PepsiCo releases version 2 of OT3098 reference genome
By: Dr. Mandy Waters, PepsiCo

Method comparison between OT3098 v1 and v2:
Written by Dr. Kevin Fengler, Corteva Agrisciences

The v1 contig assembly was built from 35x PacBio HiFi (read N50 = 17.3 kb, min predicted accuracy > .992) using the HiCanu assembler which resulted in 1,890 contigs with a contig N50= 30.3 Mb and a total assembly size of 10.97 Gb. The contigs were then scaffolded with BioNano genome maps (450 maps, map N50= 98.5, total length = 11.2 Gb) on the BioNano Access platform to create hybrid scaffolds (n= 100, scaffold N50 = 293.9 Mb, total length = 10.84 Gb). Hybrid scaffolds were tentatively placed into pseudomolecules using oat consensus markers. The final ordering of scaffolds was refined using HiC data and contact map visualization using JuiceBox and the Juicer tools.

The v2 release is based on an improved contig assembly using the same PacBio HiFi dataset, but with the hifiasm assembler. The v2 assembly exhibited a significant improvement in contiguity (1,343 contigs, contig N50 = 71 Mb, total size= 10.75 Gb). This assembly was similarly scaffolded with the same BioNano maps as v1, however, the improved contigs significantly improved the contiguity of the scaffolds as well (84 scaffolds, scaffold N50= 374 kb, total size = 10.83 Gb). The v2 assembly was manually curated to resolve overlapping contigs that were not addressed during BioNano hybrid scaffolding (this was not done for v1). Scaffolds were placed into pseudomolecules using the v1 chromosomes as a guide and were validated by HiC contact map visualization as before.

Nomenclature, orientation, and stats:
Chromosomes in v1 were oriented such that the short arm was on top. The OT3098 v2 pseudomolecules use chromosome names and orientations adopted by PanOat in an effort to maintain consistency with future assemblies. The PanOat consortium is currently constructing chromosome-scale reference for multiple oat (*Avena sativa*) genotypes and has therefore adopted a chromosome nomenclature (naming and orientation). Chromosome names specify homologous group (1 to 7) and sub genome (A/C/D). Chromosomes were oriented in such a way that proximal regions of homoeologous chromosomes are collinear. PepsiCo would like to thank Dr. Martin Mascher for updating the v2 pseudomolecule orientation and naming in accordance with PanOat.

Below are the names and lengths of chromosomes in v1 and v2. Chromosomes highlighted in bold were swapped between v1 and v2.

v1_chr_name	v1_chr_length	v2_chr_name	v2_chr_length
1A	542795238	chr1A	540897063
2A	454026946	chr2A	449127287
3A	426317889	chr3A	425675180
4A	462057589	chr4A	463192880
5A	485535456	chr5A	485323027
6A	431567647	chr6A	448461343
7A	493489733	chr7A	493511962
1C	463431985	chr1C	462796039

2C	585391692	chr2C	589118817
3C	636099650	chr3C	638425132
7C	731989224	**chr4C**	716105986
5C	612252875	chr5C	613160974
6C	624915216	chr6C	626220839
4C	552251759	**chr7C**	551718542
1D	485732902	chr1D	484215583
2D	532459853	chr2D	532103454
3D	467934025	chr3D	480949782
4D	424978419	chr4D	455353809
5D	502323219	chr5D	499214392
6D	301592285	chr6D	298028472
7D	529301501	chr7D	528225653
Chr00	235316112	ChrUn	58832055

Future release of data:
PepsiCo is currently working to annotate a set of PacBio transcripts. After annotation is complete researchers will be able to access these data as a track on the genome browser. Additional files including a GFF, ORF nucleotide fasta, and ORF translated peptide fasta will also be available via the download site on GrainGenes. These data are anticipated to be available mid-May 2021. Additionally, Dr. Nick Tinker has mapped relevant SNP markers to v2, which will be available as a track on the genome browser, GFF for download, and visualized as CMaps physical maps.

Use:
Researchers are free to use and publish with all OT3098 genomic resources shared on GrainGenes. Given that no direct publication will be submitted for this individual genome assembly we choose to opt out of the Toronto Agreement, so researchers can freely use these resources as they become available:
- Genome Browser: https://wheat.pw.usda.gov/jb?data=/ggds/oat-ot3098v2-pepsico
- BLAST: https://wheat.pw.usda.gov/blast/ (select "PepsiCo OT3098 Hexaploid Oat v2 pseudomolecules (2021)" under the "Oat Selections")
- Data Download: https://wheat.pw.usda.gov/GG3/content/pepsico-oat-oc3098-v2-files-2021

Citation:
If you use these resources, please cite:
"*Avena sativa* – OT3098 v2, PepsiCo, https://wheat.pw.usda.gov/jb?data=/ggds/oat-ot3098v2-pepsico"