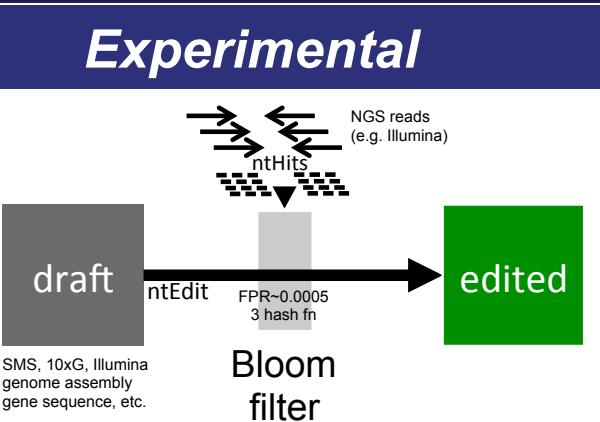
0: best substitution, or first supported indel
1: best substitution, or best indel
2: best edit overall (exhaustive)

Testing

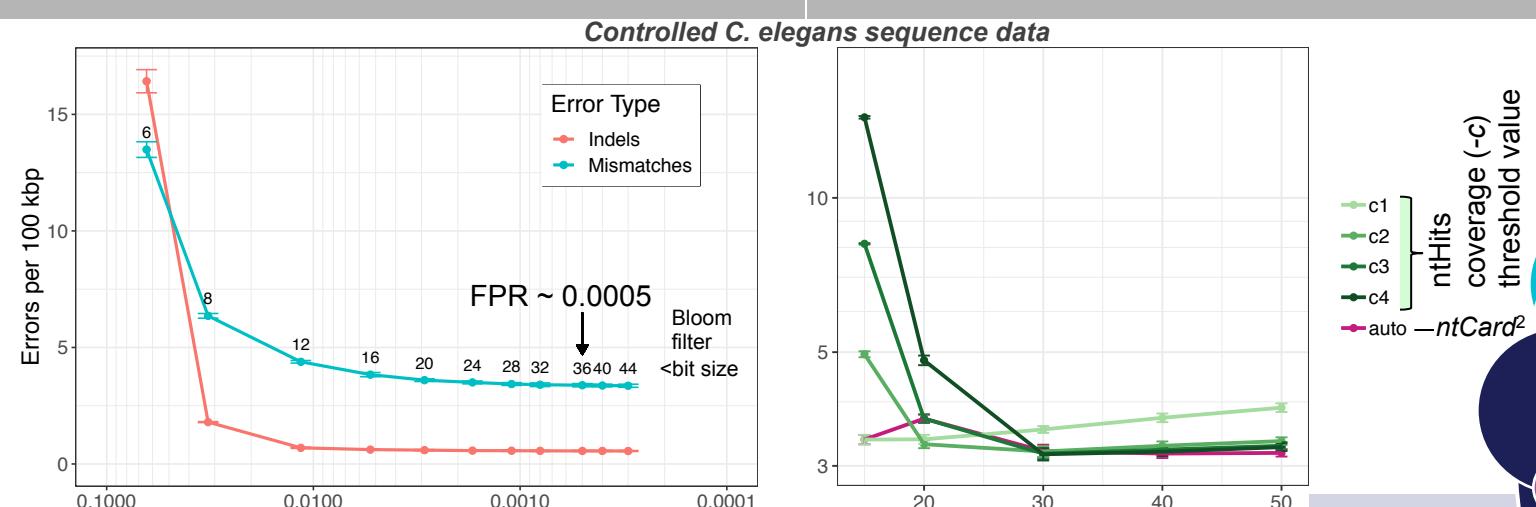
Controlled



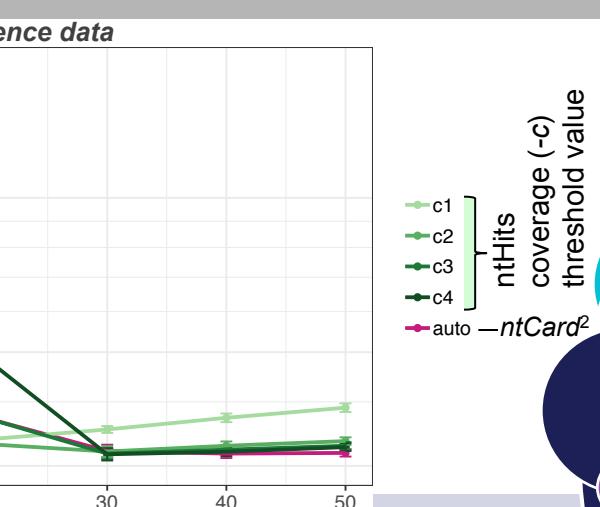
Experimental



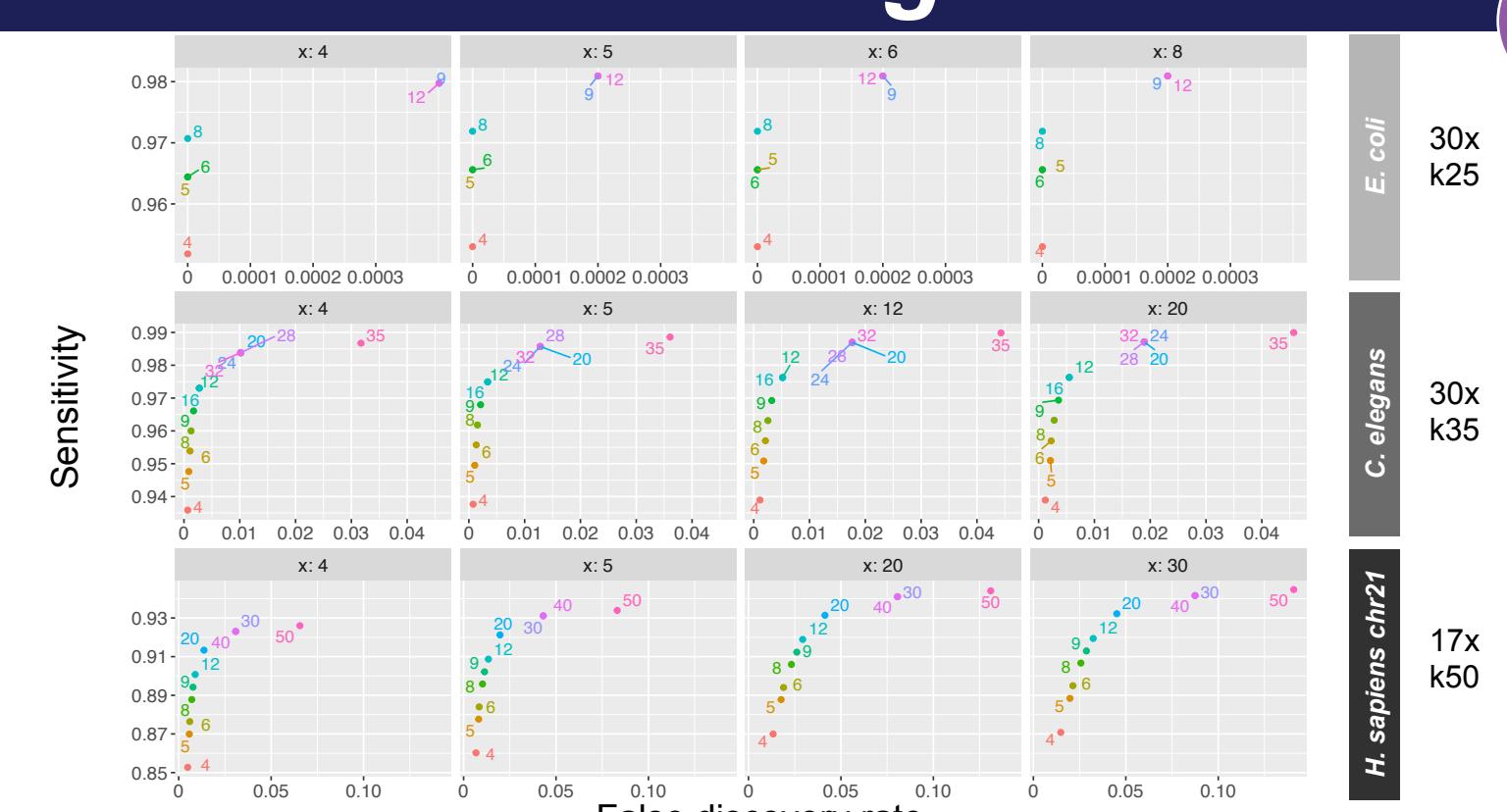
Effect of Bloom filter FPR



Threshold error kmers



Tuning



Reference

Warren et al. 2019. *Bioinformatics*. DOI: 10.1093/bioinformatics/btz400

Software

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

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

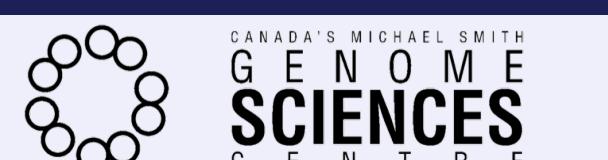
Funding

NIH

Genome British Columbia

Genome Canada

BC CANCER FOUNDATION



<http://birol-lab.ca>

<http://renewarren.ca>