Run GENESPACE vignettes- 17 Vertebrate genomes

JTLovell

12-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 <- 6
path2of <- "orthofinder" # since orthofinder is in the path via conda
```

2. Set parameters

```
vertebrates <- list(
  wd = file.path(baseDir, "vertebrates"),
  speciesIDs = "vertebrates",
  genomes = data.table(do.call("rbind", list(
    # lizards
    c(genome = "garterSnake", version = "Thamnophis_elegans_rThaEle1.pri", ploidy = 1),
    c("sandLizard", "Lacerta_agilis_rLacAgi1.pri", 1),
    # birds
    c("chicken", "Gallus_gallus_bGalGal1.mat.broiler.GRCg7b", 1),
    c("hummingbird", "Calypte_anna_bCalAnn1_v1.p", 1),
    c("budgie", "Melopsittacus_undulatus_bMelUnd1.mat.Z", 1),
    c("swan", "Cygnus_olor_bCygOlo1.pri.v2", 1),
    c("zebrafinch", "Taeniopygia_guttata_bTaeGut1.4.pri", 1),
    # placental mammals
    c("human", "Homo_sapiens_GRCh38.p13", 1),
    c("mouse", "Mus_musculus_GRCm39", 1),
    c("sloth", "Choloepus_didactylus_mChoDid1.pri", 1),
    c("horseshoeBat", "Rhinolophus_ferrumequinum_mRhiFer1_v1.p", 1),
    c("dolphin", "Tursiops_truncatus_mTurTru1.mat.Y", 1),
    # non-placentals
    c("echidna", "Tachyglossus_aculeatus_mTacAcu1.pri", 1),
    c("platypus", "Ornithorhynchus_anatinus_mOrnAna1.pri.v4", 1),
    c("brushtailPossum", "Trichosurus_vulpecula_mTriVul1.pri", 1),
    c("opossum", "Monodelphis_domestica_MonDom5", 1),
    c("tasmaniandevil", "Sarcophilus_harrisii_mSarHar1.11", 1)))))
```

3 Initialize the Verts run with the above specified parameters

```
gparVerts <- with(vertebrates, init_genespace(
  genomeIDs = genomes$genome,
  versionIDs = genomes$version,
  ploidy = genomes$ploidy,
  speciesIDs = rep(speciesIDs, length(genomes$genome)),
  pepString = "translated_cds",
  wd = wd,
  nCores = nThreads,
  path2orthofinder = path2of,
  path2mcscanx = mcscanDir,
  rawGenomeDir = rawAnnotationDir))
```

```
## set working directory to /Users/jlovell/Desktop/GENESPACE_data/results/vertebrates
##
## found raw gff files:
##    /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Thamnophis_elegans_rThaEle1.pri/annotation/GCF_0(
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Lacerta_agilis_rLacAgi1.pri/annotation/GCF_0098195
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Gallus_gallus_bGalGal1.mat.broiler.GRCg7b/annotati
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Calypte_anna_bCalAnn1_v1.p/annotation/GCF_00395755
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Melopsittacus_undulatus_bMelUnd1.mat.Z/annotation/
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Cygnus_olor_bCygOlo1.pri.v2/annotation/GCF_0097690
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Taeniopygia_guttata_bTaeGut1.4.pri/annotation/GCF_
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Homo_sapiens_GRCh38.p13/annotation/GCF_000001405.
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Mus_musculus_GRCm39/annotation/GCF_000001635.27_GF
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Choloepus_didactylus_mChoDid1.pri/annotation/GCF_(
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Rhinolophus_ferrumequinum_mRhiFer1_v1.p/annotatior
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Tursiops_truncatus_mTurTru1.mat.Y/annotation/GCF_(
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Tachyglossus_aculeatus_mTacAcu1.pri/annotation/GCF
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Ornithorhynchus_anatinus_mOrnAna1.pri.v4/annotatic
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Trichosurus_vulpecula_mTriVul1.pri/annotation/GCF_
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Monodelphis_domestica_MonDom5/annotation/GCF_00000
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Sarcophilus_harrisii_mSarHar1.11/annotation/GCF_90
##
## found raw peptide files:
##    /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Thamnophis_elegans_rThaEle1.pri/annotation/GCF_0(
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Lacerta_agilis_rLacAgi1.pri/annotation/GCF_0098195
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Gallus_gallus_bGalGal1.mat.broiler.GRCg7b/annotati
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Calypte_anna_bCalAnn1_v1.p/annotation/GCF_00395755
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Melopsittacus_undulatus_bMelUnd1.mat.Z/annotation/
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Cygnus_olor_bCygOlo1.pri.v2/annotation/GCF_0097690
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Taeniopygia_guttata_bTaeGut1.4.pri/annotation/GCF_
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Homo_sapiens_GRCh38.p13/annotation/GCF_000001405.
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Mus_musculus_GRCm39/annotation/GCF_000001635.27_GF
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Choloepus_didactylus_mChoDid1.pri/annotation/GCF_(
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Rhinolophus_ferrumequinum_mRhiFer1_v1.p/annotatior
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Tursiops_truncatus_mTurTru1.mat.Y/annotation/GCF_(
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Tachyglossus_aculeatus_mTacAcu1.pri/annotation/GCF
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Ornithorhynchus_anatinus_mOrnAna1.pri.v4/annotatic
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Trichosurus_vulpecula_mTriVul1.pri/annotation/GCF_
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Monodelphis_domestica_MonDom5/annotation/GCF_00000
##   /Users/jlovell/Desktop/GENESPACE_data/vertebrates/Sarcophilus_harrisii_mSarHar1.11/annotation/GCF_90
```

```
##
##
## Can't find all parsed annotation files ... need to run parse_annotations, parse_ncbi or parse_phytozo
##
## GENESPACE run initialized:
##   Initial orthofinder database generation method: default inside R
##   Orthology graph method: global
```

4 Parse the raw annotations.

Here, we need to specify some custom fields to get the right output. Also, for simplicity, we'll drop any chromosomes with the string "alt", as the ploidy in these regions is known to be > 1.

```
parse_ncbi(gsParam = gparVerts)
```

```
## Parsing annotations: garterSnake
##   Reading gff ... found 18900 protein coding genes
##   Reading peptide fasta ... found 31010 / 18900 total and unique entries
##   Merging fa and gff found 18900 matching entires
## Parsing annotations: sandLizard
##   Reading gff ... found 20325 protein coding genes
##   Reading peptide fasta ... found 39306 / 20325 total and unique entries
##   Merging fa and gff found 20325 matching entires
## Parsing annotations: chicken
##   Reading gff ... found 17807 protein coding genes
##   Reading peptide fasta ... found 59322 / 17807 total and unique entries
##   Merging fa and gff found 17805 matching entires
## Parsing annotations: hummingbird
##   Reading gff ... found 14711 protein coding genes
##   Reading peptide fasta ... found 29231 / 14711 total and unique entries
##   Merging fa and gff found 14711 matching entires
## Parsing annotations: budgie
##   Reading gff ... found 16458 protein coding genes
##   Reading peptide fasta ... found 29187 / 16458 total and unique entries
##   Merging fa and gff found 16458 matching entires
## Parsing annotations: swan
##   Reading gff ... found 16150 protein coding genes
##   Reading peptide fasta ... found 47851 / 16150 total and unique entries
##   Merging fa and gff found 16150 matching entires
## Parsing annotations: zebrafinch
##   Reading gff ... found 16520 protein coding genes
##   Reading peptide fasta ... found 41237 / 16520 total and unique entries
##   Merging fa and gff found 16520 matching entires
## Parsing annotations: human
##   Reading gff ... found 19532 protein coding genes
##   Reading peptide fasta ... found 124038 / 19532 total and unique entries
##   Merging fa and gff found 19532 matching entires
## Parsing annotations: mouse
##   Reading gff ... found 22186 protein coding genes
##   Reading peptide fasta ... found 93165 / 22186 total and unique entries
##   Merging fa and gff found 22186 matching entires
## Parsing annotations: sloth
##   Reading gff ... found 23549 protein coding genes
```

```
##   Reading peptide fasta ... found 54704 / 23549 total and unique entries
##   Merging fa and gff found 23549 matching entires
## Parsing annotations: horseshoeBat
##   Reading gff ... found 19651 protein coding genes
##   Reading peptide fasta ... found 48805 / 19651 total and unique entries
##   Merging fa and gff found 19651 matching entires
## Parsing annotations: dolphin
##   Reading gff ... found 19240 protein coding genes
##   Reading peptide fasta ... found 55802 / 19240 total and unique entries
##   Merging fa and gff found 19240 matching entires
## Parsing annotations: echidna
##   Reading gff ... found 18929 protein coding genes
##   Reading peptide fasta ... found 33397 / 18929 total and unique entries
##   Merging fa and gff found 18929 matching entires
## Parsing annotations: platypus
##   Reading gff ... found 18152 protein coding genes
##   Reading peptide fasta ... found 38847 / 18152 total and unique entries
##   Merging fa and gff found 18152 matching entires
## Parsing annotations: brushtailPossum
##   Reading gff ... found 22532 protein coding genes
##   Reading peptide fasta ... found 35644 / 22532 total and unique entries
##   Merging fa and gff found 22532 matching entires
## Parsing annotations: opossum
##   Reading gff ... found 20903 protein coding genes
##   Reading peptide fasta ... found 49302 / 20903 total and unique entries
##   Merging fa and gff found 20903 matching entires
## Parsing annotations: tasmaniandevil
##   Reading gff ... found 19966 protein coding genes
##   Reading peptide fasta ... found 46002 / 19966 total and unique entries
##   Merging fa and gff found 19966 matching entires
```

5 Run OrthoFinder from within R.

Also set the synteny parameters as default.

```
gparVerts <- set_syntenyParams(gsParam = gparVerts)
gparVerts <- run_orthofinder(gsParam = gparVerts)
```

```
##   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
##
##   2022-02-12 14:04:50 : Starting OrthoFinder 2.5.4
##   6 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//vertebrates/orthofinder/Resul
##
## Dividing up work for BLAST for parallel processing
## --------------------------------------------------
## 2022-02-12 14:04:54 : Creating diamond database 1 of 17
## 2022-02-12 14:04:54 : Creating diamond database 2 of 17
## 2022-02-12 14:04:54 : Creating diamond database 3 of 17
## 2022-02-12 14:04:55 : Creating diamond database 4 of 17
## 2022-02-12 14:04:55 : Creating diamond database 5 of 17
## 2022-02-12 14:04:55 : Creating diamond database 6 of 17
## 2022-02-12 14:04:55 : Creating diamond database 7 of 17
## 2022-02-12 14:04:56 : Creating diamond database 8 of 17
## 2022-02-12 14:04:56 : Creating diamond database 9 of 17
## 2022-02-12 14:04:56 : Creating diamond database 10 of 17
## 2022-02-12 14:04:57 : Creating diamond database 11 of 17
## 2022-02-12 14:04:57 : Creating diamond database 12 of 17
## 2022-02-12 14:04:57 : Creating diamond database 13 of 17
## 2022-02-12 14:04:57 : Creating diamond database 14 of 17
## 2022-02-12 14:04:58 : Creating diamond database 15 of 17
## 2022-02-12 14:04:58 : Creating diamond database 16 of 17
## 2022-02-12 14:04:58 : Creating diamond database 17 of 17
##
## Running diamond all-versus-all
## ------------------------------
## Using 6 thread(s)
## 2022-02-12 14:04:59 : This may take some time....
## 2022-02-12 14:04:59 : Done 0 of 289
## 2022-02-12 15:47:37 : Done 100 of 289
## 2022-02-12 17:10:20 : Done 200 of 289
## 2022-02-12 18:13:01 : Done all-versus-all sequence search
##
## Running OrthoFinder algorithm
## -----------------------------
## 2022-02-12 18:13:02 : Initial processing of each species
## 2022-02-12 18:14:04 : Initial processing of species 0 complete
## 2022-02-12 18:14:47 : Initial processing of species 1 complete
## 2022-02-12 18:15:32 : Initial processing of species 2 complete
## 2022-02-12 18:16:20 : Initial processing of species 3 complete
## 2022-02-12 18:17:13 : Initial processing of species 4 complete
## 2022-02-12 18:18:06 : Initial processing of species 5 complete
## 2022-02-12 18:19:01 : Initial processing of species 6 complete
## 2022-02-12 18:19:54 : Initial processing of species 7 complete
## 2022-02-12 18:20:35 : Initial processing of species 8 complete
## 2022-02-12 18:21:36 : Initial processing of species 9 complete
## 2022-02-12 18:22:35 : Initial processing of species 10 complete
## 2022-02-12 18:23:26 : Initial processing of species 11 complete
## 2022-02-12 18:24:21 : Initial processing of species 12 complete
## 2022-02-12 18:25:23 : Initial processing of species 13 complete
## 2022-02-12 18:26:05 : Initial processing of species 14 complete
## 2022-02-12 18:26:59 : Initial processing of species 15 complete
## 2022-02-12 18:27:40 : Initial processing of species 16 complete
## 2022-02-12 18:28:46 : Connected putative homologues
```

```
## 2022-02-12 18:28:55 : Written final scores for species 0 to graph file
## 2022-02-12 18:29:02 : Written final scores for species 1 to graph file
## 2022-02-12 18:29:09 : Written final scores for species 2 to graph file
## 2022-02-12 18:29:16 : Written final scores for species 3 to graph file
## 2022-02-12 18:29:24 : Written final scores for species 4 to graph file
## 2022-02-12 18:29:31 : Written final scores for species 5 to graph file
## 2022-02-12 18:29:39 : Written final scores for species 6 to graph file
## 2022-02-12 18:29:46 : Written final scores for species 7 to graph file
## 2022-02-12 18:29:52 : Written final scores for species 8 to graph file
## 2022-02-12 18:30:01 : Written final scores for species 9 to graph file
## 2022-02-12 18:30:09 : Written final scores for species 10 to graph file
## 2022-02-12 18:30:16 : Written final scores for species 11 to graph file
## 2022-02-12 18:30:24 : Written final scores for species 12 to graph file
## 2022-02-12 18:30:33 : Written final scores for species 13 to graph file
## 2022-02-12 18:30:39 : Written final scores for species 14 to graph file
## 2022-02-12 18:30:47 : Written final scores for species 15 to graph file
## 2022-02-12 18:30:53 : Written final scores for species 16 to graph file
## 2022-02-12 18:32:36 : Ran MCL
##
## Writing orthogroups to file
## ---------------------------
## OrthoFinder assigned 317999 genes (97.7% of total) to 19355 orthogroups. Fifty percent of all genes
##
## 2022-02-12 18:32:59 : Done orthogroups
##
## Analysing Orthogroups
## =====================
##
## Calculating gene distances
## --------------------------
## 2022-02-12 18:45:31 : Done
## 2022-02-12 18:45:35 : Done 0 of 17630
## 2022-02-12 18:45:47 : Done 1000 of 17630
## 2022-02-12 18:45:51 : Done 2000 of 17630
## 2022-02-12 18:45:56 : Done 3000 of 17630
## 2022-02-12 18:46:00 : Done 4000 of 17630
## 2022-02-12 18:46:04 : Done 5000 of 17630
## 2022-02-12 18:46:09 : Done 6000 of 17630
## 2022-02-12 18:46:13 : Done 7000 of 17630
## 2022-02-12 18:46:18 : Done 8000 of 17630
## 2022-02-12 18:46:22 : Done 9000 of 17630
## 2022-02-12 18:46:27 : Done 10000 of 17630
## 2022-02-12 18:46:31 : Done 11000 of 17630
## 2022-02-12 18:46:35 : Done 12000 of 17630
## 2022-02-12 18:46:40 : Done 13000 of 17630
## 2022-02-12 18:46:44 : Done 14000 of 17630
## 2022-02-12 18:46:48 : Done 15000 of 17630
## 2022-02-12 18:46:52 : Done 16000 of 17630
## 2022-02-12 18:46:56 : Done 17000 of 17630
##
## Inferring gene and species trees
## --------------------------------
##
## 10059 trees had all species present and will be used by STAG to infer the species tree
```

```
##
##  Best outgroup(s) for species tree
##  --------------------------------
##  2022-02-12 18:51:22 : Starting STRIDE
##  2022-02-12 18:51:32 : Done STRIDE
##  Observed 366 well-supported, non-terminal duplications. 355 support the best root and 11 contradict
##  Best outgroup for species tree:
##    chicken, swan, garterSnake, sandLizard, hummingbird, zebrafinch, budgie
##
##  Reconciling gene trees and species tree
##  ---------------------------------------
##  Outgroup: chicken, swan, garterSnake, sandLizard, hummingbird, zebrafinch, budgie
##  2022-02-12 18:51:32 : Starting Recon and orthologues
##  2022-02-12 18:51:32 : Starting OF Orthologues
##  2022-02-12 18:51:34 : Done 0 of 17630
##  2022-02-12 18:52:04 : Done 1000 of 17630
##  2022-02-12 18:52:20 : Done 2000 of 17630
##  2022-02-12 18:52:34 : Done 3000 of 17630
##  2022-02-12 18:52:46 : Done 4000 of 17630
##  2022-02-12 18:53:00 : Done 5000 of 17630
##  2022-02-12 18:53:12 : Done 6000 of 17630
##  2022-02-12 18:53:25 : Done 7000 of 17630
##  2022-02-12 18:53:38 : Done 8000 of 17630
##  2022-02-12 18:53:51 : Done 9000 of 17630
##  2022-02-12 18:54:04 : Done 10000 of 17630
##  2022-02-12 18:54:17 : Done 11000 of 17630
##  2022-02-12 18:54:30 : Done 12000 of 17630
##  2022-02-12 18:54:41 : Done 13000 of 17630
##  2022-02-12 18:54:52 : Done 14000 of 17630
##  2022-02-12 18:55:01 : Done 15000 of 17630
##  2022-02-12 18:55:08 : Done 16000 of 17630
##  2022-02-12 18:55:11 : Done 17000 of 17630
##  2022-02-12 18:55:13 : Done OF Orthologues
##
##  Writing results files
##  =====================
##  2022-02-12 18:55:21 : Done orthologues
##
##  Results:
##      /Users/jlovell/Desktop/GENESPACE_data/results/vertebrates/orthofinder/Results_Feb12/
##
##  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

```
gparVerts <- find_orthofinderResults(gsParam = gparVerts)
gparVerts <- synteny(gsParam = gparVerts)
```

```
## Can't find gff-like file, but overwriteGff = TRUE, setting to FALSE
## Can't find all synhits files, but overwriteHits = TRUE, setting to FALSE
## Can't find blk coords, but overwriteBlks = TRUE, setting to FALSE
## Parsing the gff files ...
##  Reading the gffs and adding orthofinder IDs ... Done!
##  Found 26865 global OGs for 325509 genes
##  QC-ing genome to ensure chromosomes/scaffolds are big enough...
##            Genome: n. chrs PASS/FAIL, n. genes PASS/FAIL, n. OGs PASS/FAIL
##      brushtailPossum: 20/46, 22460/72, 18729/65
##      budgie: 58/139, 16223/235, 15212/211
##      chicken: 56/84, 17536/269, 15448/149
##      dolphin: 27/33, 19166/74, 17985/50
##      echidna: 78/114, 18668/261, 16909/196
##      garterSnake: 30/162, 18631/269, 15861/248
##      horseshoeBat: 36/26, 19604/47, 18024/31
##      human: 25/58, 19441/91, 17284/69
##      hummingbird: 47/66, 14556/155, 13901/116
##      mouse: 22/15, 22134/52, 17859/21
##      opossum: 56/447, 20213/690, 16987/585
##      platypus: 38/46, 18037/115, 16522/69
##      sandLizard: 21/2, 20323/2, 17512/2
##      sloth: 51/37, 23482/67, 20447/60
##      swan: 60/65, 15930/220, 15074/122
##      tasmaniandevil: 10/14, 19948/18, 17146/15
##      zebrafinch: 47/123, 16286/234, 15406/200
##  All look good!
##  Defining collinear orthogroup arrays ...
##  Found the following counts of arrays / genome:
##      brushtailPossum: 3947 genes in 960 collinear arrays
##      budgie: 1455 genes in 453 collinear arrays
##      chicken: 2727 genes in 587 collinear arrays
##      dolphin: 1616 genes in 576 collinear arrays
##      echidna: 2526 genes in 731 collinear arrays
##      garterSnake: 3641 genes in 995 collinear arrays
##      horseshoeBat: 2166 genes in 685 collinear arrays
##      human: 2958 genes in 869 collinear arrays
##      hummingbird: 1051 genes in 372 collinear arrays
##      mouse: 5173 genes in 1000 collinear arrays
##      opossum: 3893 genes in 947 collinear arrays
##      platypus: 2190 genes in 657 collinear arrays
##      sandLizard: 3596 genes in 906 collinear arrays
##      sloth: 3819 genes in 1031 collinear arrays
##      swan: 1359 genes in 435 collinear arrays
##      tasmaniandevil: 3314 genes in 831 collinear arrays
##      zebrafinch: 1325 genes in 472 collinear arrays
## Pulling synteny for 153 unique pairwise combinations of genomes
##  Running 26 chunks of 6 combinations each:
##  Chunk 1 / 26 (18:55:51) ... Done!
```

```
##  sloth   -sloth   : 326443 (tot), 71817/88 (reg), 23541/88 (blk)
##  sloth   -brushtai: 321166 (tot), 53491/164 (reg), 15390/329 (blk)
##  sloth   -mouse   : 313919 (tot), 58491/169 (reg), 16485/269 (blk)
##  brushtai-brushtai: 332929 (tot), 82147/66 (reg), 22525/66 (blk)
##  brushtai-mouse   : 309490 (tot), 53951/254 (reg), 15598/431 (blk)
##  sloth   -opossum : 303256 (tot), 51752/176 (reg), 14644/399 (blk)
##  Chunk 2 / 26 (18:57:59) ... Done!
##  mouse   -mouse   : 321499 (tot), 102641/37 (reg), 22163/37 (blk)
##  sloth   -sandLiza: 304203 (tot), 37596/172 (reg), 13362/604 (blk)
##  sloth   -tasmania: 296358 (tot), 52232/153 (reg), 15439/332 (blk)
##  brushtai-opossum : 300904 (tot), 69665/83 (reg), 16331/235 (blk)
##  sloth   -horsesho: 287624 (tot), 49089/97 (reg), 16845/165 (blk)
##  mouse   -opossum : 296252 (tot), 49324/258 (reg), 14779/501 (blk)
##  Chunk 3 / 26 (19:01:01) ... Done!
##  sloth   -human   : 287486 (tot), 50328/78 (reg), 16952/191 (blk)
##  brushtai-sandLiza: 301054 (tot), 38213/191 (reg), 13595/597 (blk)
##  sloth   -dolphin : 281692 (tot), 36101/73 (reg), 16325/154 (blk)
##  mouse   -sandLiza: 288468 (tot), 33299/256 (reg), 13280/674 (blk)
##  brushtai-tasmania: 292650 (tot), 69967/40 (reg), 17045/92 (blk)
##  sloth   -echidna : 291402 (tot), 38535/234 (reg), 14084/500 (blk)
##  Chunk 4 / 26 (19:03:48) ... Done!
##  sloth   -garterSn: 288683 (tot), 32084/161 (reg), 12363/563 (blk)
##  brushtai-horsesho: 283020 (tot), 45826/185 (reg), 15845/350 (blk)
##  mouse   -tasmania: 287919 (tot), 52705/244 (reg), 15564/429 (blk)
##  brushtai-human   : 281624 (tot), 46087/170 (reg), 15932/375 (blk)
##  mouse   -horsesho: 269433 (tot), 45244/180 (reg), 16776/270 (blk)
##  opossum -opossum : 297137 (tot), 73813/501 (reg), 20887/501 (blk)
##  Chunk 5 / 26 (19:05:41) ... Done!
##  brushtai-dolphin : 276370 (tot), 32502/161 (reg), 15459/355 (blk)
##  mouse   -human   : 273944 (tot), 47101/170 (reg), 16905/289 (blk)
##  sloth   -platypus: 278140 (tot), 38804/233 (reg), 14503/497 (blk)
##  brushtai-echidna : 287329 (tot), 40038/237 (reg), 14202/490 (blk)
##  brushtai-garterSn: 285508 (tot), 36340/177 (reg), 12623/572 (blk)
##  mouse   -dolphin : 263762 (tot), 33004/165 (reg), 16285/256 (blk)
##  Chunk 6 / 26 (19:07:19) ... Done!
##  sloth   -chicken : 261938 (tot), 24281/143 (reg), 12373/527 (blk)
##  opossum -sandLiza: 286974 (tot), 36943/186 (reg), 12771/608 (blk)
##  mouse   -echidna : 281283 (tot), 37114/311 (reg), 14144/572 (blk)
##  mouse   -garterSn: 274277 (tot), 29067/245 (reg), 12299/630 (blk)
##  opossum -tasmania: 281852 (tot), 66824/74 (reg), 16484/225 (blk)
##  brushtai-platypus: 273443 (tot), 41481/240 (reg), 14686/489 (blk)
##  Chunk 7 / 26 (19:10:18) ... Done!
##  sandLiza-sandLiza: 286341 (tot), 70453/23 (reg), 20325/23 (blk)
##  opossum -horsesho: 265855 (tot), 43990/198 (reg), 14975/408 (blk)
##  opossum -human   : 269016 (tot), 43558/176 (reg), 15079/439 (blk)
##  mouse   -platypus: 266826 (tot), 37425/310 (reg), 14583/569 (blk)
##  brushtai-chicken : 256621 (tot), 26576/151 (reg), 12581/511 (blk)
##  sandLiza-tasmania: 280440 (tot), 36784/186 (reg), 13534/612 (blk)
##  Chunk 8 / 26 (19:11:54) ... Done!
##  opossum -dolphin : 259666 (tot), 31222/176 (reg), 14641/422 (blk)
##  sloth   -zebrafin: 246769 (tot), 21890/152 (reg), 11949/545 (blk)
##  sloth   -budgie  : 250818 (tot), 26436/159 (reg), 12014/555 (blk)
##  mouse   -chicken : 244391 (tot), 23394/221 (reg), 12267/582 (blk)
##  sandLiza-horsesho: 256812 (tot), 30313/193 (reg), 13512/615 (blk)
```

```
##   tasmania-tasmania: 273218 (tot), 69306/24 (reg), 19958/24 (blk)
##   Chunk 9 / 26 (19:12:46) ... Done!
##   sandLiza-human   : 258829 (tot), 31451/177 (reg), 13369/615 (blk)
##   opossum -echidna : 275162 (tot), 38186/242 (reg), 13448/551 (blk)
##   opossum -garterSn: 274296 (tot), 33625/177 (reg), 11908/584 (blk)
##   sloth   -swan    : 249510 (tot), 27783/148 (reg), 12660/508 (blk)
##   tasmania-horsesho: 257635 (tot), 43857/170 (reg), 15792/341 (blk)
##   sandLiza-dolphin : 248788 (tot), 25493/175 (reg), 13230/611 (blk)
##   Chunk 10 / 26 (19:14:27) ... Done!
##   tasmania-human   : 261475 (tot), 44005/150 (reg), 15881/363 (blk)
##   horsesho-horsesho: 244509 (tot), 43591/62 (reg), 19649/62 (blk)
##   sandLiza-echidna : 271307 (tot), 31082/195 (reg), 13234/627 (blk)
##   sandLiza-garterSn: 268996 (tot), 61419/50 (reg), 15277/254 (blk)
##   tasmania-dolphin : 251719 (tot), 30656/144 (reg), 15413/340 (blk)
##   horsesho-human   : 243250 (tot), 40326/99 (reg), 17075/188 (blk)
##   Chunk 11 / 26 (19:16:49) ... Done!
##   human   -human   : 253468 (tot), 49936/82 (reg), 19520/82 (blk)
##   opossum -platypus: 260803 (tot), 38041/245 (reg), 13903/534 (blk)
##   brushtai-zebrafin: 240680 (tot), 23035/160 (reg), 12178/524 (blk)
##   brushtai-budgie  : 244517 (tot), 26985/162 (reg), 12191/552 (blk)
##   tasmania-echidna : 269098 (tot), 39819/237 (reg), 14361/500 (blk)
##   horsesho-dolphin : 234653 (tot), 32494/79 (reg), 16591/147 (blk)
##   Chunk 12 / 26 (19:18:24) ... Done!
##   tasmania-garterSn: 266236 (tot), 33659/172 (reg), 12640/563 (blk)
##   human   -dolphin : 235002 (tot), 34648/66 (reg), 16507/176 (blk)
##   opossum -chicken : 246792 (tot), 24415/146 (reg), 11916/518 (blk)
##   mouse   -zebrafin: 228903 (tot), 21940/229 (reg), 11891/593 (blk)
##   brushtai-swan    : 243304 (tot), 28971/154 (reg), 12884/482 (blk)
##   mouse   -budgie  : 231831 (tot), 24919/247 (reg), 11960/623 (blk)
##   Chunk 13 / 26 (19:19:34) ... Done!
##   horsesho-echidna : 249396 (tot), 34817/249 (reg), 14479/514 (blk)
##   horsesho-garterSn: 242108 (tot), 27706/177 (reg), 12568/579 (blk)
##   dolphin -dolphin : 226293 (tot), 32919/60 (reg), 19231/60 (blk)
##   sandLiza-platypus: 250620 (tot), 28748/197 (reg), 13395/603 (blk)
##   human   -echidna : 251713 (tot), 33998/230 (reg), 14381/523 (blk)
##   human   -garterSn: 243809 (tot), 28368/166 (reg), 12438/570 (blk)
##   Chunk 14 / 26 (19:20:46) ... Done!
##   mouse   -swan    : 231175 (tot), 23716/224 (reg), 12574/567 (blk)
##   sloth   -hummingb: 236324 (tot), 20790/152 (reg), 11442/509 (blk)
##   dolphin -echidna : 242219 (tot), 25889/222 (reg), 14103/521 (blk)
##   dolphin -garterSn: 232865 (tot), 23944/160 (reg), 12291/562 (blk)
##   sandLiza-chicken : 232870 (tot), 27012/82 (reg), 13136/435 (blk)
##   tasmania-platypus: 254264 (tot), 39302/233 (reg), 14703/486 (blk)
##   Chunk 15 / 26 (19:22:16) ... Done!
##   echidna -echidna : 257261 (tot), 45231/192 (reg), 18923/192 (blk)
##   echidna -garterSn: 255960 (tot), 26942/189 (reg), 12292/594 (blk)
##   horsesho-platypus: 229092 (tot), 31585/248 (reg), 14690/503 (blk)
##   garterSn-garterSn: 257444 (tot), 66994/191 (reg), 18896/191 (blk)
##   tasmania-chicken : 238838 (tot), 24511/146 (reg), 12520/508 (blk)
##   human   -platypus: 231400 (tot), 30761/233 (reg), 14697/522 (blk)
##   Chunk 16 / 26 (19:23:09) ... Done!
##   horsesho-chicken : 216409 (tot), 22094/163 (reg), 12506/532 (blk)
##   opossum -zebrafin: 225710 (tot), 20774/154 (reg), 11457/530 (blk)
##   dolphin -platypus: 219433 (tot), 24196/223 (reg), 14272/512 (blk)
```

```
## opossum -budgie  : 229310 (tot), 25124/164 (reg), 11490/559 (blk)
## human   -chicken : 219915 (tot), 23598/145 (reg), 12374/531 (blk)
## brushtai-hummingb: 229536 (tot), 22199/156 (reg), 11602/498 (blk)
## Chunk 17 / 26 (19:24:04) ... Done!
## echidna -platypus: 247817 (tot), 40623/59 (reg), 16240/107 (blk)
## opossum -swan    : 228214 (tot), 25285/152 (reg), 12192/511 (blk)
## garterSn-platypus: 234376 (tot), 25078/189 (reg), 12454/574 (blk)
## dolphin -chicken : 204851 (tot), 21030/144 (reg), 12188/513 (blk)
## mouse   -hummingb: 218064 (tot), 21257/227 (reg), 11389/564 (blk)
## sandLiza-zebrafin: 223818 (tot), 23744/102 (reg), 12911/480 (blk)
## Chunk 18 / 26 (19:25:07) ... Done!
## sandLiza-budgie  : 225892 (tot), 24561/107 (reg), 12906/497 (blk)
## echidna -chicken : 228368 (tot), 24224/159 (reg), 12081/505 (blk)
## garterSn-chicken : 217801 (tot), 23633/82 (reg), 12353/435 (blk)
## tasmania-zebrafin: 222003 (tot), 21989/157 (reg), 12173/530 (blk)
## sandLiza-swan    : 225061 (tot), 28015/90 (reg), 13449/428 (blk)
## tasmania-budgie  : 225871 (tot), 23863/152 (reg), 12200/557 (blk)
## Chunk 19 / 26 (19:25:59) ... Done!
## platypus-platypus: 229124 (tot), 41262/84 (reg), 18140/84 (blk)
## horsesho-zebrafin: 204301 (tot), 20981/167 (reg), 12152/544 (blk)
## tasmania-swan    : 225020 (tot), 24551/151 (reg), 12749/478 (blk)
## horsesho-budgie  : 208637 (tot), 22693/184 (reg), 12192/570 (blk)
## human   -zebrafin: 203964 (tot), 22016/147 (reg), 12092/540 (blk)
## human   -budgie  : 208534 (tot), 23199/167 (reg), 12202/569 (blk)
## Chunk 20 / 26 (19:26:48) ... Done!
## platypus-chicken : 205680 (tot), 22376/162 (reg), 12276/507 (blk)
## horsesho-swan    : 206526 (tot), 23431/164 (reg), 12746/507 (blk)
## dolphin -zebrafin: 192752 (tot), 20147/148 (reg), 12057/534 (blk)
## dolphin -budgie  : 197088 (tot), 20713/159 (reg), 12148/552 (blk)
## human   -swan    : 206942 (tot), 23559/152 (reg), 12572/503 (blk)
## opossum -hummingb: 214555 (tot), 21086/148 (reg), 10963/522 (blk)
## Chunk 21 / 26 (19:27:38) ... Done!
## chicken -chicken : 231133 (tot), 57225/140 (reg), 17795/140 (blk)
## echidna -zebrafin: 220487 (tot), 20270/171 (reg), 11880/534 (blk)
## garterSn-zebrafin: 208606 (tot), 21582/100 (reg), 12121/460 (blk)
## dolphin -swan    : 195335 (tot), 20930/146 (reg), 12393/503 (blk)
## echidna -budgie  : 224205 (tot), 23325/178 (reg), 12050/569 (blk)
## garterSn-budgie  : 209920 (tot), 23270/100 (reg), 12268/499 (blk)
## Chunk 22 / 26 (19:28:22) ... Done!
## echidna -swan    : 222165 (tot), 24413/169 (reg), 12289/496 (blk)
## garterSn-swan    : 209210 (tot), 25080/87 (reg), 12606/417 (blk)
## sandLiza-hummingb: 212708 (tot), 22215/97 (reg), 12362/440 (blk)
## tasmania-hummingb: 212052 (tot), 20247/151 (reg), 11586/494 (blk)
## platypus-zebrafin: 197046 (tot), 20189/175 (reg), 12099/529 (blk)
## platypus-budgie  : 201084 (tot), 22304/176 (reg), 12143/557 (blk)
## Chunk 23 / 26 (19:29:12) ... Done!
## horsesho-hummingb: 194435 (tot), 19722/167 (reg), 11622/515 (blk)
## chicken -zebrafin: 195001 (tot), 25915/63 (reg), 13682/299 (blk)
## platypus-swan    : 199300 (tot), 22219/167 (reg), 12558/495 (blk)
## chicken -budgie  : 197415 (tot), 27547/75 (reg), 13594/328 (blk)
## human   -hummingb: 194140 (tot), 21110/146 (reg), 11555/511 (blk)
## chicken -swan    : 198253 (tot), 28161/58 (reg), 14172/211 (blk)
## Chunk 24 / 26 (19:30:01) ... Done!
## dolphin -hummingb: 183076 (tot), 19228/147 (reg), 11570/509 (blk)
```

```
##   echidna -hummingb: 211125 (tot), 19799/168 (reg), 11382/508 (blk)
##   garterSn-hummingb: 197073 (tot), 21126/94 (reg), 11666/434 (blk)
##   zebrafin-zebrafin: 191901 (tot), 27135/170 (reg), 16517/170 (blk)
##   zebrafin-budgie  : 189737 (tot), 25083/96 (reg), 13859/357 (blk)
##   budgie  -budgie  : 198685 (tot), 30785/197 (reg), 16455/197 (blk)
##   Chunk 25 / 26 (19:30:35) ... Done!
##   platypus-hummingb: 187434 (tot), 19570/170 (reg), 11601/511 (blk)
##   zebrafin-swan    : 186997 (tot), 24830/71 (reg), 13729/263 (blk)
##   budgie  -swan    : 190780 (tot), 25684/84 (reg), 13808/297 (blk)
##   chicken -hummingb: 181763 (tot), 24458/62 (reg), 13120/294 (blk)
##   swan    -swan    : 189788 (tot), 29562/125 (reg), 16144/125 (blk)
##   zebrafin-hummingb: 175365 (tot), 22583/72 (reg), 13116/300 (blk)
##   Chunk 26 / 26 (19:31:13) ... Done!
##   budgie  -hummingb: 180203 (tot), 23996/89 (reg), 13085/318 (blk)
##   swan    -hummingb: 176792 (tot), 23979/71 (reg), 13378/261 (blk)
##   hummingb-hummingb: 167795 (tot), 24555/112 (reg), 14709/112 (blk)
## Defining synteny-constrained orthogroups ...
##   Found 43087 synteny-split OGs for 325509 genes
## Found 43087 OGs across 325509 genes. gff3-like text file written to:
##   /Users/jlovell/Desktop/GENESPACE_data/results//vertebrates/results/gffWithOgs.txt.gz
## Calculating syntenic block breakpoints ...
##   Found 128186 blocks. Text file written to:
##   /Users/jlovell/Desktop/GENESPACE_data/results//vertebrates/results/syntenicBlocks.txt.gz:
```

7 Build pangenomes against each genome

```r
humanPg <- pangenome(gsParam = gparVerts, refGenome = "human")
```

```
## Building reference-anchored scaffold against human
##   n. ref positions = 17443
##   Reading in hits against human ... found 243801
##   Interpolating positions ... n. genes mapped: 1x = 242648, 2+x = 5337, 0x = 43275
##   Forming ref.-anchored db ... found 230909 genes for 17426 placements
## Completing the pan-genome annotation ...
##   Adding non-anchor entries ... found 12647 genes and 4715 placements
##   Checking missing direct ref. syn. OGs ... found 8226 genes and 3546 placements
##   Adding indirect syn. OGs ... found 858 genes and 412 placements
##   Adding syn. OGs without ref. anchor ... found 27093 genes and 17124 placements
##   Adding missing genes by synOG identity ... found 11527 genes and 222 placements
## Annotating and formatting pan-genome
##   Adding non-anchor entries ... found 34249 genes and 4063 placements
##   Adding non-syn. orthologs ... found 65149 genes and 11584 placements
##   Writing pangenome to results/human_pangenomeDB.txt.gz
##   Returning wide-format with only syntenic array reps
##   Done!
```

```r
chickPg <- pangenome(gsParam = gparVerts, refGenome = "chicken")
```

```
## Building reference-anchored scaffold against chicken
##   n. ref positions = 15665
##   Reading in hits against chicken ... found 220047
##   Interpolating positions ... n. genes mapped: 1x = 226451, 2+x = 2667, 0x = 62142
##   Forming ref.-anchored db ... found 207240 genes for 15650 placements
```

```
## Completing the pan-genome annotation ...
##   Adding non-anchor entries ... found 12635 genes and 3993 placements
##   Checking missing direct ref. syn. OGs ... found 8264 genes and 2835 placements
##   Adding indirect syn. OGs ... found 834 genes and 302 placements
##   Adding syn. OGs without ref. anchor ... found 37748 genes and 16200 placements
##   Adding missing genes by synOG identity ... found 24539 genes and 564 placements
## Annotating and formatting pan-genome
##   Adding non-anchor entries ... found 34249 genes and 3572 placements
##   Adding non-syn. orthologs ... found 62399 genes and 10760 placements
##   Writing pangenome to results/chicken_pangenomeDB.txt.gz
##   Returning wide-format with only syntenic array reps
##   Done!
```
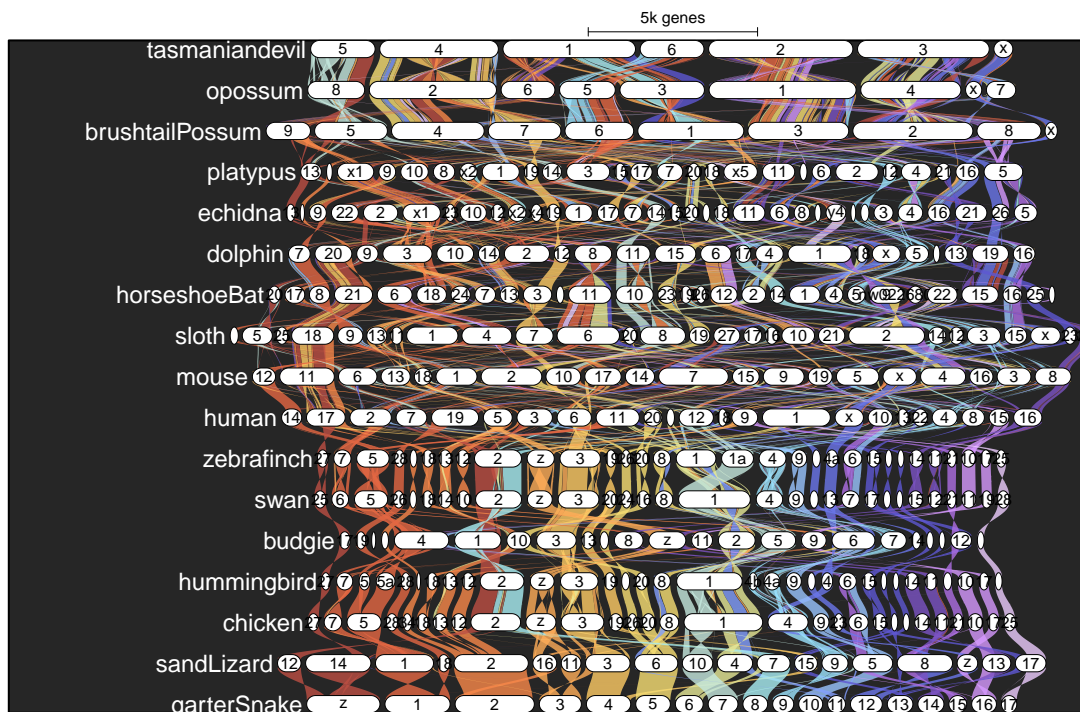
8 Make riparian plot

Also, write the parameters to file

```
plot_riparian(gsParam = gparVerts)
```

```
##   Loading the gff ... Done!
##   Mapping genes against garterSnake chromosomes ... Done!
##   Projecting linear coordinate system ... Done!
##   Generating block coordinates ... Done!
##   Rendering plot ...
```


Chromosomes scaled by gene rank order

```
## Done!
```

```
save(gparVerts, file = file.path(baseDir, "gparVerts.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
```