Hi Josh *et. al*.,

Just want to give you all an brief update on the *G kirkii* assembly.

Here is the basic assembly metrics,

|  |  |
| --- | --- |
| #TITLE: Basic Assembly Stats | |
| #Name | G kirkii |
| Assembler | Hybrid |
| Contigs | 2,361 |
| Max Contig | 9,372,454 |
| Mean Contig | 224,262 |
| Contig N50 | 616,392 |
| Contig N90 | 132,372 |
| Total Contig Length | 529,482,079 |
| Assembly GC | 33.29 |
| Scaffolds | 793 |
| Max Scaffold | 58,731,782 |
| Mean Scaffold | 667,797 |
| Scaffold N50 | 41,165,770 |
| Scaffold N90 | 37,043,456 |
| Total Scaffold Length | 529,563,066 |
| Captured Gaps | 1,568 |
| Max Gap | 100 |
| Mean Gap | 52 |
| Gap N50 | 100 |
| Total Gap Length | 80,987 |

Out of the 793 scaffolds 12 scaffolds represents 98% (518,784,543 Mbps) of the assembled genome. Please see table below,

|  |  |
| --- | --- |
| Contig0|quiver | 58,731,882 |
| Contig13|quiver | 52,648,298 |
| Contig18|quiver | 50,316,412 |
| Contig10|quiver | 44,751,930 |
| Contig19|quiver | 41,945,566 |
| Contig2|quiver | 41,165,870 |
| Contig5|quiver | 40,234,764 |
| Contig9|quiver | 39,207,618 |
| Contig14|quiver | 38,671,273 |
| Contig16|quiver | 38,027,831 |
| Contig48|quiver | 37,043,556 |
| Contig3|quiver | 36,039,543 |
|  | **518,784,543** |

The rest of the genome is in small chunks (781) representing 2% (10,779,723 Mbps) of the genome. Please see table below,

|  |  |
| --- | --- |
| **Unplaced Scaffolds** | |
| Number of Sequences | 781 |
| Max Legnth | 118,062 |
| Min Length | 236 |
| Mean Length | 13,802 |

Annotating the genome using MAKER-P predicted 21,258 gene models.

Also I ran **BUSCO** (***B***enchmarking ***U***niversal ***S***ingle-***C***opy ***O***rtholog; <http://busco.ezlab.org/>) to assess the gene completeness of the genome. Using a plant-specific database of 1,440 genes, BUSCO analysis determined a completeness of 94 % of plant genes. Please table below a short summary of BUSCO results,

|  |
| --- |
| # BUSCO version is: 2.0 beta 4 |
| # The lineage dataset is: embryophyta\_odb9 (Creation date: 2016-11-01, number of species: 30, number of BUSCOs: 1,440) |
| # To reproduce this run: python BUSCO.py -i consensus.fasta -o G.kirkii.v1\_BUSCO\_embryophyta\_odb9 -l /home/projects/moonstone/kirkii\_12-scaffolds-PGA-data-Nov152016/12-scaffolds-Final-20161130T183251Z/Final/PBJELLY2/PBJelly\_Out/QUIVER\_Polishing/017670/BUSCO/embryophyta\_odb9/ -m genome -c 28 -sp arabidopsis |
| # |
| # Summarized benchmarking in BUSCO notation for file consensus.fasta |
| # BUSCO was run in mode: genome |
|  |
| C:93.6%[S:84.2%,D:9.4%],F:1.5%,M:4.9%,n:1440 |
|  |
| 1,349 Complete BUSCOs (C) |
| 1,213 Complete and single-copy BUSCOs (S) |
| 136 Complete and duplicated BUSCOs (D) |
| 21 Fragmented BUSCOs (F) |
| 70 Missing BUSCOs (M) |
| 1,440 Total BUSCO groups searched |

If we can have BioNano whole genome maps to to do final round of validation on this assembly that would be great!

Please let me know if you have further questions.

Thanks,  
Thiru.