# Resource Allocation

## How much resources do I need?

ARTICLE

https://doi.org/10.1038/s41467-020-16037-7 OPEN

Effect of sequence depth and length in long-read assembly of the maize inbred NC358

Shujun Ou[1], Jianing Liu[2], Kapeel M. Chougule[3], Arkarachai Fungtammasan[4], Arun S. Seetharam[1,5], Joshua C. Stein[3], Victor Llaca[6], Nancy Manchanda[1], Amanda M. Gilbert[7], Sharon Wei[3], Chen-Shan Chin[4], David E. Hufnagel[1], Sarah Pedersen[1], Samantha J. Snodgrass[1], Kevin Fengler[6], Margaret Woodhouse[8], Brian P. Walenz[9], Sergey Koren[9], Adam M. Phillippy[9], Brett T. Hannigan[4], R. Kelly Dawe[2✉], Candice N. Hirsch[7✉], Matthew B. Hufford[1✉] & Doreen Ware[3,10✉]

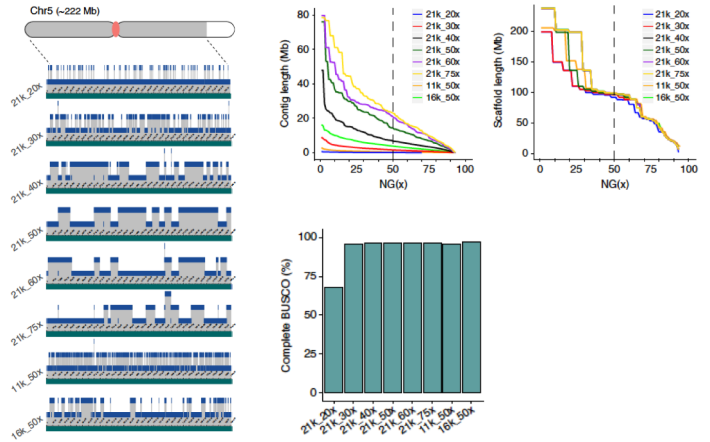
## What do I get with different resources?

**Table 1 Summary statistics for NC358 assemblies.**

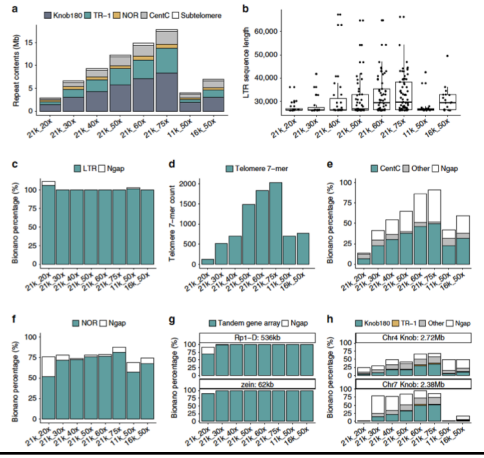
| Experiment[a] | 21k_20× | 21k_30× | 21k_40× | 21k_50× | 21k_60× | 21k_75× | 11k_50× | 16k_50× |
|---|---|---|---|---|---|---|---|---|
| Subreads size (Gb) | 45.62 | 68.16 | 91.01 | 113.89 | 136.80 | 171.08 | 113.63 | 113.60 |
| Subread coverage | 20× | 30× | 40× | 50× | 60× | 75× | 50× | 50× |
| Max read length (kb) | 89.6 | 103.3 | 103.3 | 103.3 | 103.3 | 103.3 | 88.3 | 69.8 |
| Subread N25 (kb) | 30.1 | 30.1 | 30.1 | 30.1 | 30.1 | 30.1 | 14.5 | 21.6 |
| Subread N50 (kb) | 21.2 | 21.2 | 21.2 | 21.2 | 21.2 | 21.2 | 11.1 | 16.8 |
| Corrected reads (Gb) | 25.11 | 48.13 | 66.05 | 82.96 | 88.93 | 100.90 | 79.26 | 80.22 |
| Corrected coverage | 11× | 21× | 29× | 37× | 39× | 44× | 35× | 35× |
| Corrected read N50 (kb) | 18.42 | 17.13 | 17.10 | 17.25 | 18.80 | 20.05 | 10.37 | 14.48 |
| Contig number | 10,563 | 2015 | 641 | 407 | 360 | 327 | 5683 | 1036 |
| Contig total (Gb) | 1.60 | 2.11 | 2.12 | 2.12 | 2.13 | 2.13 | 2.10 | 2.12 |
| Longest contig (Mb) | 1.06 | 11.50 | 47.89 | 76.00 | 79.68 | 78.40 | 4.37 | 21.45 |
| Contig N50 (Mb) | 0.18 | 1.82 | 7.48 | 16.27 | 22.12 | 24.54 | 0.56 | 4.24 |
| Longest scaffold (Mb) | 198.5 | 198.7 | 237.1 | 237.2 | 237.1 | 237.3 | 205.4 | 237.6 |
| Superscaffold N50 (Mb) | 95.3 | 96.9 | 99.2 | 98.5 | 99.4 | 99.2 | 98.5 | 99.4 |
| Assembled (%)[b] | 70.4% | 92.8% | 93.3% | 93.3% | 93.7% | 93.7% | 92.4% | 93.2% |
| Assembly gaps (%) | 24.50% | 0.90% | 0.43% | 0.34% | 0.31% | 0.31% | 2.01% | 0.48% |
| Effective assembly size (Gb)[c] | 1.33 | 1.67 | 1.70 | 1.72 | 1.74 | 1.75 | 1.68 | 1.70 |
| Optical map conflict[d] | 594 | 125 | 56 | 31 | 22 | 21 | 386 | 107 |
| Complete BUSCOs[e] | 68.0% | 95.5% | 96.5% | 96.4% | 96.2% | 96.3% | 95.7% | 96.7% |
| LTR Assembly Index (LAI) | 12.2 | 19.8 | 20.4 | 20.2 | 20.4 | 20.6 | 19.1 | 21.0 |
| Falcon CPU hour | 1563 | 4162 | 6363 | 10,693 | 12,386 | 32,950 | 9721 | 9224 |
| Falcon RAM (Gb) | 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 |
| Canu CPU hour | 1860 | 4036 | 5959 | 7914 | 8849 | 11,520 | 6400 | 7174 |
| Canu RAM (Gb) | 61 | 112 | 149 | 177 | 201 | 120 | 183 | 174 |

[a]Each dataset was assembled only once with the Falcon-Canu hybrid approach (see Methods).
[b]Calculated based on total contig size and the estimated genome size of 2.2724 Gb.
[c]Sum of unique 150-mers.
[d]The optical map was generated using the Direct Label and Stain (DLS) approach with enzyme DLE-1.
[e]Pilon-polished assemblies were used to calculate Benchmarking Universal Single-Copy Orthologs (BUSCO) scores.
CPU central processing.

## What do I get with different resources?



## What do I get with different resources?




## How much would it cost for sequencing for a short read assembly in maize?

- Assume you need ~250x depth
- Genome size is ~2.3Gb = 2,300,000,000bp
- Pricing at https://genomics.umn.edu
- What machine/chemistry will you use?
- How much will this cost in the least expensive scenario?

## How much would it cost for sequencing for a PacBio assembly in maize?

- Assume you need ~20x depth with HiFi reads
- Genome size is ~2.3Gb = 2,300,000,000bp
- Pricing at https://genomics.umn.edu