

6 June 2014

Description of Medicago R108 genome assembly data

Medicago HapMap project

University of Minnesota, NCGR and JCVI

www.medicagohapmap.org

Joann Mudge, Jason Miller, Thiru Ramaraj, Peng Zhou, Joseph Guhlin, Kevin Silverstein and Nevin Young

R108 SOURCE MATERIAL

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 0.9

An initial draft assembly was created using ALLPATHS (Gnerre et al., 2011) in February 2013 using default parameters. The following are the assembly statistics obtained:

Assembly Statistics:

num_scaffolds = 6,118

num_contigs = 32,949

total genome length incl. gaps = 369,383,725

total genome length w/o gaps = 334,415,625

Contig N50 = 22,849

Scaffold N50 = 351,578

Max scaffold size = 2,244,698

Min scaffold size = 920

INPUT DATA

- 166,118,124 (~60 X) SIPE reads, 2 x 100 bp from the HiSeq 2000, insert size 142 bp
- 82,323,745 (~30 X) LIPE (i.e. mate pair) reads, 2 x 100 bp from the HiSeq 2000 with insert size 6500 bp

LIST OF FILES

Medicago_R108_HM340_v0.9_assembly.fasta.gz (102 MB)

REFERENCES

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011 Jan 25;108(4):1513-8. doi: 10.1073/pnas.1017351108. Epub 2010 Dec 27. <http://www.ncbi.nlm.nih.gov/pubmed/21187386>