



in Science: beginner's course

R is a high-level open source programming language and an environment for data analysis and graphics. This is a course about how to use R & RStudio and how to perform applications in statistics, phylogenetics, and data visualizations. The course includes several sessions and each contains a presentation and an exercise for practical training.

Day 1

Oral presentation about R Essentials of the R language

- basic commands
- data import
- data export
- data re-arrangement

Day 2

Statistics and data visualization

- correlation
- GLMM (generalized linear mixed models)
- multiple comparison analysis in ANOVA (analysis of variance)
- PCA (principle component analysis)

Phylogenetics

- data import
- data modification
- data analysis
- data export

```
#####  
#combine Fst, DD and Nielsen into one table#   
#####  
wholedata<-  
-data.frame(datafst$Chromosome,dataDD$start_pos,dataDD$end_pos,  
dataDD$res,datafst$Fst,data2dsfs$Nielsen_diversity)  
names(wholedata)<-  
c("Chromosome","Start_pos","End_pos","DD_residuals","Fst","Nielsen")  
  
#####  
#kill windows bigger than 100 kb#   
#####  
wholedata<-wholedata[wholedata$End_pos-  
wholedata$Start_pos<100000,]  
write.table(wholedata,"AllpostUGtests.csv",col.names=TRUE,sep="\t")  
  
#####  
#calculate average and median of window size#   
#####  
mean(wholedata$End_pos-wholedata$Start_pos)  
median(wholedata$End_pos-wholedata$Start_pos)  
  
#####  
#open list of A. lyrata genes#   
#####  
genelist<-read.table("Alyrata_107_geneS.gff3")  
#V4=start_pos  
#V5=end_pos  
names(genelist)<-  
c("Scaffold","gene_no","Start_pos","End_pos","?","?","?","ID")  
Chromosome1<-genelist[genelist$Scaffold=="scaffold_1",]  
Chromosome2<-genelist[genelist$Scaffold=="scaffold_2",]  
Chromosome3<-genelist[genelist$Scaffold=="scaffold_3",]  
Chromosome4<-genelist[genelist$Scaffold=="scaffold_4",]
```