

Allelic state plot for sex-linked SNPs. First the heterozygosity for sex-linked SNPs was extracted from each vcf files using vcfutils in 012 format.0 means heterozygous allele, 2 means homozygous for alternate allele and 2 means missing."

Set the working directory and load the data

```
r setwd("C:/Users/khana/OneDrive - Texas Tech University/Documents/Texas_Tech/research/first_chapter/sphlebophylla/gwas/reanalysis") heterozygosity
```

Load the required packages

```
suppressMessages(library(cowplot))
library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.4    ✓ readr      2.1.5
## ✓ forcats    1.0.0    ✓ stringr   1.5.1
## ✓ ggplot2     3.5.1    ✓ tibble    3.2.1
## ✓ lubridate  1.9.4    ✓ tidyr     1.3.1
## ✓ purrr      1.0.4
## — Conflicts ————— tidyverse_conflicts() —
## X dplyr::filter() masks stats::filter()
## X dplyr::lag()     masks stats::lag()
## X lubridate::stamp() masks cowplot::stamp()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##      smiths
```

Convert the data into a “tidy” format using the reshape2 library

```
mydata_tidy <- melt(heterozygosity_file, id.vars="Ind", variable.name="SNP", value.name="AllelicState")
head(mydata_tidy)
```

```
##      Ind      SNP AllelicState
## 1 Male1 X2908969          0
## 2 Male2 X2908969          0
## 3 Male3 X2908969          0
## 4 Male4 X2908969          0
## 5 Male5 X2908969          0
## 6 Male6 X2908969          0
```

```
tail(mydata_tidy)
```

```
##      Ind      SNP AllelicState
## 2663 Female20 X8037834          2
## 2664 Female21 X8037834          2
## 2665 Female22 X8037834          1
## 2666 Female23 X8037834          0
## 2667 Female24 X8037834          1
## 2668 Female25 X8037834          1
```

```
view(mydata_tidy)
mydata_tidy$AllelicState <- factor(mydata_tidy$AllelicState,
                                   levels = c("0", "1", "2", "-1", "3"))
```

Add a column to identify Male and Female groups

```
mydata_tidy$Group <- ifelse(grepl("Male", mydata_tidy$Ind), "Male", "Female")
```

Arrange the data so that males come before females, and order within each group

```
mydata_tidy <- mydata_tidy %>%
  arrange(Group, Ind)
```

Set the levels of Ind to maintain the new ordering

```
mydata_tidy$Ind <- factor(mydata_tidy$Ind, levels = unique(mydata_tidy$Ind))
```

```
ggplot(mydata_tidy, aes(x = as.factor(SNP), y = Ind, fill = AllelicState)) +  
  geom_tile(color = NA) + # Remove the tile borders  
  scale_fill_manual(values = c("skyblue", "darkgreen", "skyblue", "white", "peachpuff"), na.value = "white") +  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +  
  labs(x = "SNP", y = "Individual", fill = "Allelic State")
```

