

Construct BSgenome Object

exomePeak2 requires BSgenome object of matched species. However, when there is no available BSgenome object, we have to manually construct one. Therefore, this page provides instructions for constructing BSgenome object from scratch by using *Zea mays* (Corn) genome as an example.

1. Download Genome FASTA

Download *Zea mays* (Corn) reference genome "AGPv4" from [Illumina iGenome](https://www.illumina.com/genomes/igenomes/illumina_iGenome.html), and find genome fasta file.

```
$ wget http://igenomes.illumina.com/s3-website-us-east-1.amazonaws.com/Zea_mays/Ensembl/AGPv4/Zea_mays_Ensembl_AGPv4.tar.gz
$ tar -xvzf Zea_mays_Ensembl_AGPv4.tar.gz
```

2. Convert fasta to 2bit file format

Download "[faToTwoBit](#)" tool by

```
$ rsync -aP \
  rsync://hgdownload.soe.ucsc.edu/genome/admin/exe/linux.x86_64/faToTwoBit
./
```

Convert file format

```
$ mkdir
/home/zhen.di/corn/Zea_mays/Ensembl/AGPv4/Sequence/wholeGenomeFasta/preparation/
$ cd
/home/zhen.di/corn/Zea_mays/Ensembl/AGPv4/Sequence/wholeGenomeFasta/preparation/
$ cp ../genome.fa zmAGPv4.fa
$ faToTwoBit zmAGPv4.fa zmAGPv4.2bit
```

3. Extract chromosome names

```
$ less -S zmAGPv4.fa | grep ">" | awk '{print $1}' | sed 's/^>//g' >
zmAGPv4.chromName.txt
```

Find chromosome name by

```
$ cat zmAGPv4.chromName.txt
```

4. Prepare seed file "BSgenome.Zmays.Ensemble.zmAGPv4-seed" in the same directory

```
Package: BSgenome.Zmays.Ensemble.zmAGPv4
Title: Genome sequences for Zea mays (Ensemble AGPv4)
Description: A BSgenome package containing the full genome sequences for Zea mays (Maize) as provided by Ensemble (B73 AGPv4, Sept. 2020) and stored in Biostrings objects.
Version: 1.0
organism: Zea mays
```

```

common_name: Maize
provider: Ensemble
provider_version: zmAGPv4
release_date: Sept. 2020
release_name: Maize Genome Sequencing B73 RefGen_v4.0
#source_url: ftp://ftp.ensemblgenomes.org/pub/plants/release-
49/fasta/zea_mays/dna/
circ_seqs: c("Mt","Pt")
organism_biocview: Zea_mays
BSgenomeObjname: Zmays
seqs_srcdir:
/home/zhen.di/corn/Zea_mays/Ensembl/AGPv4/Sequence/wholeGenomeFasta/preparat
ion/
seqfiles_suffix: .fa
seqnames: c("10", "1", "2", "3", "4", "5", "6", "7", "8", "9", "MT", "Pt")
seqfile_name: zmAGPv4.2bit

```

Please make sure the information included in the seed file is correct by referring to [BSgenomeForge doc](#) and [description](#).

5. Construct BSgenome data package

```

# Construct BSgenome data package in R
library(rtracklayer)
library(BSgenome)
setwd("/home/zhen.di/corn/Zea_mays/Ensembl/AGPv4/Sequence/wholeGenomeFasta/p
reparation")
forgeBSgenomeDataPkg("BSgenome.Zmays.Ensemble.zmAGPv4-seed")

## The construction completes successfully with messages:

# Creating package in ./BSgenome.Zmays.Ensemble.zmAGPv4
# Loading
'/home/zhen.di/corn/Zea_mays/Ensembl/AGPv4/Sequence/wholeGenomeFasta/prepara
tion//zmAGPv4.2bit' ... DONE
# Writing sequences to
'./BSgenome.Zmays.Ensemble.zmAGPv4/inst/extdata/single_sequences.2bit' ...
DONE

```

6. Construct R package

```
$ R CMD build BSgenome.Zmays.Ensemble.zmAGPv4
```

7. Check whether the tar.gz file has been successfully created by

```
$ R CMD check BSgenome.Zmays.Ensemble.zmAGPv4_1.0.tar.gz
```

You should receive messages showing everything is OK.

8. Install package into R

```
$ R CMD INSTALL BSgenome.Zmays.Ensemble.zmAGPv4_1.0.tar.gz
```

9. Construct TxDb object

```
library(AnnotationHub) # help to build the annotation object
library(biomaRt)
library(GenomicFeatures)
# library(exomePeak2)
library(rtracklayer)
library(BSgenome)
library(BSgenome.Zmays.Ensemble.zmAGPv4)

maize_txdb<- makeTxDbFromBiomart(biomart = "plants_mart",dataset =
"zmays_eg_gene",host = "http://plants.ensembl.org")
saveDb(maize_txdb, file="maize_v4.sqlite")
maize_txdb <- loadDb("maize_v4.sqlite")
```

10. Use TxDb object in exomePeak2 function

```
exomePeak2(bam_ip = c("IP_1.bam", "IP_2.bam"),
            bam_input = c("Input_1.bam", "Input_2.bam"),
            txdb = maize_txdb,
            bsgenome = BSgenome.Zmays.Ensemble.zmAGPv4,
            paired_end = TRUE)
```

Reference

1. Blog - by Zhanmin Liang

<https://www.jianshu.com/p/1fa3c7eb850e>

2. How to forge a BSgenome data package

<https://www.bioconductor.org/packages/devel/bioc/vignettes/BSgenome/inst/doc/BSgenomeForge.pdf>

3. Writing R extensions

<https://cran.r-project.org/doc/manuals/r-release/R-exts.html#The-DESCRIPTION-file>