Speculomics PBS

Population Branch Stats

I'm interested in using population branch statistics (PBS) to measure differentiation (and possible selection) between multiple ecotypes. Particularly, we seem to be observing forest:nonforest differentiation at the CYP9K locus on chrX. I calculated the 2dsfs between all ecotypes (rainforest, deciduous forest, mangrove and savannah) and then the PBS between then using realSFS in ANGSD.

Set env, load externally developed packages, create filelist and names

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: viridis
```

```
## Loading required package: viridisLite
```

```
## Loading required package: data.table
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
## Loading required package: cowplot
```

```
## Loading required package: RColorBrewer
```

```
## [[1]]
## [1] TRUE
##
## [[2]]
## [1] TRUE
##
## [[3]]
## [1] TRUE
##
## [[4]]
## [1] TRUE
##
## [[5]]
## [1] TRUE
```

Def function for plotting pbs

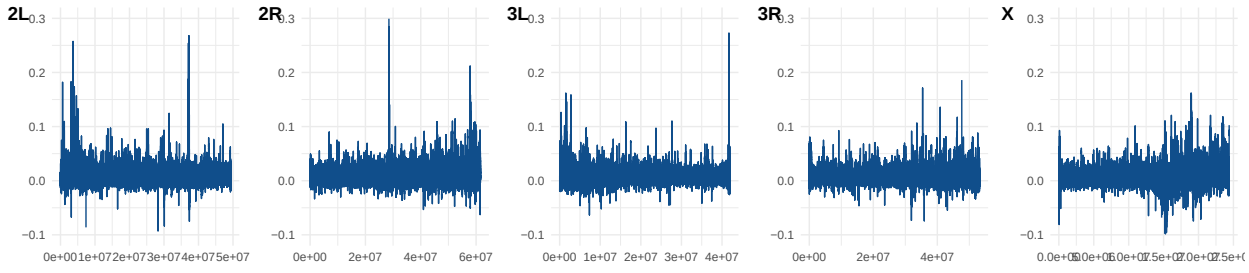
```
list = list.files(pattern = '*.windowedpbs')
names = gsub('.windowedpbs', '', list)
names = gsub('decid_forest', 'decid-forest', names)
#func for plotting pbs
plot_pbs <- function(pbsfile) {
  pbsdf = fread(pbsfile)
  pnam = gsub('.1kwin200step-windowedpbs', '', pbsfile)
  ggplot(pbsdf, aes(x=midPos,y=PBS1))+
  theme_minimal()+
  ylim(-0.1, 0.3)+
  geom_line(color = '#104e8b')+
  labs(x='', y='')
}

plots = lapply(list, plot_pbs)
```

Then, plot all of our PBS comparisons (I excluded one because decid_forest and rainforest are essentially the same)

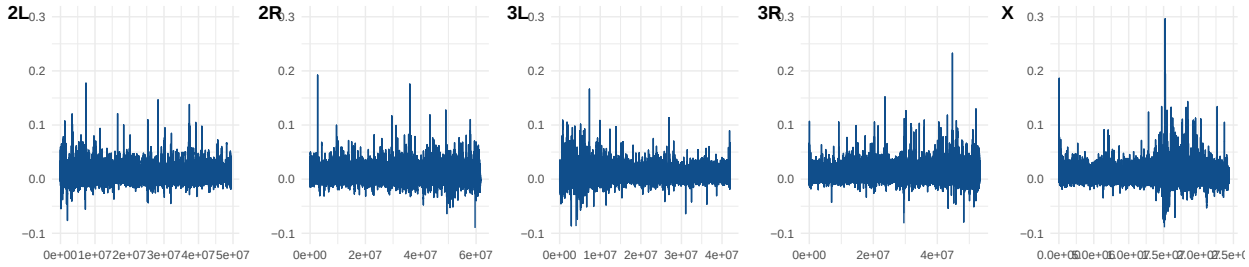
M, decid-forest, savannah, mangrove

```
labels = c('2L', '2R', '3L', '3R', 'X')
plot_grid(plotlist = plots[1:5], ncol = 5, nrow = 1, labels = labels)
```



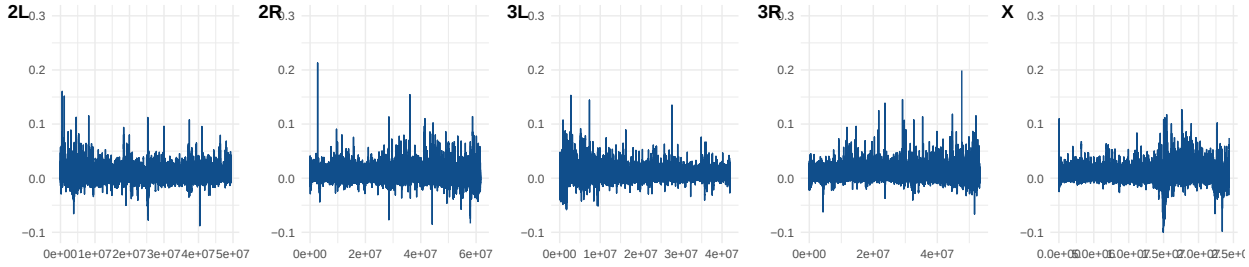
m_rainforest_decid-forest_mangrove

```
labels = c('2L', '2R', '3L', '3R', 'X')
plot_grid(plotlist = plots[6:10], ncol = 5, nrow = 1, labels = labels)
```



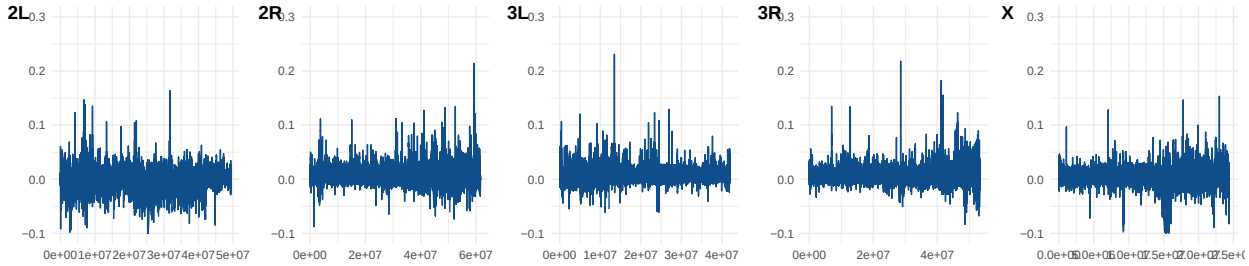
m_rainforest_decid-forest_savannah

```
labels = c('2L', '2R', '3L', '3R', 'X')
plot_grid(plotlist = plots[11:15], ncol = 5, nrow = 1, labels = labels)
```



s_rainforest_decid-forest_savannah

```
labels = c('2L', '2R', '3L', '3R', 'X')
plot_grid(plotlist = plots[21:25], ncol = 5, nrow = 1,  labels = labels)
```



There's some peaks here. I've extracted the gene annotation at the highest points of each peak > 0.3.

```
#func for extracting pbs 'peaks' of > 0.15, assign condition name and assign 'peak id' to each one chos
extract_pbs <- function(pbsfile, threshold) {
  pbsdf = fread(pbsfile)
  pnam = gsub('.1kwin200step-windowedpbs', '', pbsfile)
  filtered_site = pbsdf[pbsdf$PBS2 > 0.15, ]
  fs1 = mutate(filtered_site, name = pnam)
  fs2 = mutate(fs1, peak = paste0(min(midPos),'_', max(midPos)))
}
```

```
#get filtered windows, concat into bigdf
windows = lapply(list, extract_pbs)
```

```
## Warning in min(midPos): no non-missing arguments to min; returning Inf
```

```
## Warning in max(midPos): no non-missing arguments to max; returning -Inf
```

```
windows_comb = do.call(rbind, windows)
#create bed start/end coords from midpoint
windows_comb$chromStart = windows_comb$midPos - 500
windows_comb$chromEnd = windows_comb$midPos + 500
#write df of bed file
bed_prec = as.data.frame(windows_comb)[c("chr", "chromStart", "chromEnd", "name", "PBS2", "peak")]
write_delim(bed_prec, file = "/Users/tristanpwdennis/Projects/MOVE/td_je_angam_2022/annotations/files/pb

#go to dir in question
setwd('/Users/tristanpwdennis/Projects/MOVE/td_je_angam_2022/annotations/files/')
#read the output of bedtools intersect to get the annotations overlapping the peaks
intersect_ann = fread(text = system('bedtools intersect -a pbs2.bed -b VectorBase-55_AgambiaePEST.gff -l

#filter to get protein coding gene annotations only
pcgs = intersect_ann[V9 == 'protein_coding_gene']
#aggregate by peak, and by condition, then select the highest PBS value
highest_genes = setDT(pcgs)[, .SD[which.max(V5)], by=c('V4', 'V6')]
highest_genes = data.frame(highest_genes)[c('V4','V6', 'V1','V2','V3','V10', 'V11', 'V15')]
highest_genes
```

```
##                                           V4                V6        V1
## 1    m_decid_forest_savannah_mangrove_AgamP4_2L   606300_47080900 AgamP4_2L
## 2    m_decid_forest_savannah_mangrove_AgamP4_2R 10444500_48884300 AgamP4_2R
## 3    m_decid_forest_savannah_mangrove_AgamP4_3L  9709700_41710500 AgamP4_3L
## 4     m_decid_forest_savannah_mangrove_AgamP4_X   195300_19210100  AgamP4_X
## 5  m_rainforest_decid_forest_mangrove_AgamP4_2L  3061300_40862300 AgamP4_2L
## 6  m_rainforest_decid_forest_mangrove_AgamP4_3R 15530500_51394500 AgamP4_3R
## 7   m_rainforest_decid_forest_mangrove_AgamP4_X 12838100_21163700  AgamP4_X
## 8  m_rainforest_decid_forest_savannah_AgamP4_2L  3651300_37240100 AgamP4_2L
## 9  m_rainforest_decid_forest_savannah_AgamP4_2R 28470300_58842700 AgamP4_2R
## 10  m_rainforest_decid_forest_savannah_AgamP4_X 14877100_23821100  AgamP4_X
## 11     m_rainforest_savannah_mangrove_AgamP4_2L   606300_47080900 AgamP4_2L
## 12     m_rainforest_savannah_mangrove_AgamP4_2R 28703900_59624300 AgamP4_2R
## 13     m_rainforest_savannah_mangrove_AgamP4_3L  2786100_41710500 AgamP4_3L
## 14      m_rainforest_savannah_mangrove_AgamP4_X 15113900_21164100  AgamP4_X
## 15 s_rainforest_decid_forest_savannah_AgamP4_2L   219100_44839900 AgamP4_2L
## 16 s_rainforest_decid_forest_savannah_AgamP4_2R 15038900_59211500 AgamP4_2R
## 17 s_rainforest_decid_forest_savannah_AgamP4_3L 12165100_12165300 AgamP4_3L
## 18 s_rainforest_decid_forest_savannah_AgamP4_3R  6979100_53176100 AgamP4_3R
## 19  s_rainforest_decid_forest_savannah_AgamP4_X  1076300_16703300  AgamP4_X
##          V2       V3      V10      V11
## 1   3577600  3578600  3577207  3578328
## 2  10444000 10445000 10333268 10458861
## 3  41710000 41711000 41710048 41711592
```

```
## 4  18452400 18453400 18426678 18467864
## 5  30097400 30098400 30093121 30106185
## 6  32033800 32034800 32033307 32034744
## 7  15240600 15241600 15240572 15242864
## 8  25461400 25462400 25448585 25461509
## 9  28490600 28491600 28491415 28493141
## 10 15239600 15240600 15240572 15242864
## 11  3060800  3061800  3057997  3061949
## 12 48883600 48884600 48854615 48947869
## 13 41710000 41711000 41710048 41711592
## 14 15188200 15189200 15188580 15189244
## 15 15450400 15451400 15429893 15458971
## 16 59210800 59211800 59201123 59235934
## 17 12164800 12165800 12157568 12189797
## 18 52326000 52327000 52301033 52349893
## 19 15333000 15334000 15327411 15333680
##                                                                                  V15
## 1                                          ID=AGAP004791;description=high mobility group 20A
## 2                                              ID=AGAP001786;description=unspecified product
## 3                           ID=AGAP012394;description=peptide-methionine (S)-S-oxide reductase
## 4     ID=AGAP000962;Name=alpha7;description=nicotinic acetylcholine receptor subunit alpha 7
## 5                                              ID=AGAP006345;description=unspecified product
## 6                                       ID=AGAP009392;Name=Or46;description=odorant receptor 46
## 7                                         ID=AGAP000818;Name=CYP9K1;description=cytochrome P450
## 8  ID=AGAP006030;Name=mfrn;description=Mitoferrin [Source:UniProtKB/TrEMBL%3BAcc:A0A1S4GRT4]
## 9                                         ID=AGAP002865;Name=CYP6P3;description=cytochrome P450
## 10                                        ID=AGAP000818;Name=CYP9K1;description=cytochrome P450
## 11                                             ID=AGAP004753;description=unspecified product
## 12                                             ID=AGAP004052;description=unspecified product
## 13                          ID=AGAP012394;description=peptide-methionine (S)-S-oxide reductase
## 14                                             ID=AGAP029884;description=unspecified product
## 15                                             ID=AGAP005433;description=unspecified product
## 16                                      ID=AGAP004646;description=homeobox protein HoxA/B/C/D4
## 17                                             ID=AGAP010867;description=unspecified product
## 18                            ID=AGAP010292;description=guanine nucleotide exchange factor VAV
## 19                                                  ID=AGAP000823;description=CD81 antigen
```

```
colnames(highest_genes) = c('comparison', 'peak_start_end', 'chrom', 'window_start', 'window_end', 'gene
#wo
write_csv(highest_genes, file='/Users/tristanpwdennis/Projects/MOVE/td_je_angam_2022/annotations/files/
highest_genes
```

```
##                                      comparison    peak_start_end     chrom
## 1    m_decid_forest_savannah_mangrove_AgamP4_2L   606300_47080900 AgamP4_2L
## 2    m_decid_forest_savannah_mangrove_AgamP4_2R 10444500_48884300 AgamP4_2R
## 3    m_decid_forest_savannah_mangrove_AgamP4_3L  9709700_41710500 AgamP4_3L
## 4     m_decid_forest_savannah_mangrove_AgamP4_X   195300_19210100  AgamP4_X
## 5  m_rainforest_decid_forest_mangrove_AgamP4_2L  3061300_40862300 AgamP4_2L
## 6  m_rainforest_decid_forest_mangrove_AgamP4_3R 15530500_51394500 AgamP4_3R
## 7   m_rainforest_decid_forest_mangrove_AgamP4_X 12838100_21163700  AgamP4_X
## 8  m_rainforest_decid_forest_savannah_AgamP4_2L  3651300_37240100 AgamP4_2L
## 9  m_rainforest_decid_forest_savannah_AgamP4_2R 28470300_58842700 AgamP4_2R
## 10  m_rainforest_decid_forest_savannah_AgamP4_X 14877100_23821100  AgamP4_X
## 11     m_rainforest_savannah_mangrove_AgamP4_2L   606300_47080900 AgamP4_2L
```

```
## 12     m_rainforest_savannah_mangrove_AgamP4_2R 28703900_59624300 AgamP4_2R
## 13     m_rainforest_savannah_mangrove_AgamP4_3L  2786100_41710500 AgamP4_3L
## 14      m_rainforest_savannah_mangrove_AgamP4_X 15113900_21164100  AgamP4_X
## 15 s_rainforest_decid_forest_savannah_AgamP4_2L   219100_44839900 AgamP4_2L
## 16 s_rainforest_decid_forest_savannah_AgamP4_2R 15038900_59211500 AgamP4_2R
## 17 s_rainforest_decid_forest_savannah_AgamP4_3L 12165100_12165300 AgamP4_3L
## 18 s_rainforest_decid_forest_savannah_AgamP4_3R  6979100_53176100 AgamP4_3R
## 19  s_rainforest_decid_forest_savannah_AgamP4_X  1076300_16703300  AgamP4_X
##    window_start window_end gene_start gene_end
## 1       3577600    3578600    3577207  3578328
## 2      10444000   10445000   10333268 10458861
## 3      41710000   41711000   41710048 41711592
## 4      18452400   18453400   18426678 18467864
## 5      30097400   30098400   30093121 30106185
## 6      32033800   32034800   32033307 32034744
## 7      15240600   15241600   15240572 15242864
## 8      25461400   25462400   25448585 25461509
## 9      28490600   28491600   28491415 28493141
## 10     15239600   15240600   15240572 15242864
## 11      3060800    3061800    3057997  3061949
## 12     48883600   48884600   48854615 48947869
## 13     41710000   41711000   41710048 41711592
## 14     15188200   15189200   15188580 15189244
## 15     15450400   15451400   15429893 15458971
## 16     59210800   59211800   59201123 59235934
## 17     12164800   12165800   12157568 12189797
## 18     52326000   52327000   52301033 52349893
## 19     15333000   15334000   15327411 15333680
##                                                                                annotation
## 1                                          ID=AGAP004791;description=high mobility group 20A
## 2                                              ID=AGAP001786;description=unspecified product
## 3                           ID=AGAP012394;description=peptide-methionine (S)-S-oxide reductase
## 4      ID=AGAP000962;Name=alpha7;description=nicotinic acetylcholine receptor subunit alpha 7
## 5                                              ID=AGAP006345;description=unspecified product
## 6                                       ID=AGAP009392;Name=Or46;description=odorant receptor 46
## 7                                        ID=AGAP000818;Name=CYP9K1;description=cytochrome P450
## 8  ID=AGAP006030;Name=mfrn;description=Mitoferrin [Source:UniProtKB/TrEMBL%3BAcc:A0A1S4GRT4]
## 9                                        ID=AGAP002865;Name=CYP6P3;description=cytochrome P450
## 10                                       ID=AGAP000818;Name=CYP9K1;description=cytochrome P450
## 11                                             ID=AGAP004753;description=unspecified product
## 12                                             ID=AGAP004052;description=unspecified product
## 13                          ID=AGAP012394;description=peptide-methionine (S)-S-oxide reductase
## 14                                             ID=AGAP029884;description=unspecified product
## 15                                             ID=AGAP005433;description=unspecified product
## 16                                      ID=AGAP004646;description=homeobox protein HoxA/B/C/D4
## 17                                             ID=AGAP010867;description=unspecified product
## 18                            ID=AGAP010292;description=guanine nucleotide exchange factor VAV
## 19                                                 ID=AGAP000823;description=CD81 antigen
```