

# Popgen

## Loading libraries

## Loading previous results

```
setwd("~/GitHub/devnetwork/")
load("results/DEtests.RData")
load("results/collectedPhylo.RData")
tau <- read.csv("results/bee_tau.csv")
sexGenes <- read.csv("results/dmel_sexGenes.csv")
antConn <- read.csv("results/antConnectivity.csv")
beeConn <- read.csv("results/beeConnectivity.csv")
beePi <- read.csv("results/apis.gene.pi.csv")
beeSub <- read.csv("results/substitutions.csv")
antEvol <- read.csv("data/MpharAnn.csv")
antConstraint <- read.csv("data/Mphar_constraint.csv")
beeSnipre <- read.csv("results/bayesian_results_apis.csv")
alphaResults <- read.csv("results/collectedAlpha.csv")
colnames(antConstraint) = c("Gene", "f")
antSub = antEvol[!is.na(antEvol$Fixed.Non.Synonymous), c(1, 13, 12, 15, 14, 17, 16)]
colnames(antSub) = c("Gene", "FN", "FS", "PN", "PS", "Trepl", "Tsil")
antSub = merge(antSub, antConstraint, by="Gene")
beeConstraint = read.table("results/MKtest_out")
beeSub = cbind(beeSub, f=as.numeric(as.character((t(beeConstraint[2, 4:(ncol(beeConstraint) - 1)]))))))
antGamma = antEvol[!is.na(antEvol$BSnIPRE.gamma), c(1, 20)]
beeGamma = beeSnipre[!is.na(beeSnipre$BSnIPRE.gamma), c("gene", "BSnIPRE.gamma")]
colnames(beeGamma)[1] = "Gene"
antSub = merge(antSub, antGamma, by="Gene")
beeAnn = read.csv("results/annotation.csv", header=F)
beeAnn = beeAnn[!duplicated(beeAnn$V5), ]
beeGamma = merge(beeGamma, beeAnn, by.x="Gene", by.y="V5")
beeSub = merge(beeSub, beeGamma[, c(2, 6)], by.x = "Gene", by.y="V4")

beeT <- read.table("data/bees.tpm.txt", header=TRUE)
antT <- read.table("data/ants.tpm.txt", header=TRUE)
modifyDF <- function(data){
  rownames(data)=data[, 1]
  return(data[!grepl("ERCC", rownames(data)), -c(1)])
}
beeT <- modifyDF(beeT)
antT <- modifyDF(antT)
antT = antT[rowSums(antT) > 0, ]
beeT = beeT[rowSums(beeT) > 0, ]
```

## Collecting Data

```
beeS = merge(beeRes_allstage[[2]], beeSub, by="Gene") %>% melt(id.vars=colnames(beeSub))
antS = merge(antRes_allstage[[2]], antSub, by="Gene") %>% melt(id.vars=colnames(antSub))
beeS$species = "bee"
```

```

antS$species = "ant"
beeS$divRank = rank(beeS$FN/(beeS$FS+beeS$FN+1))
antS$divRank = rank(antS$FN/(antS$FS+antS$FN+1))

allS = rbind(beeS,antS)

```

## Part 1: Divergence of worker and queen genes

```

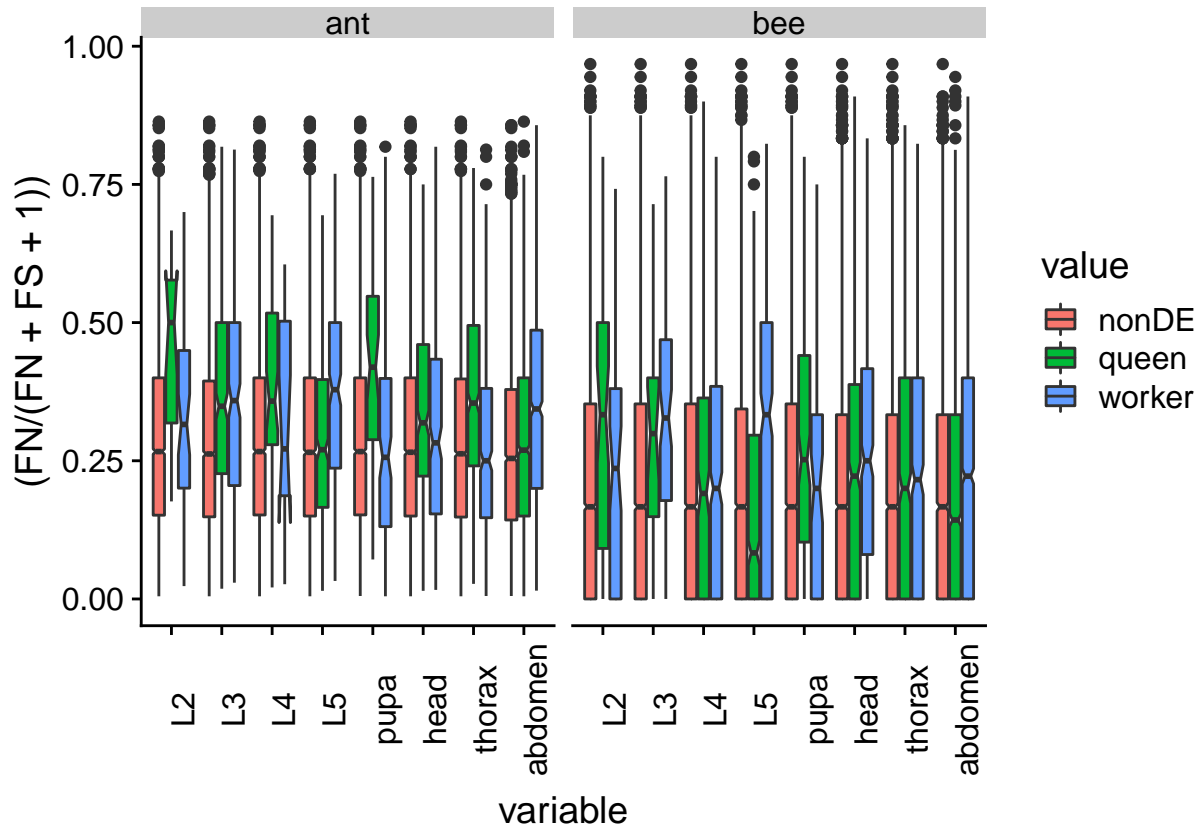
ggplot(allS,aes(x = variable,fill=value,y=(FN/(FN+FS+1))))+
  geom_boxplot(notch=T)+
  facet_wrap(. ~ species)+
  theme(axis.text.x=element_text(angle=90))

```

```

## notch went outside hinges. Try setting notch=FALSE.
## notch went outside hinges. Try setting notch=FALSE.

```



```

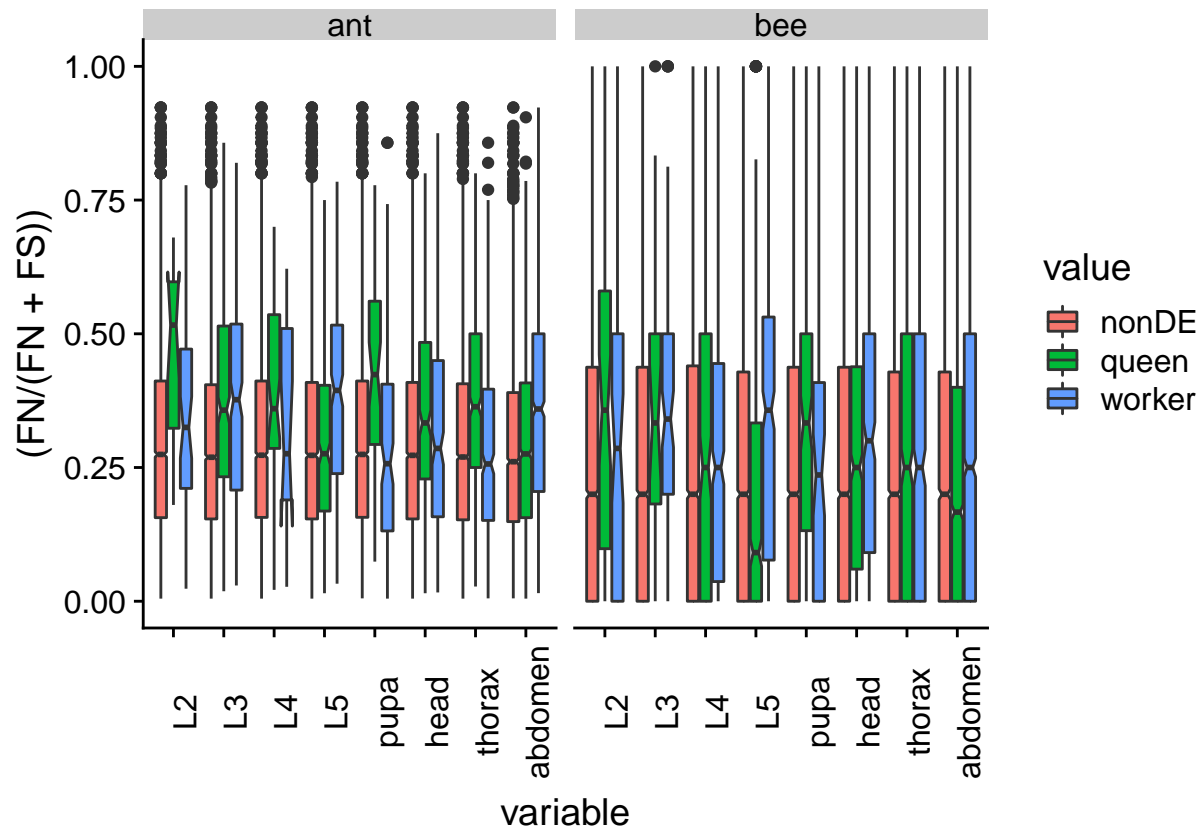
#This removes genes with zero substitutions
ggplot(allS,aes(x = variable,fill=value,y=(FN/(FN+FS))))+
  geom_boxplot(notch=T)+
  facet_wrap(. ~ species)+
  theme(axis.text.x=element_text(angle=90))

```

```

## Warning: Removed 1248 rows containing non-finite values (stat_boxplot).
## notch went outside hinges. Try setting notch=FALSE.
## notch went outside hinges. Try setting notch=FALSE.

```

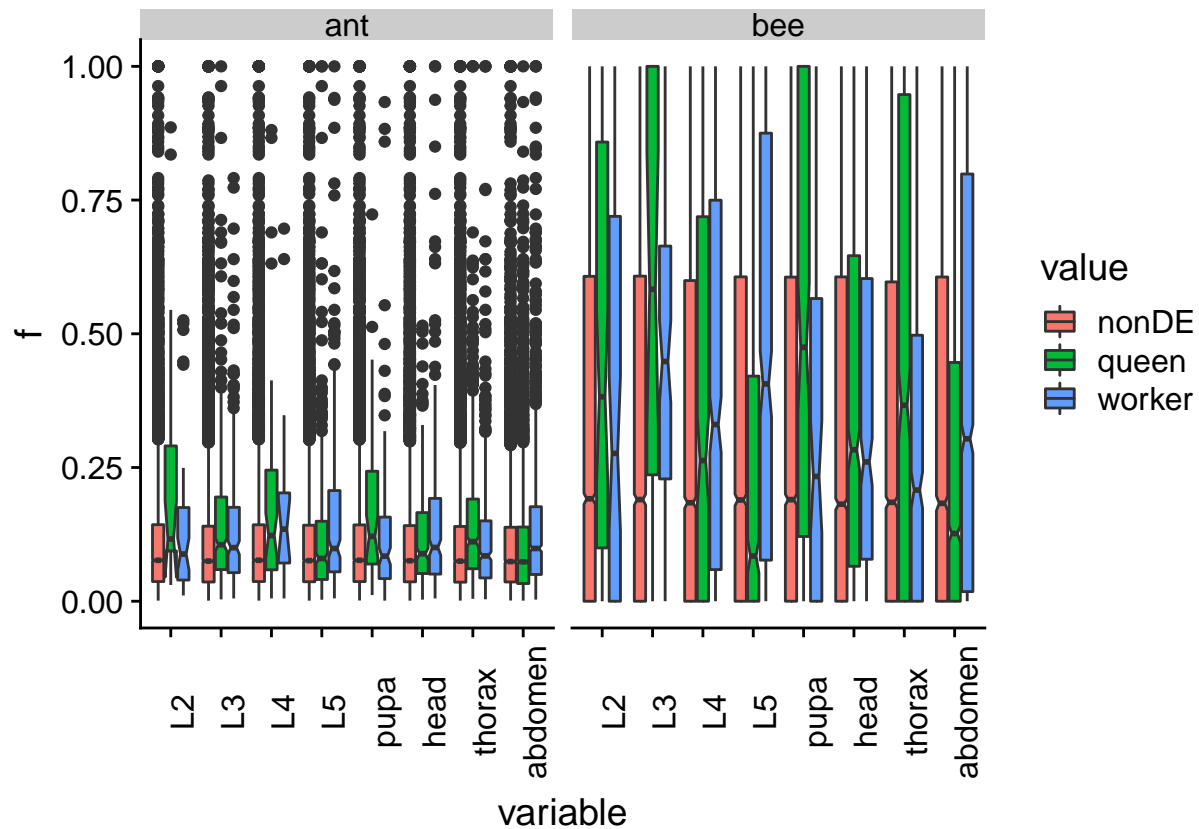


## Part 2: Selective Constraint and Pi

- note that for now only have pi for bees

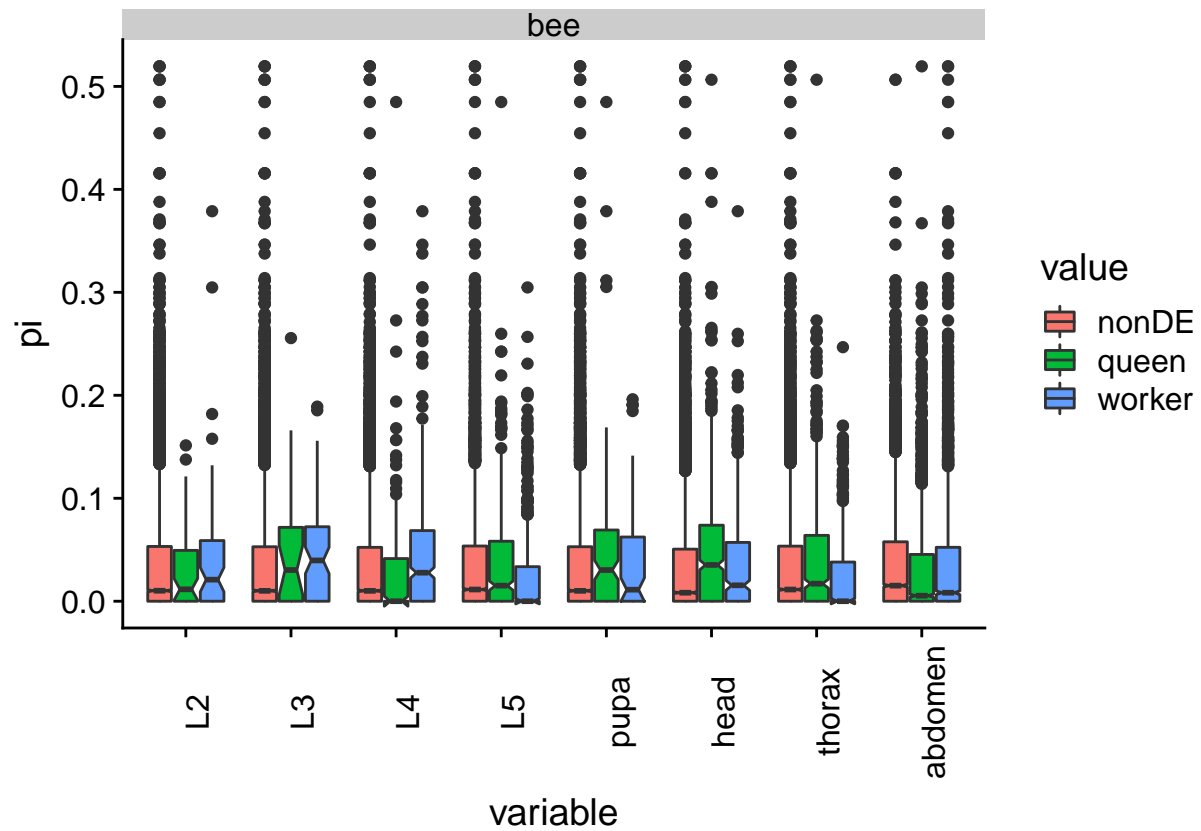
```
#constraint
ggplot(allS,aes(x = variable,fill=value,y=f))+
  geom_boxplot(notch=T)+
  facet_wrap(. ~ species)+
  theme(axis.text.x=element_text(angle=90))
```

## notch went outside hinges. Try setting notch=FALSE.



```
#pi (just bees)
Bpi = merge(beePi, beeS, by="Gene")
ggplot(Bpi, aes(x=variable, y=pi, fill=value)) +
  geom_boxplot(notch=T) +
  facet_wrap(. ~ species) +
  theme(axis.text.x=element_text(angle=90))
```

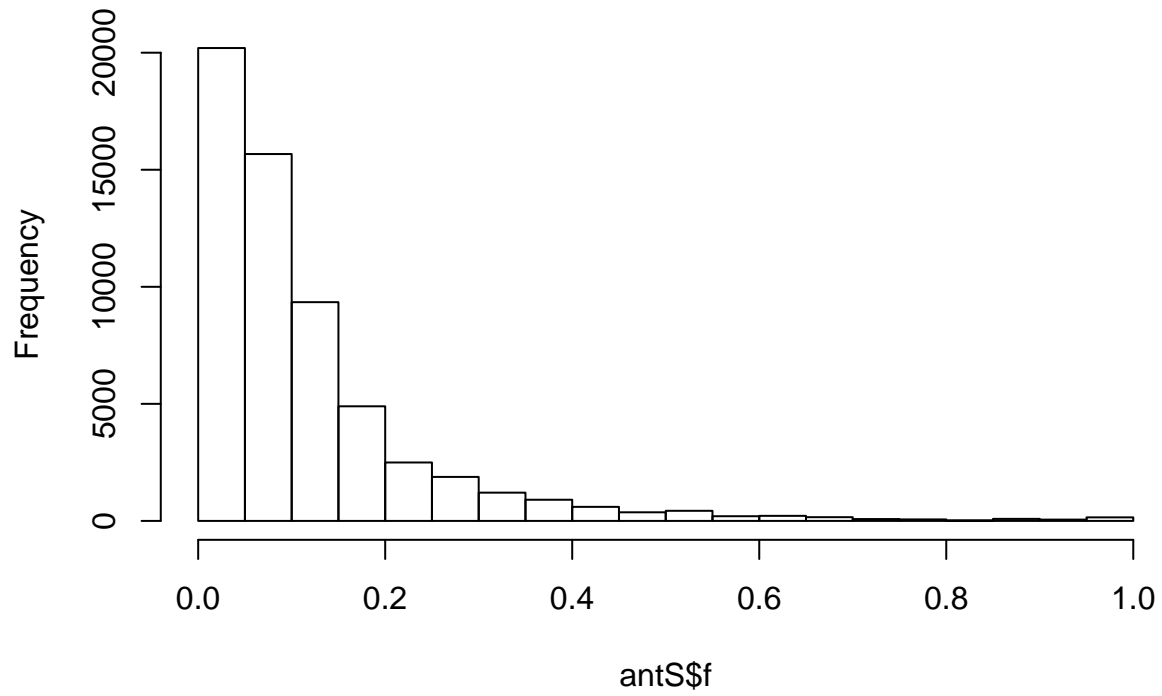
```
## notch went outside hinges. Try setting notch=FALSE.
## notch went outside hinges. Try setting notch=FALSE.
## notch went outside hinges. Try setting notch=FALSE.
## notch went outside hinges. Try setting notch=FALSE.
```



Note: the distributions of f look weird (but are consistent with more diversity in pharaonis sequences)

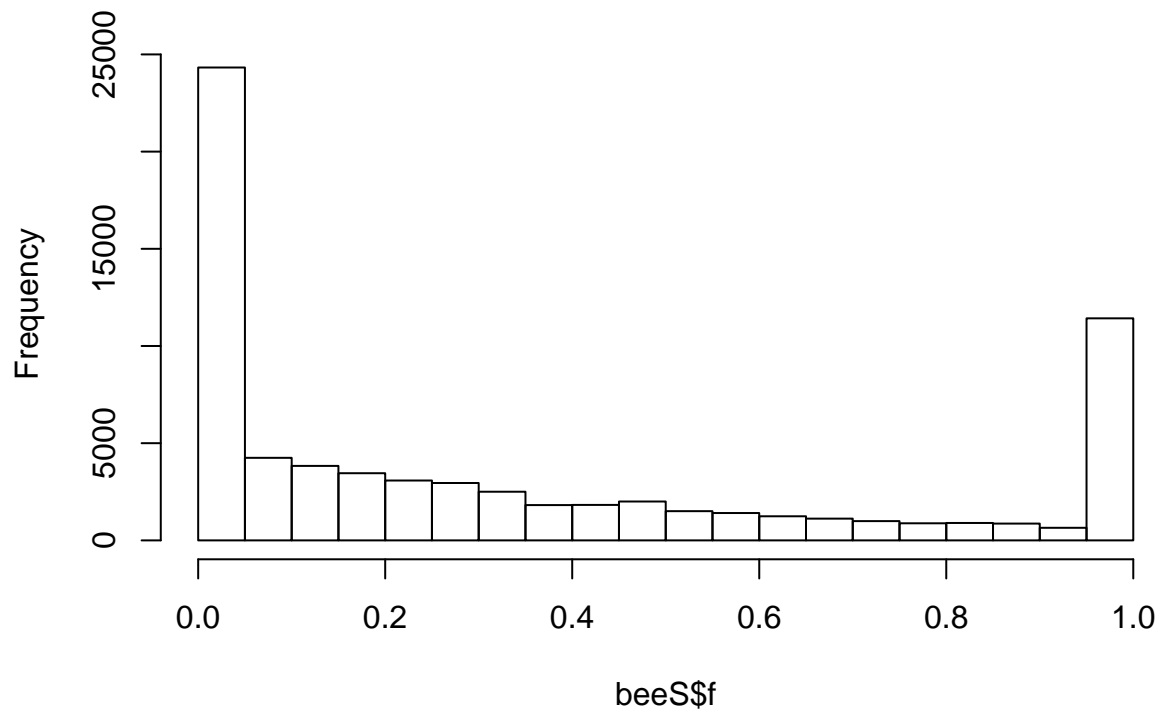
```
hist(antS$f)
```

**Histogram of antS\$f**



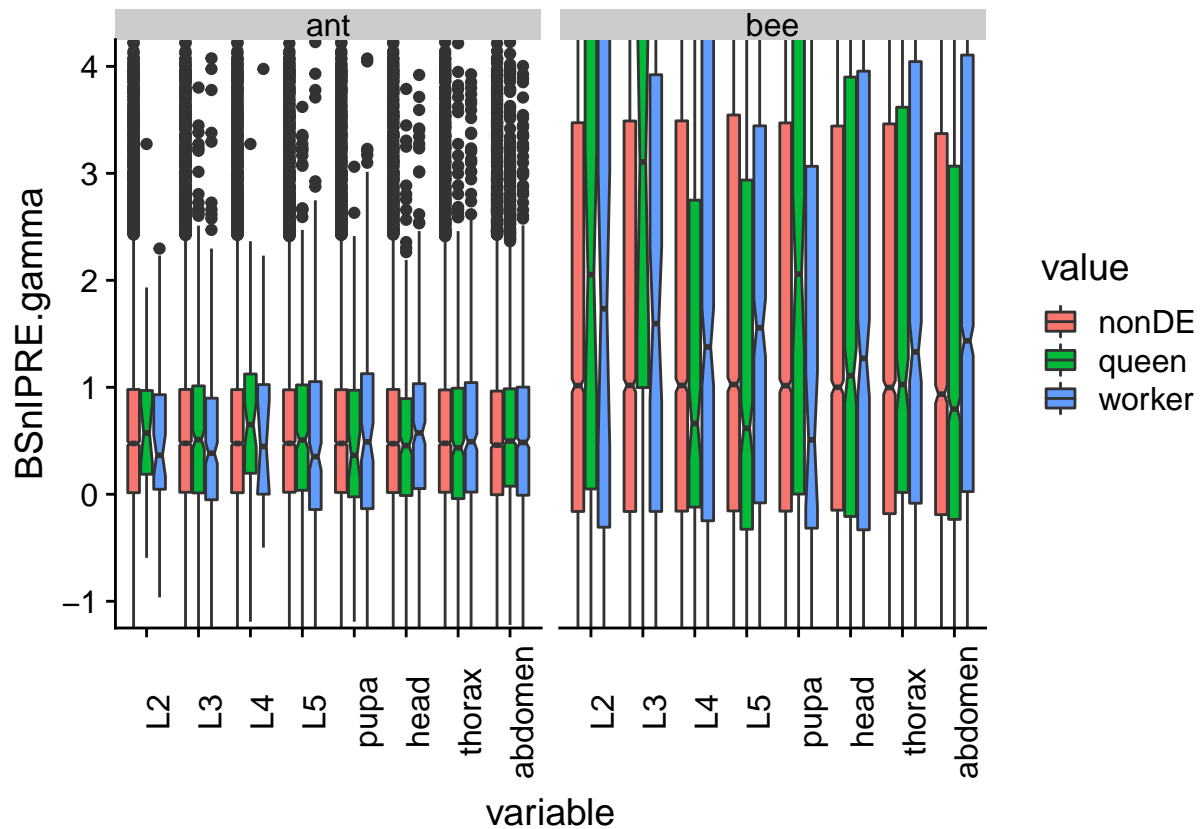
```
hist(beeS$f)
```

**Histogram of beeS\$f**



### Part 3: Positive selection, Gamma

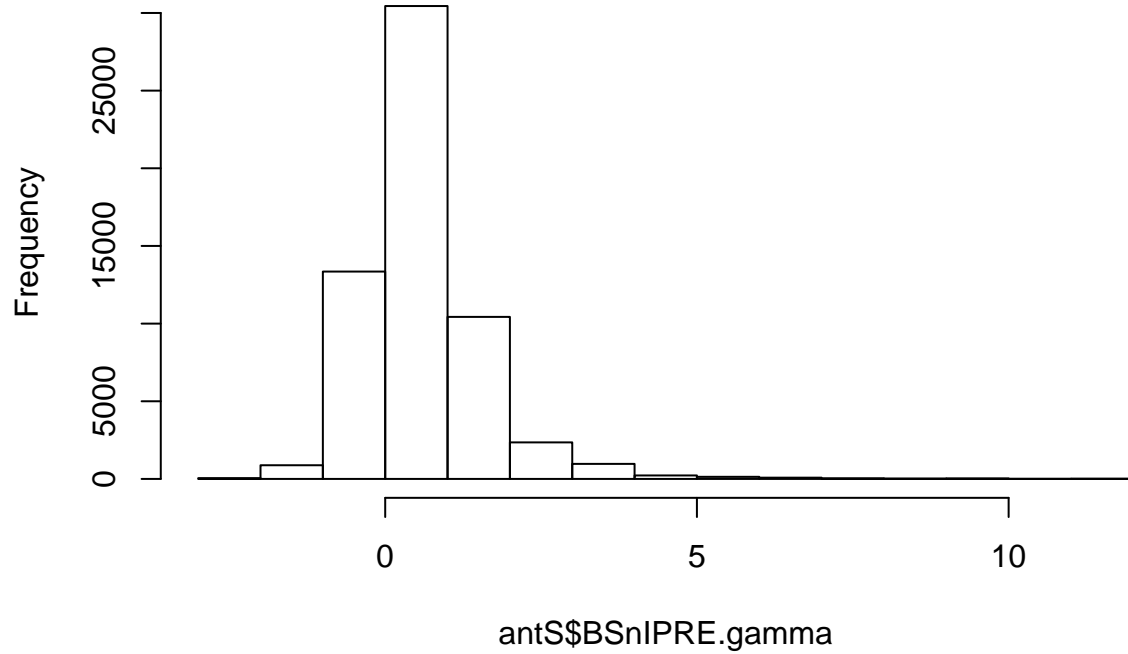
```
ggplot(allS,aes(x = variable,fill=value,y=BSnIPRE.gamma))+  
  geom_boxplot(notch=T)+  
  facet_wrap(. ~ species)+  
  coord_cartesian(ylim = c(-1,4))+  
  theme(axis.text.x=element_text(angle=90))
```



Honey bee gamma distribution also looks weird

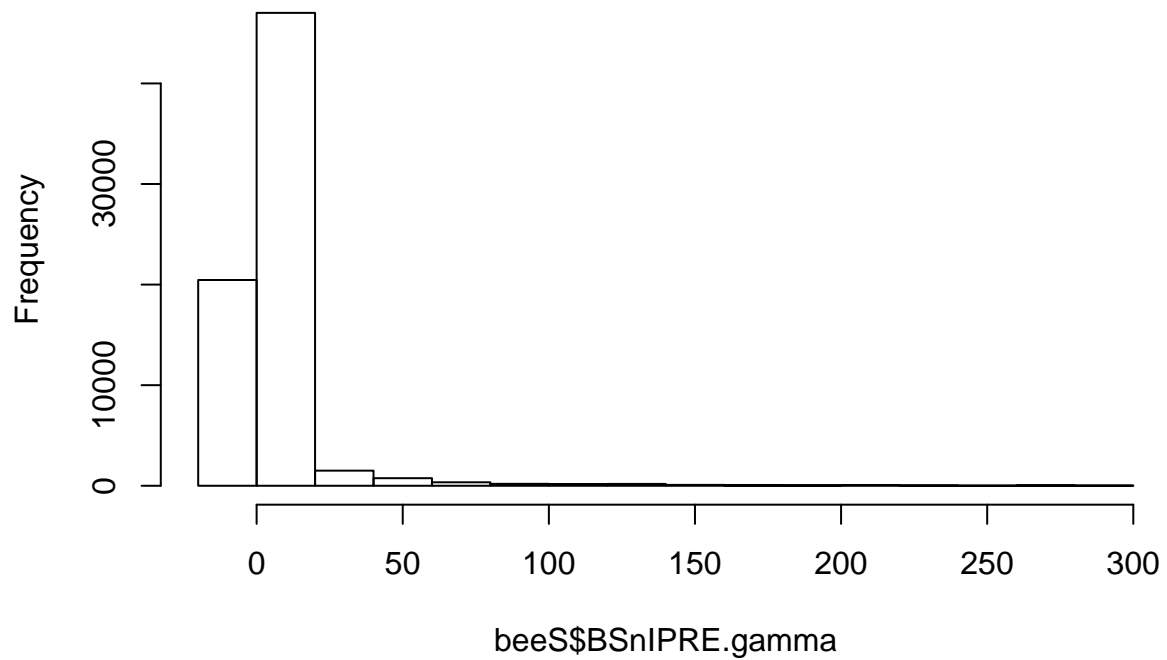
```
hist(antS$BSnIPRE.gamma)
```

**Histogram of ant\$BSnIPRE.gamma**



```
hist(bee$BSnIPRE.gamma)
```

**Histogram of bee\$BSnIPRE.gamma**





### Part 3: Positive selection, alpha

```
ggplot(alphaResults, aes(x=stage, y=alpha, fill=caste)) +  
  geom_bar(stat="identity") +  
  geom_errorbar(aes(ymin=c1, ymax=c2), width=0.2) +  
  facet_wrap(. ~ species) +  
  theme(axis.text.x = element_text(angle=90))
```

