

24 June 2016

Description of *Medicago truncatula de novo* assemblies

Medicago HapMap project

University of Minnesota, National Center for Genome Resources (NCGR) and J. Craig Venter Institute (JCVI)

www.medicagohapmap.org

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

Genome Assemblies May 2015 Release

Twenty *Medicago truncatula* accessions were sequenced and assembled *de novo* to facilitate genome-wide comparisons of structural variation, copy-number variation, gene family architecture, and validation of called SNP variation.

June 24, 2016 Update

HM018a (A20) sequence has been added for download.

May 15, 2015 Update

Three additional accessions were added for download: HM005, HM010, and HM022.

Please Note: The Hapmap project intended to sequence HM018, but downstream SNP analysis revealed an exact match (within sequencing error) to HM023. Presumably there was a mix-up during germplasm or DNA processing. Hence, the reported sequence assembly has been re-named "HM023a".

Restrictions on Use

These resources are made available to the research community by the *Medicago* HapMap consortium under the *Toronto Agreement* [<http://www.nature.com/nature/journal/v461/n7261/full/461168a.html>]. As producers of these data, we reserve the right to be the first to publish a genome-wide, multiple accession comparative analysis of the data.

The pre-publication data released here is embargoed for publication except for analyses of single gene loci or small (< 10 kb) genome regions, which may involve multiple *Medicago* accessions. Researchers are encouraged to contact us if there are queries about referencing or publishing analyses based on pre-publication data obtained via this website. Researchers are also invited to consider collaborations with the *Medicago* Hapmap consortium for larger studies or if the limitations here restrict further work.

Please note that no gene annotation is provided for these assemblies.

Assemblies were generated from Illumina Sequencing and ALLPATHS-NG. The following table lists assembly statistics for each accession.

Accession	Scaffold N50 (in kb)	Contig N50 (in kb)	Number of Scaffolds	Total Contig Length	Total Scaffold Length
HM002	1,265	34.4	3,836	354,262,477	412,786,159
HM004	1,057	39.9	3,494	360,726,300	398,065,827
HM005	493	43.3	3,979	355,212,816	382,558,815
HM006	464	39.1	4,583	351,903,053	383,166,944

ASSEMBLY STATISTICS

HM010	698	39.0	3,159	365,654,634	402,576,858
HM017	706	39.8	3,489	338,805,835	369,879,279
HM018a	669	23.3	4,995	356,828,485	402,812,623
HM020	591	36.4	3,767	359,431,501	405,236,562
HM022	1,153	41.7	2,565	343,345,336	373,691,354
HM023a	422	24.4	3,728	361,243,464	403,142,311
HM026	542	36.3	3,999	357,046,471	402,599,446
HM034	471	26.7	3,267	362,786,782	396,665,787
HM050	1,375	42.6	3,234	366,110,832	404,587,980
HM056	511	19.2	3,486	362,971,816	406,705,336
HM058	374	18.5	4,349	355,004,629	409,729,257
HM060	452	22.5	3,634	363,308,695	403,209,823
HM095	526	24.7	3,711	367,894,112	410,354,770
HM125	517	28.1	3,666	371,005,289	406,998,713
HM129	523	21.0	3,213	367,625,895	398,468,296
HM185	1,653	25.1	3,335	367,670,036	428,228,061
HM324	267	41.5	7,057	343,081,677	421,810,942

HM340 / R108

The most recent release of R108 / HM340 is available separately at:

<http://www.medicagohapmap.org/downloads/r108>

LIST OF FILES

all-assemblies-24June2016-release.tgz
 HM002-29Apr2015-release.fasta.gz
 HM004-29Apr2015-release.fasta.gz
 HM005-15May2015-release.fasta.gz
 HM006-29Apr2015-release.fasta.gz
 HM010-15May2015-release.fasta.gz
 HM017-29Apr2015-release.fasta.gz

HM018a-3May2016-release.fasta.gz
HM020-29Apr2015-release.fasta.gz
HM022-15May2015-release.fasta.gz
HM023-29Apr2015-release.fasta.gz
HM026-29Apr2015-release.fasta.gz
HM034-29Apr2015-release.fasta.gz
HM050-29Apr2015-release.fasta.gz
HM056-29Apr2015-release.fasta.gz
HM058-29Apr2015-release.fasta.gz
HM060-29Apr2015-release.fasta.gz
HM095-29Apr2015-release.fasta.gz
HM125-29Apr2015-release.fasta.gz
HM129-29Apr2015-release.fasta.gz
HM185-29Apr2015-release.fasta.gz
HM324-29Apr2015-release.fasta.gz

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>