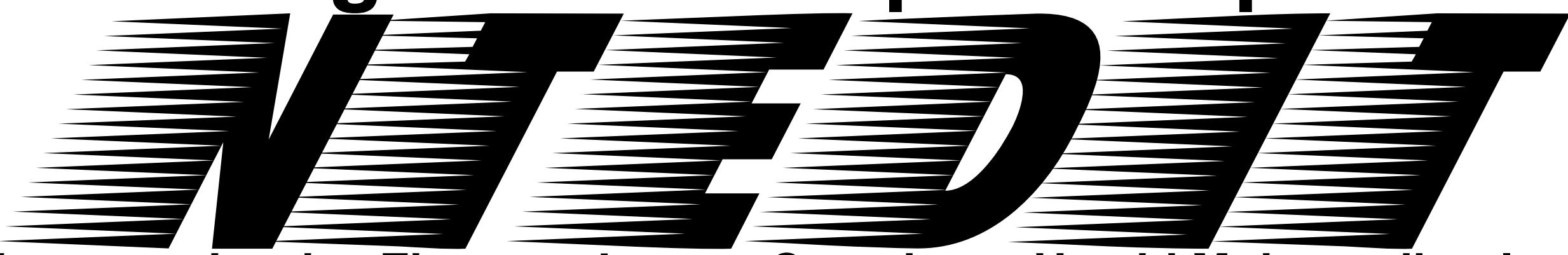
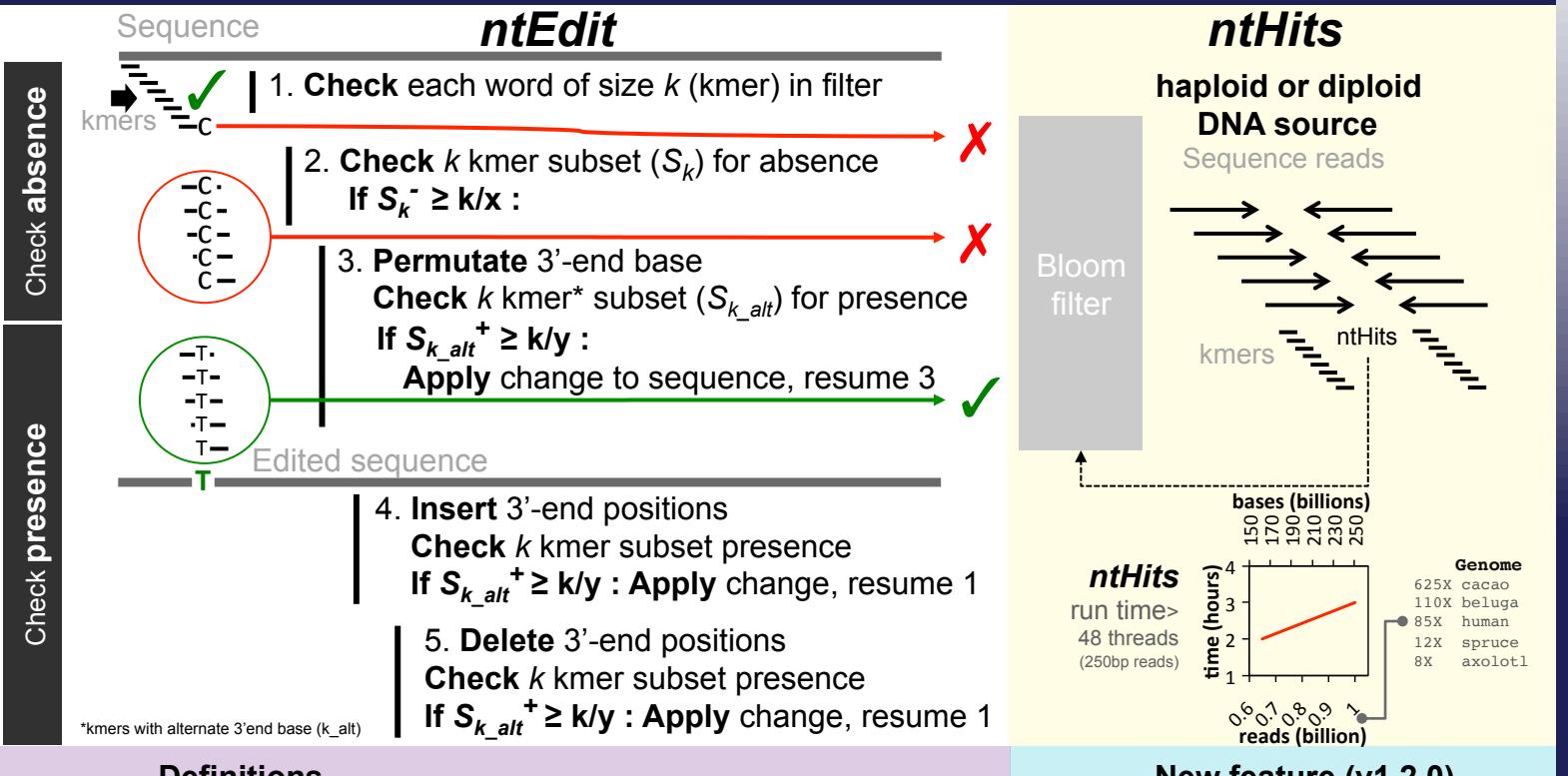


# scalable genome sequence polishing

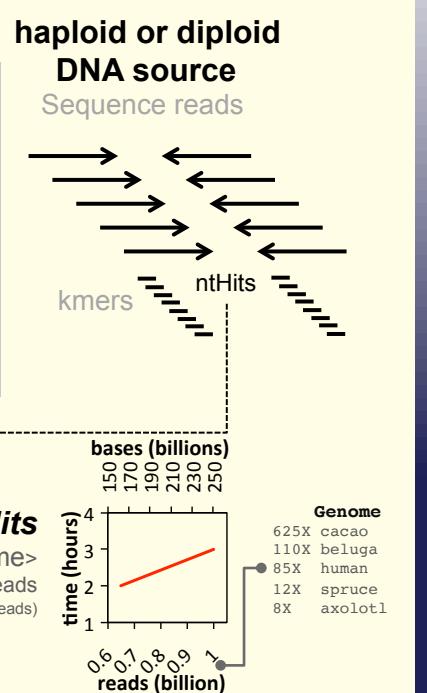


René Warren • Jessica Zhang • Lauren Coombe • Hamid Mohamadi • Inanç Birol

## Method



## ntHits



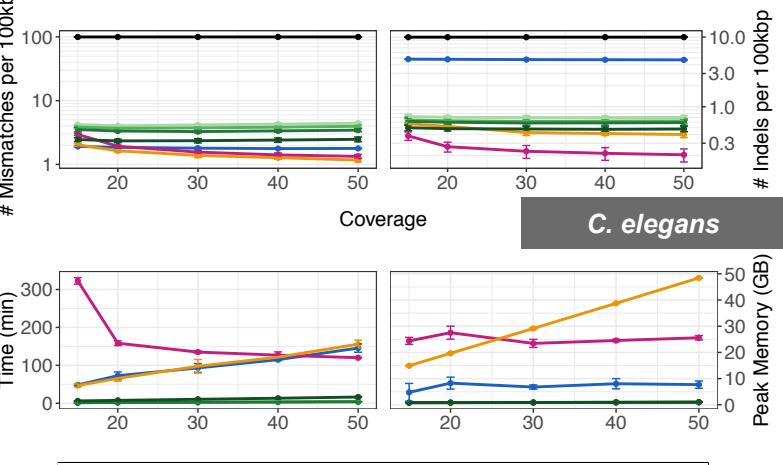
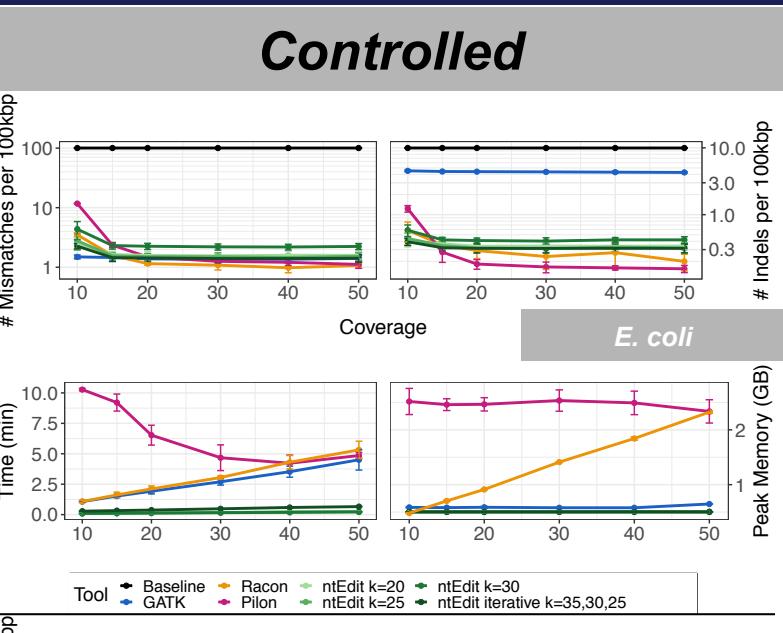
## New feature (v1.2.0)

-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)

## Results

### Controlled



### Experimental

#### I. Polishing

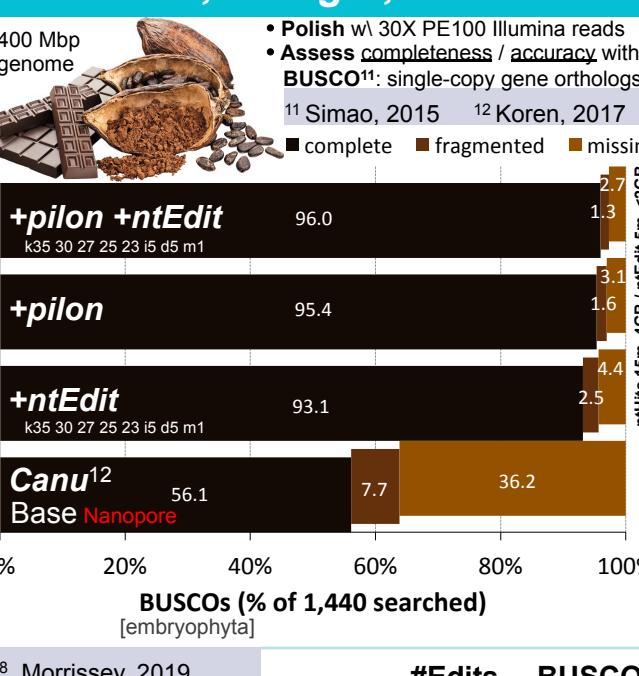
##### Human\*

Polish w/ 54X Illumina	Time**	Edits (M)	BUSCO (%)	Baseline
GATK <sup>4</sup>	41h45	0.97	5,654 (91.3)	7
ntEdit***	2h10	0.95	5,670 (91.6)	23
Racon <sup>5</sup>	45h54	N/A	5,681 (91.7)	34
GATK	42h21	2.66	5,285 (85.4)	0
ntEdit***	2h10	3.63	5,651 (91.3)	366
Racon	40h55	N/A	5,670 (91.6)	385
PacBio <sup>6</sup>			5,647 (91.2)	854

<sup>4</sup>McKenna, 2010 <sup>5</sup>Vaser, 2017 <sup>6</sup>Jain, 2018 <sup>7</sup>Pendleton, 2015

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

##### Cacao<sup>8</sup>, Beluga<sup>9</sup>, Axolotl<sup>10</sup>



<sup>8</sup> Morrissey, 2019 <sup>9</sup> Jones, 2017 <sup>10</sup> Nowoshilow, 2018

#Edits (M) BUSCO (%)



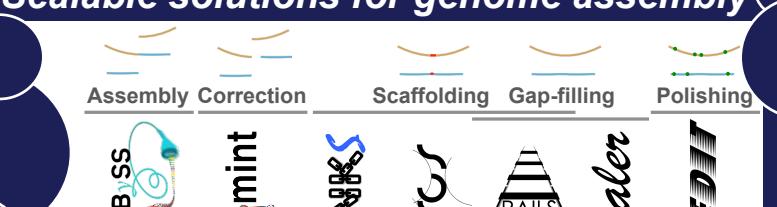
<sup>11</sup> Simao, 2015 <sup>12</sup> Koren, 2017 <sup>13</sup> Warren, 2015

2.7 <2GB 1.3 3.4 1.6 4.4 2.5

ntHits 15m, 4GB ntEdit 5m, 2GB

## 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://birol-lab.ca>

<http://renewarren.ca>

<https://bcgsc.ca>

<https://canadiangenome.ca>

<https://bcgsc.ca>

<https://bcgsc.ca>

<https://bcgsc.ca>

<https://bcgsc.ca>

<https://bcgsc.ca>

<a href="https://bcg