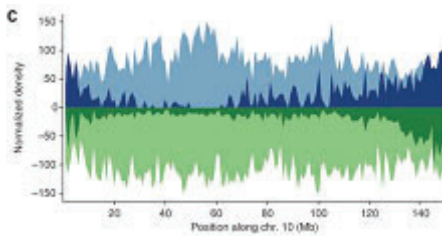


Genotypes

From this page you can obtain the latest public "flat file" versions of prepackaged, genotypic data sets. All of these can be opened with the TASSEL 5 (GUI or command line). The genotypes were obtained by whole genome sequencing ("HapMap"), GBS, or with MaizeSNP50 chip or traditional SNP assays.

Maize HapMap genotypes



In collaboration with several projects, we have performed whole-genome, Illumina sequencing on hundreds of maize lines, uncovering millions of polymorphisms. These high density genotypes have enabled GWAS in maize. Several maize HapMap versions are available, the most recent of which is maize HapMapV3. The bam files used for construction of HapMapV3 are available through CyVerse Data Store. [Click here for instruction to access the bam files.](#)

Citation: Construction of the third generation Zea mays haplotype map. *Gigascience*. 2017 Dec 30. doi: [10.1093/gigascience/gix134](https://doi.org/10.1093/gigascience/gix134). [\[Epub ahead of print\]](#)

- **Maize 282 association panel genotypes (7x, AGPv3 coordinates) [\[more\]](#)**

Genotypes of 282 association panel based on the whole genome sequencing data. This data set include the same markers as in HapMap V3.2.1, but called from higher coverage sequencing data than the HapMapV3.2.1 release. The VCF files can be downloaded from CyVerse in the directory: /iplant/home/shared/commons_repo/curated/

Qi_Sun_Zea_mays_haplotype_map_2018/282_onHmp321. The file names are

c*_282_corrected_onHmp321.vcf.gz. You can use this icommand to download all files: `iget -K -r /iplant/home/shared/commons_repo/curated/`

Qi_Sun_Zea_mays_haplotype_map_2018/282_onHmp321. Users from China can use this link to download files: <http://pan.baidu.com/s/1eRAyACM>. All raw data has been deposited to NCBI SRA [PRJNA389800](https://www.ncbi.nlm.nih.gov/sra/PRJNA389800).

- **Maize 282 association panel genotypes (7x, AGPv4 coordinates) [\[more\]](#)**

Genotypes of 282 association panel based on the whole genome sequencing data. This data set include the same markers as in HapMap V3.2.1, but called from higher coverage sequencing data than the HapMapV3.2.1 release. These files have their coordinates uplifted to agpv4 The VCF files can be downloaded from CyVerse in the directory: /iplant/home/shared/panzea/hapmap3/hmp321/unimputed/282_libs_2015/uplifted_APgv4. The file names are c*_282_corrected_onHmp321.vcf.gz. You can use this icommand to download all files: `iget -K -r /iplant/home/shared/panzea/hapmap3/hmp321/unimputed/282_libs_2015/uplifted_APgv4` Users from China can download files from this link: <http://pan.baidu.com/s/1kVfwCrt>

- **Maize HapMapV3.2.1 genotypes (with imputation by LinkImpute, AGPv3) [\[more\]](#)**

- **2016.10.5** Read depths are added to the imputed VCF files. There are minor changes in the file format to be compliant with VCF 4.1 specification. The new files with read depth and correct vcf tags are named c*_hmp321_withDPGL_imputed.vcf.gz.

The HapMapV3 study used whole genome sequence data from 1210 maize lines. From these data, genotypes were called for 83 million SNPs (30 million SNPs marked with "LLD" are high confidence markers). The imputation was performed with [LinkImpute](#). The merged_flt_c*.vcf.gz files are smaller, but have no DP (read depth) and GL (genotype likelihood) for each genotype. The c*_hmp321_withDPGL_imputed.vcf.gz files have the DP and GL tag. The bam files used for construction of HapMapV3 are available through CyVerse Data Store. [Click here for instruction to access the bam files.](#) Users from China please download from this link: <http://pan.baidu.com/s/>

[1eRNGtxw.](#)

- **Maize HapMapV3.2.1 genotypes (with imputation by LinkImpute, AGPv4) [\[more\]](#)**

Uplifted from imputed HapMap3.2.1 agpv3.

- **2016.10.5** Read depths are added to the imputed VCF files. There are minor changes in the file format to be compliant with VCF 4.1 specification. The new files with read depth and correct vcf tags are located in /iplant/home/shared/panzea/hapmap3/hmp321/imputed/uplifted_APGv4. The files are named hmp321_agpv4_chr*.vcf.gz

- **Maize HapMapV3.2.1 genotypes (without imputation, AGPv3) [\[more\]](#)**

- **2016.10.5** There are minor changes in the file format to be compliant with VCF 4.1 specification. The new files are named c*_hmp321_withGL.vcf.gz.

The HapMapV3 study used whole genome sequence data from 1210 maize lines. From these data, genotypes were called for 83 million SNPs (30 million SNPs marked with "LLD" are high confidence markers). These are files without imputation. The merged_flt_c*.vcf.gz files are smaller, they have DP (read depth) but have no GL (genotype likelihood) for each genotype. The c*_hmp321_withGL.vcf.gz files have the GL tag. Users from China please download from this link: <http://pan.baidu.com/s/1eRNGtxw>

- **Maize HapMapV3.2.1 genotypes (without imputation, AGPv4) [\[more\]](#)**

Uplifted from agpv3 genotypes.

- **2016.10.5** There are minor changes in the file format to be compliant with VCF 4.1 specification. The files are located in /iplant/home/shared/panzea/hapmap3/hmp321/unimputed/uplifted_APGv4, and named hmp321_agpv4_chr*.vcf.gz.

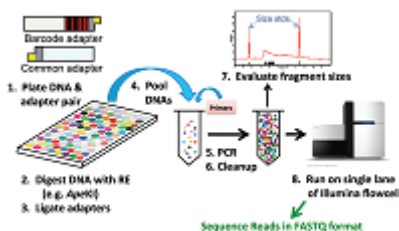
- **Maize HapMapV2 genotypes [\[more\]](#)**

The maize HapMapV2 study was a collaborative effort involving Panzea, CSHL, BGI, CAU, UC Davis, USDOE and several others. This study generated random-sheared, paired-end Illumina GAII reads from 103 maize, teosinte and maize landrace inbred lines at a depth ranging from 4-30X. High quality genotypes were obtained for 55 million SNPs and small indels.

- **Maize HapMapV1 genotypes [\[more\]](#)**

Genotypes from our original, first-generation maize HapMap study.

Maize GBS genotypes



We have performed Genotyping By Sequencing (GBS) with the enzyme *ApeKI* and the [Elshire et al. 2011 protocol](#) on more than 60K maize samples to date. Many of these genotypes are publically available.

- **ZeaGBSv2.7 [\[more\]](#)**

This is the current ZeaGBS build, containing genotypes for 955,690 SNPs at more than 60K taxa (samples), 17,280 of which are public. For these public samples, both raw and partially imputed genotypes are available as HDF5 files with AGPv2 coordinates which can be opened with TASSEL 5. In addition, the variants and genotypes were uplifted to AGPv3 and AGPv4 coordinates in hapmap and Tassel HDF5 formats.

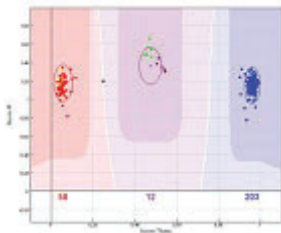
- **ZeaGBSv2.3 [\[more\]](#)**

This somewhat older GBS build consists of 2,219,843 SNP genotypes for 32,736 public and private taxa. SNP filtering was less stringent compared to v2.7, so the additional SNPs are more "noisy". However, the public data are more highly curated, as they are split into germplasm sets of interest, including NAM, IBM, the Maize282 association panel, and a teosinte/W22 backcross family. Furthermore, cleaned up, phased, and fully imputed genotypes are available for NAM and the teosinte/W22 backcross family. The genome coordinates are AGPv2.

- **ZeaGBSv1.0** [\[more\]](#)

This is the exact genotypic data set published in Romay et al. (2013) Genome Biology 14:R55. It consists of 681,257 SNPs from ZeaGBSv1.0 scored on 4,476 samples.

Legacy SNPs



We also have SNP genotypic data from more traditional SNP assays, including the [Illumina MaizeSNP50 BeadChip](#), the Illumina Golden Gate SNP assay, and the Sequenom single-base extension assay.

- **50K SNP chip genotypes** [\[more\]](#)

Illumina MaizeSNP50 BeadChip genotypes.

- **Legacy SNPs** [\[more\]](#)

Legacy SNPs designed from Sanger sequencing data (context sequences provided); Genotyping results from traditional SNP assays such as the Illumina Golden Gate SNP assay or the Sequenom single-base extension assay.
