

15 Dec 2017

Description of Medicago R108 (HM340) v1.0 genome data

Medicago HapMap project

NCGR, Montana State University, University of Minnesota and JCVI

www.medicagohapmap.org

Moll KM, Zhou P, Ramaraj T, Fajardo D, Devitt NP, Sadowsky MJ, Stupar RM, Tiffin P, Miller JR, Young ND, Silverstein KAT, Mudge J

LIST OF FILES

Medicago_R108_HM340_v1.0_assembly.fasta.gz

Medicago_R108_HM340_v1.0_annotation.gff3.gz

RESTRICTIONS ON USE

There are no restrictions on the use of this data.

R108 SOURCE MATERIAL

(Renamed HM340 as part of the Hapmap project)

Seeds were obtained from Pascal Ratet derived from *M. truncatula* R108-1 C3 in July 2012. This is the stock propagated in Gif from the original regenerable line and used by the community for transformation (i.e., to produce also the Tnt1 line originally). The same seeds were distributed to all the labs who wanted to have R108 plants or produce transgenics.

R108 ASSEMBLY VERSION 1.0

Input data and assembly strategy:

Approximately 107X coverage of PacBio RSII sequence coverage was polished using Quiver then assembled using FALCON. The PacBio contigs were scaffolded using ~588X physical coverage of Dovetail Chicago libraries and HiRise software. Another round of scaffolding was done using BioNano Irys data. Finally, PBJelly was used for gap filling. More detailed information is available in Moll et al., 2017.

Assembly Statistics:

Contigs: 1,016

Contig N50: 5,925,378

Total Contig Length: 399,348,944

Scaffolds: 909

Scaffold N50: 12, 848,239

Total Scaffold Length: 402,065,285

ANNOTATION

The MAKER-P genome annotation pipeline was used for annotation. All available *M. truncatula* R108 transcripts were assembled using the Trinity Assembler. All transcripts were from a single tissue, root, which is not ideal. Nevertheless, GMAP alignments to A17 indicate that the transcript assembly contains the majority of

genes. The resulting assembly was used as input for expressed sequence tag (EST) evidence for MAKER-P annotations. The MAKER-P pipeline aligns the provided ESTs to the genome and creates ab initio gene predictions with SNAP and Augustus using evidence-based quality values. Each assembly was divided into ten chunks and processed through MAKER-P individually. Following completion of MAKER-P runs for each of the ten chunks, fasta and gff files were combined using fasta_merge and gff3_merge, respectively, included as part of the MAKER-P package.

REFERENCES

- Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011;12:491.
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008;18(1):188–96.
- Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and Curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics*. 2014;48:4 11 11–14 11 39.
- Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004;5(1):59.
- Moll, KM, Zhou P, Ramaraj T, Fajardo D, Devitt NP, Sadowsky MJ, Stupar RM, Tiffin P, Miller JR, Young NC, Silverstein KAT, Mudge J Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago truncatula*. *BMC Genomics*, 2017; 18:578.
- Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;19(Suppl 2):ii215–25.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 2006;34(Web Server issue):W435–9.