

# Run GENESPACE vignette - grasses

JTLovell

11-Feb 2022

## 1. Global Parameters

```
baseDir <- "/Users/jlovell/Desktop/GENESPACE_data/results/"  
rawAnnotationDir <- "/Users/jlovell/Desktop/GENESPACE_data"  
mcscanDir <- "/Users/jlovell/Documents/comparative_genomics/programs/MCScanX"  
nThreads <- 4  
path2of <- "orthofinder" # since orthofinder is in the path via conda
```

## 2. Set parameters

**NOTE:** Wheat, Maize and Switchgrass WGD occurred after MRCA of all genomes below. Therefore, treat them as polyploid.

Since some ploidy > 1, run orthofinder again in syntenic blocks.

```
grasses <- list(  
  wd = file.path(baseDir, "grasses"),  
  speciesIDs = "grasses",  
  genomes = data.table(do.call("rbind", list(  
    c(genome = "Phallii", version = "Phallii_HAL2_v2.1", ploidy = 1),  
    c("switchgrass", "switchgrass_v5.1", 2),  
    c("Sviridis", "Sviridis_v2.1", 1),  
    c("Sorghum", "Sbicolor_BTx623_v3.1", 1),  
    c("maize", "maize_refgen_v5", 2),  
    c("rice", "rice_kitaake_v3.1", 1),  
    c("brachy", "Bdistachyon_v3.1", 1),  
    c("wheat", "wheat_v4.0", 3))))  
  
grassesParams <- list(  
  orthofinderInBlk = T, # Recommended when any genome has ploidy > 1.  
  pepString = "fa") # specify the file name search string for peptide fastas
```

## 3 Initialize the grasses run with the above specified parameters

```
gparGrasses <- with(grasses, init_genespace(  
  genomeIDs = genomes$genome,  
  versionIDs = genomes$version,  
  ploidy = genomes$ploidy,
```

```

speciesIDs = rep(speciesIDs, length(genomes$genome)),
orthofinderInBlk = grassesParams$orthofinderInBlk,
pepString = grassesParams$pepString,
wd = wd,
nCores = nThreads,
path2orthofinder = path2of,
path2mcscanx = mcscanDir,
rawGenomeDir = rawAnnotationDir)

## set working directory to /Users/jlovell/Desktop/GENESPACE_data/results/grasses
##
## found raw gff files:
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Phallii_HAL2_v2.1/annotation/PhalliiHAL_496_v2.1.gene.gff
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/switchgrass_v5.1/annotation/Pvиргатум_516_v5.1.gene.gff
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Sviridis_v2.1/annotation/Sviridis_500_v2.1.gene.gff3
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Sbicolor_BTx623_v3.1/annotation/Sbicolor_313_v3.1.gene.gff
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/maize_refgen_v5/annotation/GCF_902167145.1_Zm-B73-REF
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/rice_kitaake_v3.1/annotation/OsativaKitaake_499_v3.1.gene.gff
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Bdistachyon_v3.1/annotation/Bdistachyon_314_v3.1.gene.gff
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/wheat_v4.0/annotation/GCA_002220415.3_Triticum_4.0.gene.gff
##
## found raw peptide files:
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Phallii_HAL2_v2.1/annotation/PhalliiHAL_496_v2.1.protein.fasta
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/switchgrass_v5.1/annotation/Pvиргатум_516_v5.1.protein.fasta
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Sviridis_v2.1/annotation/Sviridis_500_v2.1.protein.fasta
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Sbicolor_BTx623_v3.1/annotation/Sbicolor_313_v3.1.protein.fasta
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/maize_refgen_v5/annotation/GCF_902167145.1_Zm-B73-REF
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/rice_kitaake_v3.1/annotation/OsativaKitaake_499_v3.1.protein.fasta
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/Bdistachyon_v3.1/annotation/Bdistachyon_314_v3.1.protein.fasta
##   /Users/jlovell/Desktop/GENESPACE_data/grasses/wheat_v4.0/annotation/GCA_002220415.3_Triticum_4.0.protein.fasta
##
## Can't find all parsed annotation files ... need to run parse_annotations, parse_ncbi or parse_phytozome
##
## GENESPACE run initialized:
##   Initial orthofinder database generation method: default inside R
##   Orthology graph method: inBlock

```

## 4 Parse the raw annotations.

This is a good example of the flexibility of GENESPACE annotation parsing. The maize genome is from NCBI, the wheat from an independent server with different gff specifications, and the rest from phytozome.

```

parse_ncbi(gsParam = gparGrasses, genomeIDs = "maize")

## Parsing annotations: maize
## Reading gff ... found 34295 protein coding genes
## Reading peptide fasta ... found 57578 / 34295 total and unique entries
## Merging fa and gff found 34294 matching entires

parse_phytozome(
  gsParam = gparGrasses,
  genomeIDs = c("Phallii", "switchgrass", "Sviridis", "Sorghum", "rice", "brachy"))

```

```

## Parsing annotations: Phallii
##  Reading gff ... found 33263 protein coding genes
##  Reading peptide fasta ... found 33263 / 33263 total and unique entries
##  Merging fa and gff found 33263 matching entires
## Parsing annotations: switchgrass
##  Reading gff ... found 80278 protein coding genes
##  Reading peptide fasta ... found 80278 / 80278 total and unique entries
##  Merging fa and gff found 80278 matching entires
## Parsing annotations: Sviridis
##  Reading gff ... found 38334 protein coding genes
##  Reading peptide fasta ... found 38334 / 38334 total and unique entries
##  Merging fa and gff found 38334 matching entires
## Parsing annotations: Sorghum
##  Reading gff ... found 34211 protein coding genes
##  Reading peptide fasta ... found 34211 / 34211 total and unique entries
##  Merging fa and gff found 34211 matching entires
## Parsing annotations: rice
##  Reading gff ... found 35594 protein coding genes
##  Reading peptide fasta ... found 35594 / 35594 total and unique entries
##  Merging fa and gff found 35594 matching entires
## Parsing annotations: brachy
##  Reading gff ... found 34310 protein coding genes
##  Reading peptide fasta ... found 34310 / 34310 total and unique entries
##  Merging fa and gff found 34310 matching entires

parse_annotations(
  gsParam = gparGrasses, genomeIDs = "wheat", gffEntryType = "gene",
  gffIdColumn = "ID", headerEntryIndex = 2, headerSep = "locus_tag=",
  headerStripText = "] .*", gffStripText = "gene-")

## Parsing annotation files ...
## wheat ...
##      Importing gff ... found 109335 gff entires, and 109335 gene entries
##      Importing fasta ... found 135934 fasta entires
##
## **FOUND DUPLICATE PEPTIDE FASTA HEADERS. IS THIS JUST A PRIMARY TRANSCRIPT FILE?**
## subsetting to longest model for each gene
##      107351 gff-peptide matches
## Done!

```

## 5 Run OrthoFinder from within R.

Also set the synteny parameters as default.

```

gparGrasses <- set_syntenyParams(gsParam = gparGrasses)
gparGrasses <- run_orthofinder(gsParam = gparGrasses)

## Running 'defualt' genespace orthofinder method
## #####
## Cleaning out orthofinder directory and prepping run
## Calculating blast results and running OrthoFinder
## #####
## #####
## #####
## #####
## 
```

```

## 
## OrthoFinder version 2.5.4 Copyright (C) 2014 David Emms
##
## 2022-02-11 12:35:03 : Starting OrthoFinder 2.5.4
## 4 thread(s) for highly parallel tasks (BLAST searches etc.)
## 1 thread(s) for OrthoFinder algorithm
##
## Checking required programs are installed
## -----
## Test can run "mcl -h" - ok
## Test can run "fastme -i /Users/jlovell/Desktop/GENESPACE_data/results//grasses/orthofinder/Results_1
##
## Dividing up work for BLAST for parallel processing
## -----
## 2022-02-11 12:35:06 : Creating diamond database 1 of 8
## 2022-02-11 12:35:07 : Creating diamond database 2 of 8
## 2022-02-11 12:35:07 : Creating diamond database 3 of 8
## 2022-02-11 12:35:07 : Creating diamond database 4 of 8
## 2022-02-11 12:35:08 : Creating diamond database 5 of 8
## 2022-02-11 12:35:08 : Creating diamond database 6 of 8
## 2022-02-11 12:35:08 : Creating diamond database 7 of 8
## 2022-02-11 12:35:09 : Creating diamond database 8 of 8
##
## Running diamond all-versus-all
## -----
## Using 4 thread(s)
## 2022-02-11 12:35:10 : This may take some time.....
## 2022-02-11 12:35:11 : Done 0 of 64
## 2022-02-11 13:31:24 : Done 10 of 64
## 2022-02-11 13:59:56 : Done 20 of 64
## 2022-02-11 14:17:57 : Done 30 of 64
## 2022-02-11 14:32:36 : Done 40 of 64
## 2022-02-11 14:43:50 : Done 50 of 64
## 2022-02-11 14:54:31 : Done 60 of 64
## 2022-02-11 15:01:44 : Done all-versus-all sequence search
##
## Running OrthoFinder algorithm
## -----
## 2022-02-11 15:01:46 : Initial processing of each species
## 2022-02-11 15:02:26 : Initial processing of species 0 complete
## 2022-02-11 15:03:08 : Initial processing of species 1 complete
## 2022-02-11 15:03:55 : Initial processing of species 2 complete
## 2022-02-11 15:04:36 : Initial processing of species 3 complete
## 2022-02-11 15:05:21 : Initial processing of species 4 complete
## 2022-02-11 15:06:04 : Initial processing of species 5 complete
## 2022-02-11 15:07:34 : Initial processing of species 6 complete
## 2022-02-11 15:10:12 : Initial processing of species 7 complete
## 2022-02-11 15:10:52 : Connected putative homologues
## 2022-02-11 15:10:58 : Written final scores for species 0 to graph file
## 2022-02-11 15:11:04 : Written final scores for species 1 to graph file
## 2022-02-11 15:11:11 : Written final scores for species 2 to graph file
## 2022-02-11 15:11:17 : Written final scores for species 3 to graph file
## 2022-02-11 15:11:23 : Written final scores for species 4 to graph file
## 2022-02-11 15:11:29 : Written final scores for species 5 to graph file

```

```
## 2022-02-11 15:11:42 : Written final scores for species 6 to graph file
## 2022-02-11 15:12:01 : Written final scores for species 7 to graph file
##
## WARNING: program called by OrthoFinder produced output to stderr
##
## Command: mcl /Users/jlovell/Desktop/GENESPACE_data/results//grasses/orthofinder/Results_Feb11/Working
##
## stdout
## -----
## b''
## stderr
## -----
## b'[mcl] cut <2> instances of overlap\n'
## 2022-02-11 15:13:29 : Ran MCL
##
## Writing orthogroups to file
## -----
## OrthoFinder assigned 349308 genes (87.8% of total) to 36635 orthogroups. Fifty percent of all genes
## 2022-02-11 15:13:58 : Done orthogroups
##
## Analysing Orthogroups
## -----
##
## Calculating gene distances
## -----
## 2022-02-11 15:22:27 : Done
## 2022-02-11 15:22:32 : Done 0 of 26523
## 2022-02-11 15:22:41 : Done 1000 of 26523
## 2022-02-11 15:22:48 : Done 2000 of 26523
## 2022-02-11 15:22:53 : Done 3000 of 26523
## 2022-02-11 15:22:57 : Done 4000 of 26523
## 2022-02-11 15:23:02 : Done 5000 of 26523
## 2022-02-11 15:23:07 : Done 6000 of 26523
## 2022-02-11 15:23:11 : Done 7000 of 26523
## 2022-02-11 15:23:16 : Done 8000 of 26523
## 2022-02-11 15:23:21 : Done 9000 of 26523
## 2022-02-11 15:23:25 : Done 10000 of 26523
## 2022-02-11 15:23:30 : Done 11000 of 26523
## 2022-02-11 15:23:35 : Done 12000 of 26523
## 2022-02-11 15:23:39 : Done 13000 of 26523
## 2022-02-11 15:23:44 : Done 14000 of 26523
## 2022-02-11 15:23:49 : Done 15000 of 26523
## 2022-02-11 15:23:53 : Done 16000 of 26523
## 2022-02-11 15:23:58 : Done 17000 of 26523
## 2022-02-11 15:24:03 : Done 18000 of 26523
## 2022-02-11 15:24:08 : Done 19000 of 26523
## 2022-02-11 15:24:12 : Done 20000 of 26523
## 2022-02-11 15:24:17 : Done 21000 of 26523
## 2022-02-11 15:24:22 : Done 22000 of 26523
## 2022-02-11 15:24:26 : Done 23000 of 26523
## 2022-02-11 15:24:31 : Done 24000 of 26523
## 2022-02-11 15:24:36 : Done 25000 of 26523
## 2022-02-11 15:24:40 : Done 26000 of 26523
```

```

## 
## Inferring gene and species trees
## -----
## 
## 15296 trees had all species present and will be used by STAG to infer the species tree
## 
## Best outgroup(s) for species tree
## -----
## 2022-02-11 15:29:39 : Starting STRIDE
## 2022-02-11 15:29:48 : Done STRIDE
## Observed 509 well-supported, non-terminal duplications. 454 support the best root and 55 contradict
## Best outgroup for species tree:
##     brachy, wheat
## 
## Reconciling gene trees and species tree
## -----
## Outgroup: brachy, wheat
## 2022-02-11 15:29:48 : Starting Recon and orthologues
## 2022-02-11 15:29:48 : Starting OF Orthologues
## 2022-02-11 15:29:50 : Done 0 of 26523
## 2022-02-11 15:30:05 : Done 1000 of 26523
## 2022-02-11 15:30:14 : Done 2000 of 26523
## 2022-02-11 15:30:22 : Done 3000 of 26523
## 2022-02-11 15:30:29 : Done 4000 of 26523
## 2022-02-11 15:30:35 : Done 5000 of 26523
## 2022-02-11 15:30:41 : Done 6000 of 26523
## 2022-02-11 15:30:47 : Done 7000 of 26523
## 2022-02-11 15:30:52 : Done 8000 of 26523
## 2022-02-11 15:30:57 : Done 9000 of 26523
## 2022-02-11 15:31:02 : Done 10000 of 26523
## 2022-02-11 15:31:07 : Done 11000 of 26523
## 2022-02-11 15:31:12 : Done 12000 of 26523
## 2022-02-11 15:31:17 : Done 13000 of 26523
## 2022-02-11 15:31:22 : Done 14000 of 26523
## 2022-02-11 15:31:27 : Done 15000 of 26523
## 2022-02-11 15:31:31 : Done 16000 of 26523
## 2022-02-11 15:31:36 : Done 17000 of 26523
## 2022-02-11 15:31:41 : Done 18000 of 26523
## 2022-02-11 15:31:45 : Done 19000 of 26523
## 2022-02-11 15:31:49 : Done 20000 of 26523
## 2022-02-11 15:31:53 : Done 21000 of 26523
## 2022-02-11 15:31:56 : Done 22000 of 26523
## 2022-02-11 15:31:59 : Done 23000 of 26523
## 2022-02-11 15:32:02 : Done 24000 of 26523
## 2022-02-11 15:32:04 : Done 25000 of 26523
## 2022-02-11 15:32:07 : Done 26000 of 26523
## 2022-02-11 15:32:08 : Done OF Orthologues
## 
## Writing results files
## -----
## 2022-02-11 15:32:20 : Done orthologues
## 
## Results:
##     /Users/jlovell/Desktop/GENESPACE_data/results/grasses/orthofinder/Results_Feb11/

```

```

## 
## CITATION:
## When publishing work that uses OrthoFinder please cite:
## Emms D.M. & Kelly S. (2019), Genome Biology 20:238
## 
## If you use the species tree in your work then please also cite:
## Emms D.M. & Kelly S. (2017), MBE 34(12): 3267–3278
## Emms D.M. & Kelly S. (2018), bioRxiv https://doi.org/10.1101/267914

```

## 6 Build synteny data

```

gparGrasses <- find_orthofinderResults(gsParam = gparGrasses)
gparGrasses <- synteny(gsParam = gparGrasses, overwrite = T)

## Parsing the gff files ...
## Reading the gffs and adding orthofinder IDs ... Done!
## Found 84963 global OGs for 397635 genes
## QC-ing genome to ensure chromosomes/scaffolds are big enough...
##           Genome: n. chrs PASS/FAIL, n. genes PASS/FAIL, n. OGs PASS/FAIL
##           Phallii: 10/5, 33253/10, 29545/9
##           Sorghum: 14/80, 34115/96, 29721/95
##           Sviridis: 9/1, 38330/4, 32779/1
##           brachy: 5/5, 34303/7, 28764/7
##           maize: 28/124, 34057/237, 30031/219
##           rice: 15/11, 35566/28, 30812/26
##           switchgrass: 45/68, 80108/170, 69114/138
##           wheat: 46/292, 106863/488, 82182/435
## All look good!
## Defining collinear orthogroup arrays ...
## Found the following counts of arrays / genome:
##           Phallii: 4593 genes in 1849 collinear arrays
##           Sorghum: 5569 genes in 2064 collinear arrays
##           Sviridis: 6514 genes in 2459 collinear arrays
##           brachy: 4927 genes in 1904 collinear arrays
##           maize: 3532 genes in 1412 collinear arrays
##           rice: 6341 genes in 2226 collinear arrays
##           switchgrass: 11848 genes in 4699 collinear arrays
##           wheat: 31292 genes in 11472 collinear arrays
## Pulling synteny for 36 unique pairwise combinations of genomes
## Running 9 chunks of 4 combinations each:
## Chunk 1 / 9 (15:33:21) ... Done!
## wheat -wheat : 2625895 (tot), 576934/488 (reg), 213625/1987 (blk)
## wheat -switchgr: 1902001 (tot), 158798/309 (reg), 72253/1372 (blk)
## switchgr-switchgr: 1437845 (tot), 226076/139 (reg), 125442/294 (blk)
## wheat -Sviridis: 1446237 (tot), 94894/176 (reg), 37649/806 (blk)
## Chunk 2 / 9 (15:40:35) ... Done!
## wheat -rice : 1426989 (tot), 99814/125 (reg), 38128/696 (blk)
## wheat -brachy : 1413741 (tot), 105555/152 (reg), 38561/776 (blk)
## wheat -maize : 1359495 (tot), 88837/270 (reg), 39648/1204 (blk)
## wheat -Sorghum : 1410253 (tot), 93260/139 (reg), 36566/693 (blk)
## Chunk 3 / 9 (15:43:29) ... Done!
## wheat -Phallii : 1367394 (tot), 88200/153 (reg), 37663/729 (blk)

```

```

##  switchgr-Sviridis: 883064 (tot), 100866/70 (reg), 43901/321 (blk)
##  switchgr-rice     : 835715 (tot), 85345/54 (reg), 37682/290 (blk)
##  switchgr-brachy   : 829700 (tot), 70980/88 (reg), 33696/465 (blk)
##  Chunk 4 / 9 (15:45:52) ... Done!
##  switchgr-maize    : 862406 (tot), 78972/160 (reg), 41769/699 (blk)
##  switchgr-Sorghum  : 830479 (tot), 89472/44 (reg), 39650/148 (blk)
##  switchgr-Phallii  : 852576 (tot), 94598/26 (reg), 45544/239 (blk)
##  Sviridis-Sviridis: 538615 (tot), 72625/10 (reg), 38327/10 (blk)
##  Chunk 5 / 9 (15:47:37) ... Done!
##  Sviridis-rice     : 489053 (tot), 48053/43 (reg), 19787/221 (blk)
##  Sviridis-brachy   : 476705 (tot), 39557/59 (reg), 17543/298 (blk)
##  Sviridis-maize    : 485794 (tot), 42849/102 (reg), 21420/434 (blk)
##  Sviridis-Sorghum  : 479819 (tot), 48667/38 (reg), 20451/138 (blk)
##  Chunk 6 / 9 (15:48:40) ... Done!
##  Sviridis-Phallii  : 461715 (tot), 50109/38 (reg), 22771/205 (blk)
##  rice   -rice      : 491825 (tot), 72441/26 (reg), 35579/26 (blk)
##  rice   -brachy    : 457054 (tot), 40069/32 (reg), 17728/222 (blk)
##  rice   -maize     : 470978 (tot), 38756/81 (reg), 20030/376 (blk)
##  Chunk 7 / 9 (15:49:32) ... Done!
##  rice   -Sorghum   : 461197 (tot), 42965/22 (reg), 18415/120 (blk)
##  rice   -Phallii   : 444201 (tot), 42781/26 (reg), 19344/170 (blk)
##  brachy -brachy   : 468654 (tot), 57180/10 (reg), 34306/10 (blk)
##  brachy -maize    : 448306 (tot), 35146/102 (reg), 18814/489 (blk)
##  Chunk 8 / 9 (15:50:19) ... Done!
##  maize -maize     : 557804 (tot), 66719/224 (reg), 44226/428 (blk)
##  brachy -Sorghum  : 442140 (tot), 37017/39 (reg), 16881/238 (blk)
##  maize -Sorghum   : 459158 (tot), 43999/76 (reg), 22771/346 (blk)
##  Sorghum -Sorghum : 473892 (tot), 62818/94 (reg), 34204/94 (blk)
##  Chunk 9 / 9 (15:51:14) ... Done!
##  brachy -Phallii  : 425853 (tot), 35553/42 (reg), 17269/249 (blk)
##  maize -Phallii   : 451768 (tot), 40446/82 (reg), 21399/381 (blk)
##  Sorghum -Phallii : 431866 (tot), 43719/21 (reg), 20372/95 (blk)
##  Phallii -Phallii : 445814 (tot), 55234/15 (reg), 33258/15 (blk)
## Defining synteny-constrained orthogroups ...
## Found 146402 synteny-split OGs for 397635 genes
## Running orthofinder by region ...
## genome combinat. : n. non-self genes, nOGs global/syntenic/inblk
## Phallii -Phallii : no non-self syn. regions
## switchgr-Phallii : 71664 genes, 26065 / 29436 / 27757
## switchgr-switchgr: 46791 genes, 24233 / 26680 / 24499
## switchgr-Sviridis: 69683 genes, 25538 / 28824 / 27285
## switchgr-Sorghum  : 63247 genes, 23585 / 26070 / 24824
## switchgr-maize    : 61234 genes, 21768 / 24264 / 22806
## switchgr-rice     : 61043 genes, 23217 / 25635 / 24432
## switchgr-brachy   : 56819 genes, 21768 / 24149 / 23090
## Sviridis-Phallii  : 45793 genes, 21469 / 23936 / 23817
## Sviridis-Sviridis: no non-self syn. regions
## Sviridis-Sorghum  : 42370 genes, 20325 / 22464 / 22334
## Sviridis-maize    : 41607 genes, 18538 / 20390 / 19655
## Sviridis-rice     : 41214 genes, 20162 / 22222 / 21854
## Sviridis-brachy   : 38042 genes, 18818 / 20739 / 20311
## Sorghum -Phallii  : 41444 genes, 19898 / 21844 / 21641
## Sorghum -Sorghum  : no non-self syn. regions
## maize -Phallii   : 41505 genes, 18515 / 20274 / 19513

```

```

## maize -Sorghum : 44176 genes, 19384 / 21344 / 20642
## maize -maize   : 12325 genes, 7206 / 7679 / 6751
## rice  -Phallii : 40397 genes, 19816 / 21737 / 21334
## rice  -Sorghum : 38926 genes, 19179 / 21055 / 20667
## rice  -maize   : 39463 genes, 17820 / 19548 / 18619
## rice  -rice    : no non-self syn. regions
## rice  -brachy  : 38126 genes, 18726 / 20597 / 20175
## brachy -Phallii : 37126 genes, 18431 / 20147 / 19762
## brachy -Sorghum : 36478 genes, 18165 / 19914 / 19445
## brachy -maize   : 37220 genes, 16915 / 18549 / 17646
## brachy -brachy  : no non-self syn. regions
## wheat -Phallii : 59300 genes, 21830 / 23399 / 21904
## wheat -switchgr: 75567 genes, 24307 / 26028 / 23210
## wheat -Sviridis: 59984 genes, 22138 / 23825 / 22301
## wheat -Sorghum : 59693 genes, 22113 / 23868 / 22463
## wheat -maize   : 57685 genes, 20710 / 22430 / 20775
## wheat -rice    : 61361 genes, 22363 / 24090 / 22553
## wheat -brachy  : 63940 genes, 23112 / 25303 / 24084
## wheat -wheat   : 69842 genes, 29563 / 33763 / 28412
## Combining synteny-constrained and inblock orthogroups ...
## syn OGs: 146402, inblk OGs: 163878, combined OGs: 130771
## Found the following counts of arrays / genome:
##     Phallii: 6052 genes in 2326 collinear arrays
##     Sorghum: 6792 genes in 2423 collinear arrays
##     Sviridis: 8025 genes in 2883 collinear arrays
##     brachy: 5915 genes in 2254 collinear arrays
##     maize: 4294 genes in 1733 collinear arrays
##     rice: 7448 genes in 2542 collinear arrays
##     switchgrass: 15114 genes in 5727 collinear arrays
##     wheat: 35509 genes in 12611 collinear arrays
## Pulling synteny for 36 unique pairwise combinations of genomes
## Running 9 chunks of 4 combinations each:
## Chunk 1 / 9 (17:04:44) ... Done!
## wheat -wheat   : 2625895 (tot), 577605/483 (reg), 205545/1856 (blk)
## wheat -switchgr: 1902001 (tot), 161256/305 (reg), 67280/1382 (blk)
## switchgr-switchgr: 1437845 (tot), 226326/139 (reg), 121832/281 (blk)
## wheat -Sviridis: 1446237 (tot), 96278/172 (reg), 34688/802 (blk)
## Chunk 2 / 9 (17:09:58) ... Done!
## wheat -rice    : 1426989 (tot), 100908/123 (reg), 35480/695 (blk)
## wheat -brachy  : 1413741 (tot), 106277/147 (reg), 35990/763 (blk)
## wheat -maize   : 1359495 (tot), 91688/278 (reg), 38560/1233 (blk)
## wheat -Sorghum : 1410253 (tot), 94509/136 (reg), 33842/686 (blk)
## Chunk 3 / 9 (17:12:42) ... Done!
## wheat -Phallii : 1367394 (tot), 89728/149 (reg), 34488/733 (blk)
## switchgr-Sviridis: 883064 (tot), 101141/69 (reg), 40054/317 (blk)
## switchgr-rice   : 835715 (tot), 86089/54 (reg), 34893/286 (blk)
## switchgr-brachy : 829700 (tot), 71288/88 (reg), 31564/455 (blk)
## Chunk 4 / 9 (17:14:31) ... Done!
## switchgr-maize  : 862406 (tot), 79561/161 (reg), 40257/695 (blk)
## switchgr-Sorghum: 830479 (tot), 90117/43 (reg), 36408/142 (blk)
## switchgr-Phallii: 852576 (tot), 95100/26 (reg), 41622/224 (blk)
## Sviridis-Sviridis: 538615 (tot), 72773/10 (reg), 38327/10 (blk)
## Chunk 5 / 9 (17:15:55) ... Done!
## Sviridis-rice   : 489053 (tot), 48295/43 (reg), 18254/208 (blk)

```

```

## Sviridis-brachy : 476705 (tot), 39901/57 (reg), 16449/290 (blk)
## Sviridis-maize : 485794 (tot), 42894/101 (reg), 20599/427 (blk)
## Sviridis-Sorghum : 479819 (tot), 49152/37 (reg), 18833/138 (blk)
## Chunk 6 / 9 (17:16:50) ... Done!
## Sviridis-Phallii : 461715 (tot), 50298/37 (reg), 20682/196 (blk)
## rice -rice : 491825 (tot), 72601/26 (reg), 35579/26 (blk)
## rice -brachy : 457054 (tot), 40474/33 (reg), 16791/224 (blk)
## rice -maize : 470978 (tot), 39364/85 (reg), 19482/386 (blk)
## Chunk 7 / 9 (17:17:35) ... Done!
## rice -Sorghum : 461197 (tot), 43273/22 (reg), 17172/117 (blk)
## rice -Phallii : 444201 (tot), 43061/28 (reg), 17913/168 (blk)
## brachy -brachy : 468654 (tot), 57258/10 (reg), 34306/10 (blk)
## brachy -maize : 448306 (tot), 35229/103 (reg), 18168/489 (blk)
## Chunk 8 / 9 (17:18:25) ... Done!
## maize -maize : 557804 (tot), 66931/224 (reg), 43973/432 (blk)
## brachy -Sorghum : 442140 (tot), 37199/39 (reg), 15807/223 (blk)
## maize -Sorghum : 459158 (tot), 44179/78 (reg), 21945/336 (blk)
## Sorghum -Sorghum : 473892 (tot), 62946/94 (reg), 34204/94 (blk)
## Chunk 9 / 9 (17:19:15) ... Done!
## brachy -Phallii : 425853 (tot), 35939/43 (reg), 16226/248 (blk)
## maize -Phallii : 451768 (tot), 40665/83 (reg), 20602/372 (blk)
## Sorghum -Phallii : 431866 (tot), 43956/21 (reg), 18585/93 (blk)
## Phallii -Phallii : 445814 (tot), 55374/15 (reg), 33258/15 (blk)
## Found 130771 OGs across 397635 genes. gff3-like text file written to:
## /Users/jlovell/Desktop/GENESPACE_data/results//grasses/results/gffWithOgs.txt.gz
## Calculating syntenic block breakpoints ...
## Found 29346 blocks. Text file written to:
## /Users/jlovell/Desktop/GENESPACE_data/results//grasses/results/syntenicBlocks.txt.gz:

```

## 7 Build pangenomes against each genome

```

for(i in gparGrasses$genomes$genomeIDs)
  tmp <- panGenome(gsParam = gparGrasses, refGenome = i)

## Building reference-anchored scaffold against Phallii
## n. ref positions = 29537
## Reading in hits against Phallii ... found 210847
## Interpolating positions ... n. genes mapped: 1x = 285126, 2+x = 8676, 0x = 47183
## Forming ref.-anchored db ... found 190314 genes for 29580 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 15826 genes and 7876 placements
## Checking missing direct ref. syn. OGs ... found 3876 genes and 2288 placements
## Adding indirect syn. OGs ... found 419 genes and 267 placements
## Adding syn. OGs without ref. anchor ... found 119146 genes and 95032 placements
## Adding missing genes by synOG identity ... found 11404 genes and 262 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 16334 placements
## Adding non-syn. orthologs ... found 121562 genes and 48368 placements
## Writing panGenome to results/Phallii_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## *NOTE* RefGenome switchgrass has >1x ploidy - this is fine, but will be slower.

```

```

## Depending on the size of your run, you may run into memory issues.
## Building reference-anchored scaffold against switchgrass
## n. ref positions = 70891
## Reading in hits against switchgrass ... found 427687
## Interpolating positions ... n. genes mapped: 1x = 73876, 2+x = 253126, 0x = 13983
## Forming ref.-anchored db ... found 217942 genes for 99692 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 12775 genes and 15896 placements
## Checking missing direct ref. syn. OGs ... found 3201 genes and 4363 placements
## Adding indirect syn. OGs ... found 369 genes and 510 placements
## Adding syn. OGs without ref. anchor ... found 98661 genes and 152656 placements
## Adding missing genes by synOG identity ... found 8037 genes and 398 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 32502 placements
## Adding non-syn. orthologs ... found 96207 genes and 88509 placements
## Writing pangenome to results/switchgrass_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## Building reference-anchored scaffold against Sviridis
## n. ref positions = 33192
## Reading in hits against Sviridis ... found 214612
## Interpolating positions ... n. genes mapped: 1x = 283599, 2+x = 7208, 0x = 50178
## Forming ref.-anchored db ... found 193634 genes for 33289 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 16444 genes and 8100 placements
## Checking missing direct ref. syn. OGs ... found 4265 genes and 2656 placements
## Adding indirect syn. OGs ... found 464 genes and 325 placements
## Adding syn. OGs without ref. anchor ... found 115452 genes and 91949 placements
## Adding missing genes by synOG identity ... found 10726 genes and 263 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 16288 placements
## Adding non-syn. orthologs ... found 122748 genes and 47843 placements
## Writing pangenome to results/Sviridis_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## Building reference-anchored scaffold against Sorghum
## n. ref positions = 29842
## Reading in hits against Sorghum ... found 204127
## Interpolating positions ... n. genes mapped: 1x = 287497, 2+x = 7291, 0x = 46197
## Forming ref.-anchored db ... found 184824 genes for 30003 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 15604 genes and 7355 placements
## Checking missing direct ref. syn. OGs ... found 3427 genes and 2069 placements
## Adding indirect syn. OGs ... found 378 genes and 235 placements
## Adding syn. OGs without ref. anchor ... found 126657 genes and 95821 placements
## Adding missing genes by synOG identity ... found 10095 genes and 310 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 16532 placements
## Adding non-syn. orthologs ... found 123911 genes and 49034 placements
## Writing pangenome to results/Sorghum_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## *NOTE* RefGenome maize has >1x ploidy - this is fine, but will be slower.
## Depending on the size of your run, you may run into memory issues.

```

```

## Building reference-anchored scaffold against maize
## n. ref positions = 31733
## Reading in hits against maize ... found 242739
## Interpolating positions ... n. genes mapped: 1x = 80348, 2+x = 235789, 0x = 24848
## Forming ref.-anchored db ... found 167316 genes for 49115 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 14994 genes and 11837 placements
## Checking missing direct ref. syn. OGs ... found 4718 genes and 4111 placements
## Adding indirect syn. OGs ... found 387 genes and 391 placements
## Adding syn. OGs without ref. anchor ... found 141209 genes and 172586 placements
## Adding missing genes by synOG identity ... found 12361 genes and 433 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 29779 placements
## Adding non-syn. orthologs ... found 103155 genes and 84360 placements
## Writing pangenome to results/maize_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## Building reference-anchored scaffold against rice
## n. ref positions = 30688
## Reading in hits against rice ... found 202537
## Interpolating positions ... n. genes mapped: 1x = 286444, 2+x = 8261, 0x = 46280
## Forming ref.-anchored db ... found 181661 genes for 30763 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 16259 genes and 7454 placements
## Checking missing direct ref. syn. OGs ... found 3928 genes and 2261 placements
## Adding indirect syn. OGs ... found 482 genes and 307 placements
## Adding syn. OGs without ref. anchor ... found 128663 genes and 95564 placements
## Adding missing genes by synOG identity ... found 9992 genes and 254 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 16641 placements
## Adding non-syn. orthologs ... found 124737 genes and 49135 placements
## Writing pangenome to results/rice_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## Building reference-anchored scaffold against brachy
## n. ref positions = 30649
## Reading in hits against brachy ... found 194607
## Interpolating positions ... n. genes mapped: 1x = 281115, 2+x = 5567, 0x = 54303
## Forming ref.-anchored db ... found 175135 genes for 30685 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 17243 genes and 7147 placements
## Checking missing direct ref. syn. OGs ... found 5263 genes and 2714 placements
## Adding indirect syn. OGs ... found 594 genes and 337 placements
## Adding syn. OGs without ref. anchor ... found 128110 genes and 90838 placements
## Adding missing genes by synOG identity ... found 14640 genes and 377 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 16062 placements
## Adding non-syn. orthologs ... found 123963 genes and 47081 placements
## Writing pangenome to results/brachy_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!
## *NOTE* RefGenome wheat has >1x ploidy - this is fine, but will be slower.
## Depending on the size of your run, you may run into memory issues.
## Building reference-anchored scaffold against wheat

```

```

## n. ref positions = 84453
## Reading in hits against wheat ... found 508062
## Interpolating positions ... n. genes mapped: 1x = 16500, 2+x = 274947, 0x = 49538
## Forming ref.-anchored db ... found 196231 genes for 134997 placements
## Completing the pan-genome annotation ...
## Adding non-anchor entries ... found 9997 genes and 14379 placements
## Checking missing direct ref. syn. OGs ... found 4249 genes and 5282 placements
## Adding indirect syn. OGs ... found 427 genes and 492 placements
## Adding syn. OGs without ref. anchor ... found 98010 genes and 180649 placements
## Adding missing genes by synOG identity ... found 32071 genes and 1170 placements
## Annotating and formatting pan-genome
## Adding non-anchor entries ... found 56650 genes and 43845 placements
## Adding non-syn. orthologs ... found 100496 genes and 84392 placements
## Writing pangenome to results/wheat_pangenomeDB.txt.gz
## Returning wide-format with only syntenic array reps
## Done!

```

## 8 Make riparian plot

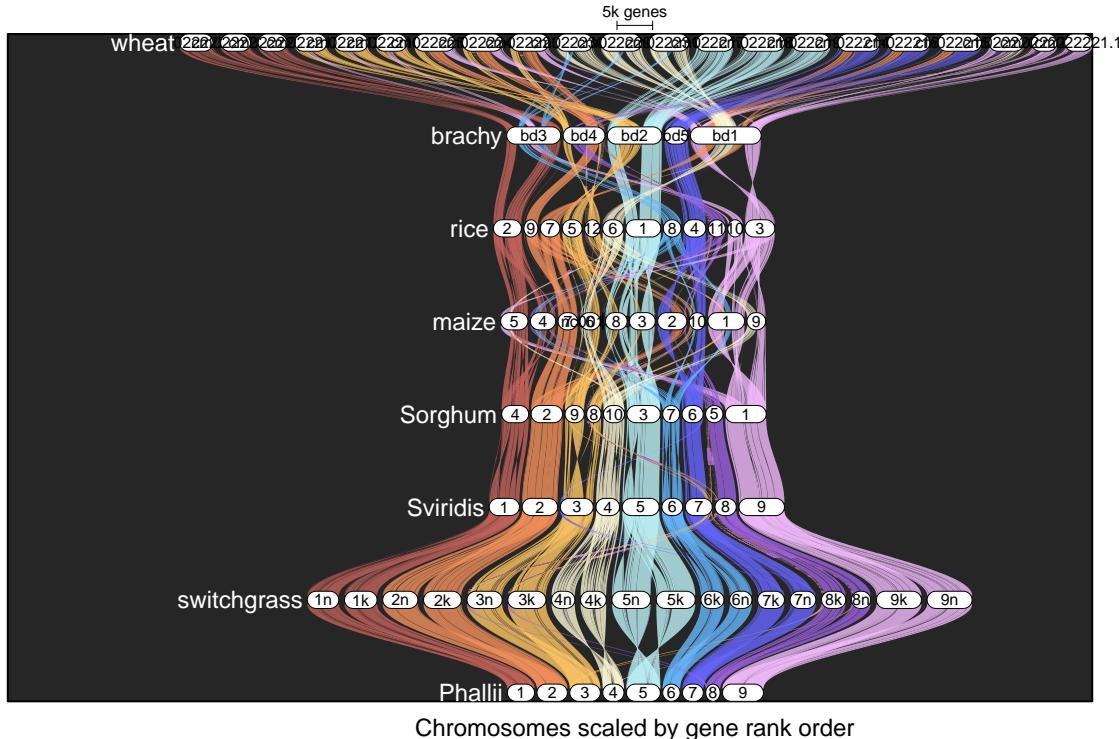
Also, write the parameters to file

```
plot_riparianHits(gsParam = gparGrasses)
```

```

## Reading hits ... Done!
## Generating plot coordinates ... Done!
## Rendering plot ...

```



```
## Done!
```

```
save(gparGrasses, file = file.path(baseDir, "gparGrasses.rda"))
```

## 9 Print session info

```
sessionInfo()

## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics   grDevices utils      datasets   methods    base
##
## other attached packages:
## [1] GENESPACE_0.9.3   data.table_1.14.2
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.8           highr_0.9
## [3] compiler_4.1.2       restfulr_0.0.13
## [5] GenomeInfoDb_1.30.1  XVector_0.34.0
## [7] MatrixGenerics_1.6.0 R.methodsS3_1.8.1
## [9] bitops_1.0-7         R.utils_2.11.0
## [11] tools_4.1.2          zlibbioc_1.40.0
## [13] digest_0.6.29        lattice_0.20-45
## [15] evaluate_0.14        pkgconfig_2.0.3
## [17] rlang_1.0.0          Matrix_1.4-0
## [19] igraph_1.2.11        DelayedArray_0.20.0
## [21] cli_3.1.1            rstudioapi_0.13
## [23] yaml_2.2.2           parallel_4.1.2
## [25] xfun_0.29            fastmap_1.1.0
## [27] GenomeInfoDbData_1.2.7 rtracklayer_1.54.0
## [29] stringr_1.4.0         knitr_1.37
## [31] Biostrings_2.62.0     S4Vectors_0.32.3
## [33] IRanges_2.28.0        grid_4.1.2
## [35] stats4_4.1.2          Biobase_2.54.0
## [37] BiocParallel_1.28.3   XML_3.99-0.8
## [39] rmarkdown_2.11         magrittr_2.0.2
## [41] matrixStats_0.61.0    GenomicAlignments_1.30.0
## [43] Rsamtools_2.10.0      GenomicRanges_1.46.1
## [45] htmltools_0.5.2       BiocGenerics_0.40.0
## [47] SummarizedExperiment_1.24.0 stringi_1.7.6
## [49] RCurl_1.98-1.5        rjson_0.2.21
## [51] crayon_1.4.2          dbscan_1.1-10
## [53] BiocIO_1.4.0          R.oo_1.24.0
```