

Transcriptomic analysis about tropospheric ozone tolerance in *Abies religiosa*.

Verónica Reyes Galindo, Lab. Genética y Evolución,
Instituto de Ecología, UNAM, México.



@Verolarrachitai



veronica.rg.pb@gmail.com



Tutora Principal: **Dra. Alicia Mastretta-Yanes** CONACYT-CONABIO.

Co-Tutor: **Dr. Juan Pablo Jaramillo Correa** Instituto de Ecología, UNAM.

Comité Tutor: **Dra. Alejandra Moreno Letelier** Instituto de Biología, UNAM.

14 Agosto 2019

Characteristics of my individual (It is called Christmas tree)

Biology

- Native tree to the mountains of the central of Mexico
- The leaves are needle-like
- Is the preferred tree for the monarch butterfly and is a important tree of forest near of CDMX

Sacred fir: *A. religiosa*



Branch

Leaf

Bioinformatics

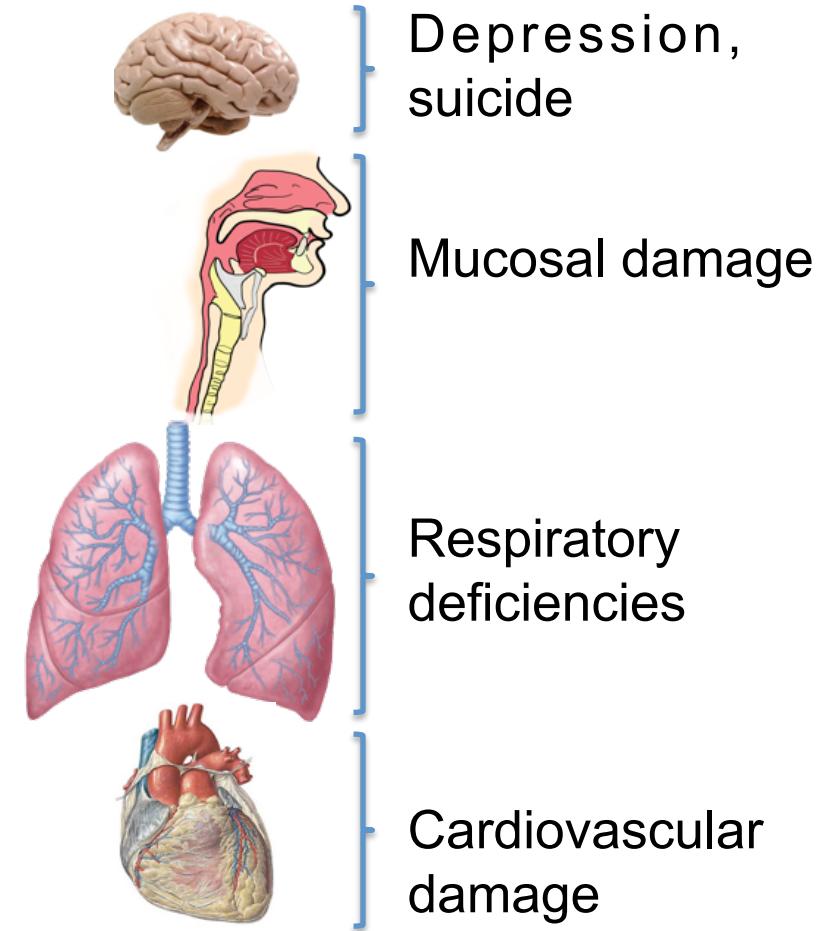
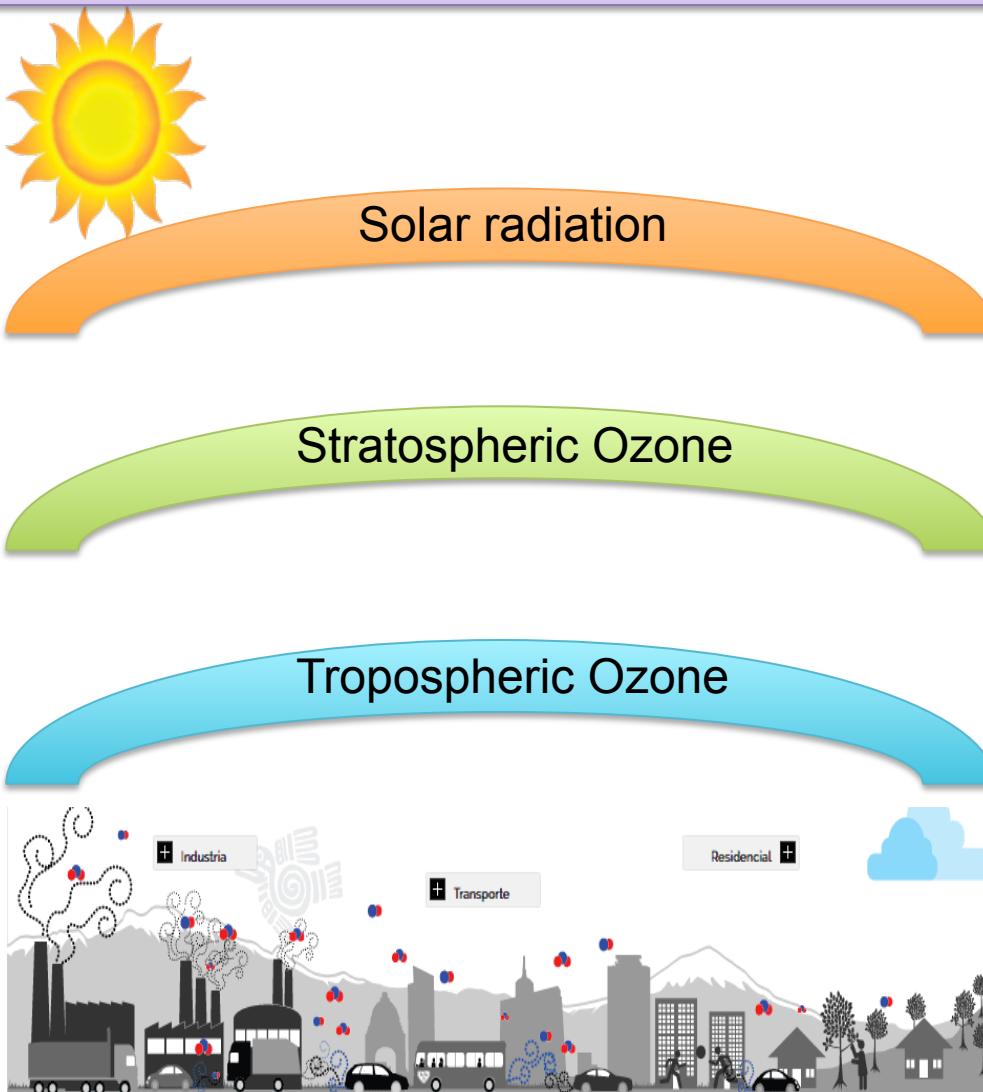
- The genome is 15-18 Gb
- Haven't reference genome
- Biallelic



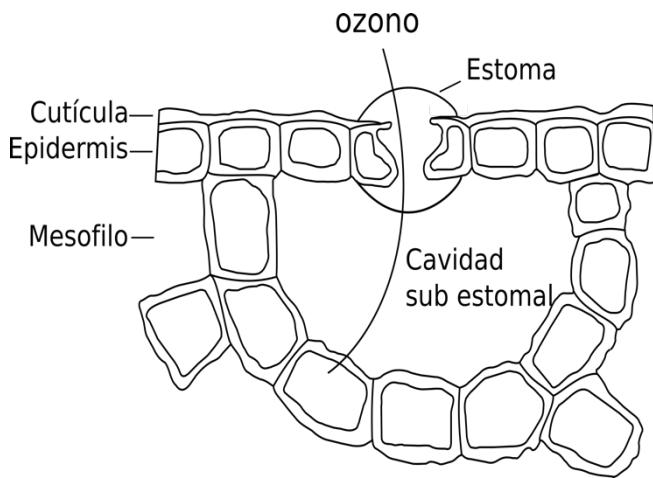
But... I don't have a box



Exposure to tropospheric ozone (O_3) is harmful to health



O_3 causes stress in plants

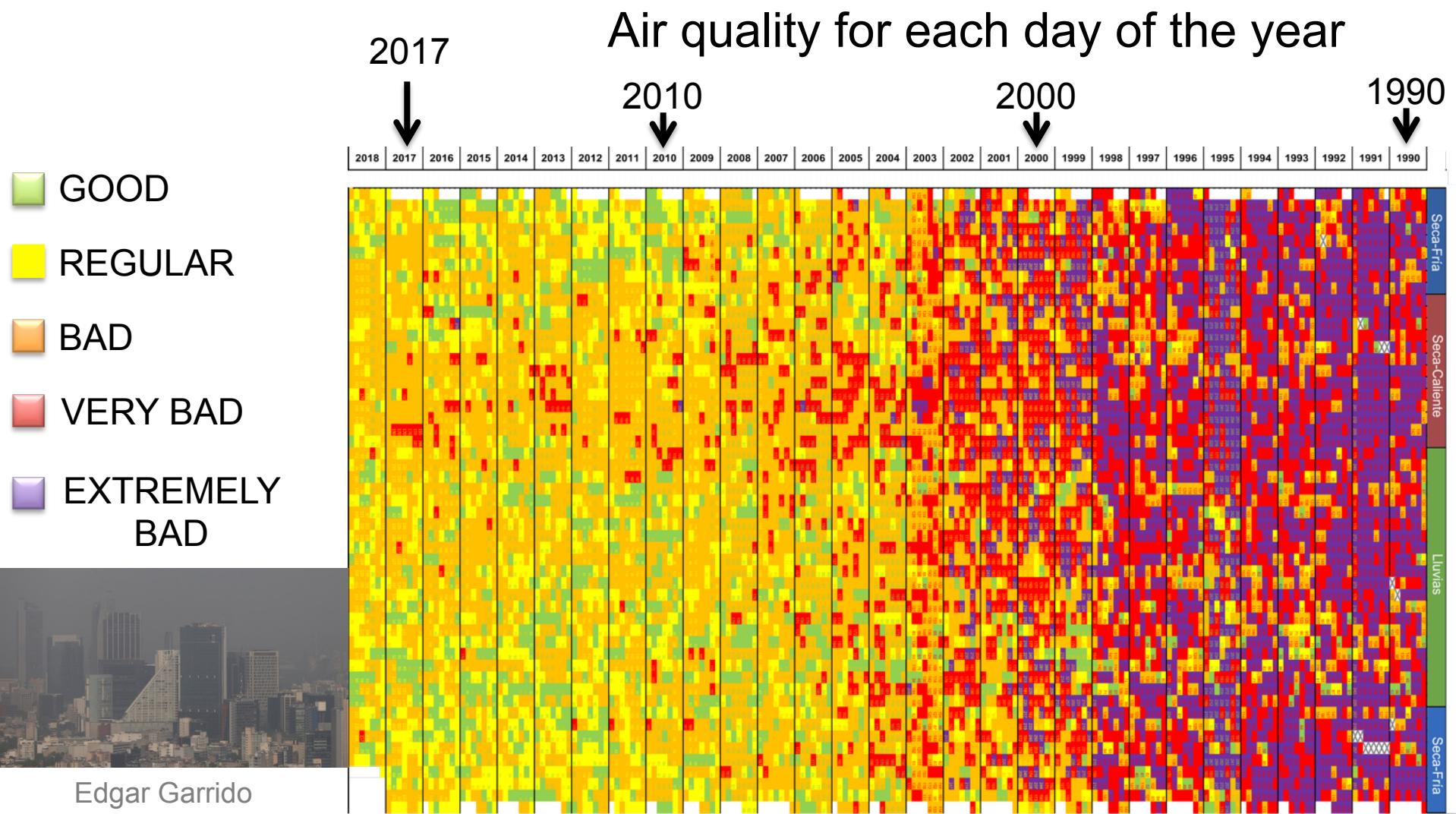


O_3 enters plant cells
through stomata

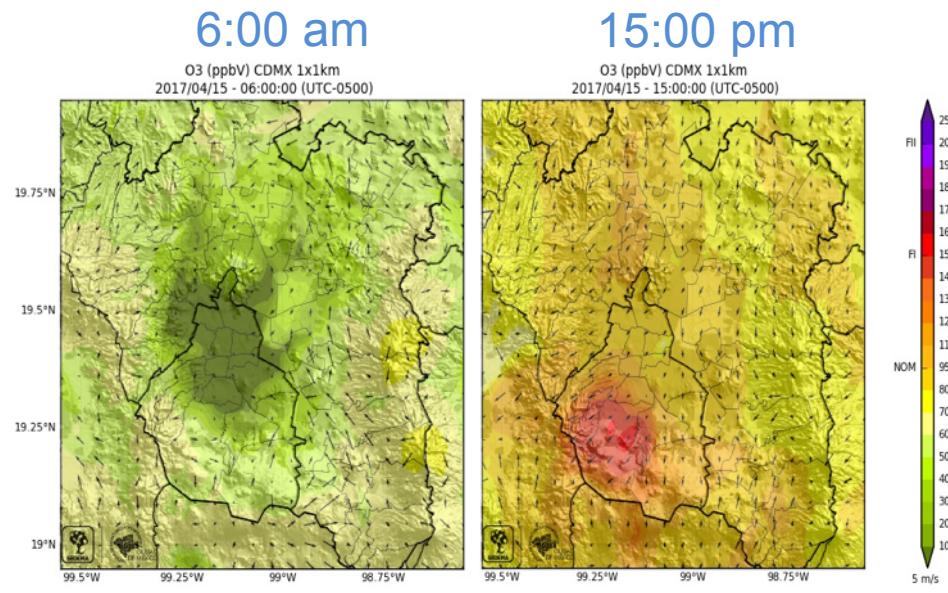


Ozone generally causes a **reddish**
color in the leaves

CDMX is the ideal place to present periods with high O₃ production



There are atmospheric drains that flow into forests near the CDMX; “Sacred fir graveyard”



Wind direction and O₃ concentration in the CDMX in the morning and afternoon.



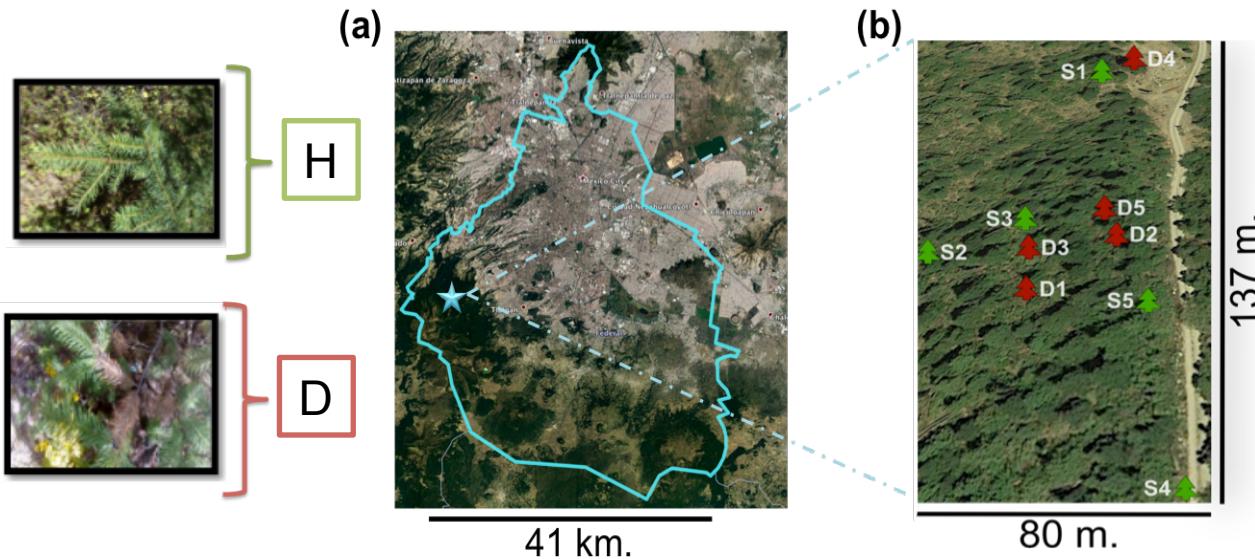
Panoramic view of an “Sacred fir graveyard” in Santa Rosa Xochiac

Principal aim

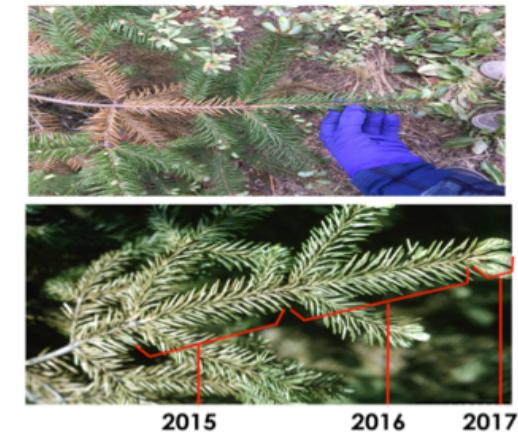
To assess whether the tolerance of sacred fir (*Abies religiosa*) to exposure by O₃, in a peripheral forest of the CDMX, is related to the genetic origin of the individuals or is given by differential expression of candidate genes.

- Evaluate the differential expression of **healthy** and **damaged** trees in two periods of [O₃].
- Identify origins of sacred fir with tolerance to O₃.
- Quantify the relative abundance of secondary metabolites in **healthy** y **damaged** trees in two periods of concentration of O₃
- Compare the histological differences in **healthy** and **damaged** trees during a contingency period (170ppb)

1. Were selected 5 individuals **healty** and 5 **damaged**.



2. The year of growth was identified in 6 branches/tree.

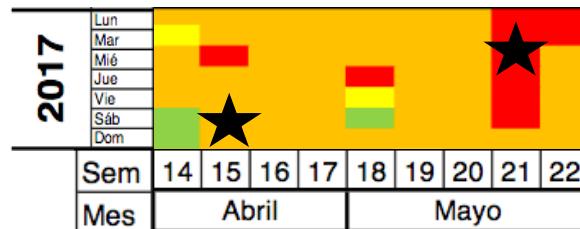


3. Two samplings of the same individuals in different periods of concentration of O₃.

15-April-2017
(87 ppb)



REGULAR



CONTINGENCY PHASE 1

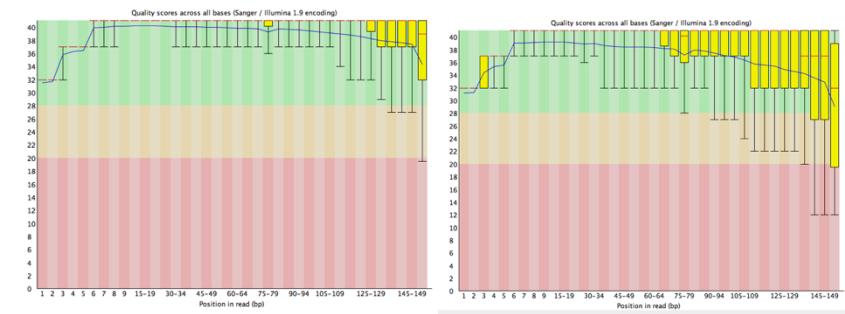
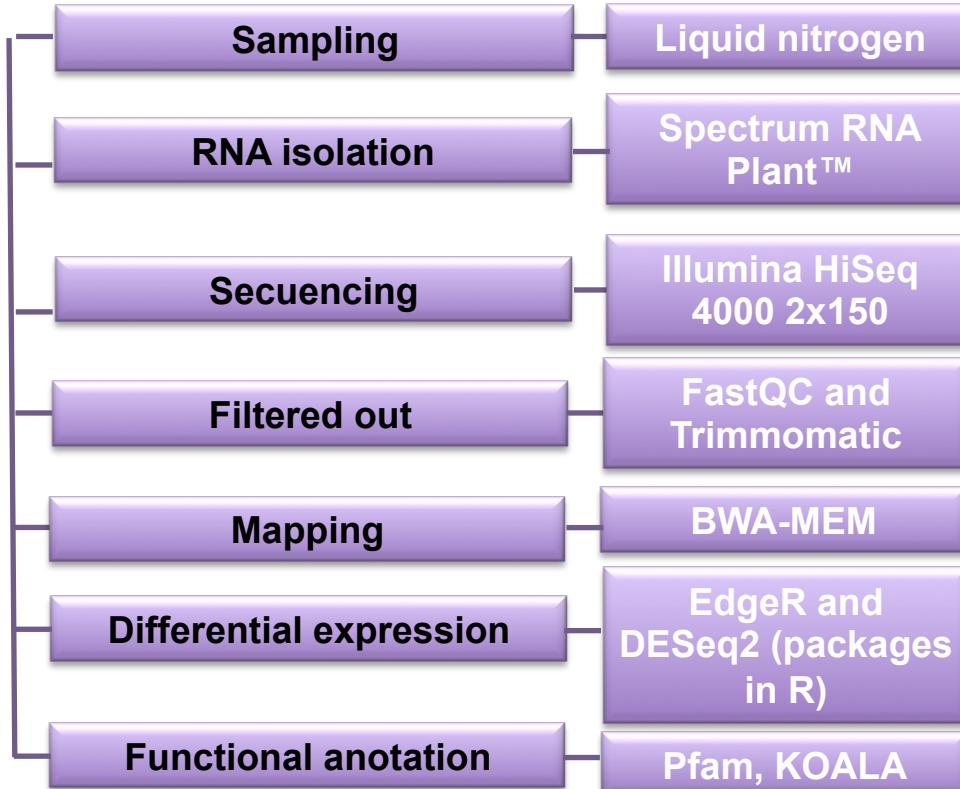
17-May-2017
(170 ppb)



VERY BAD

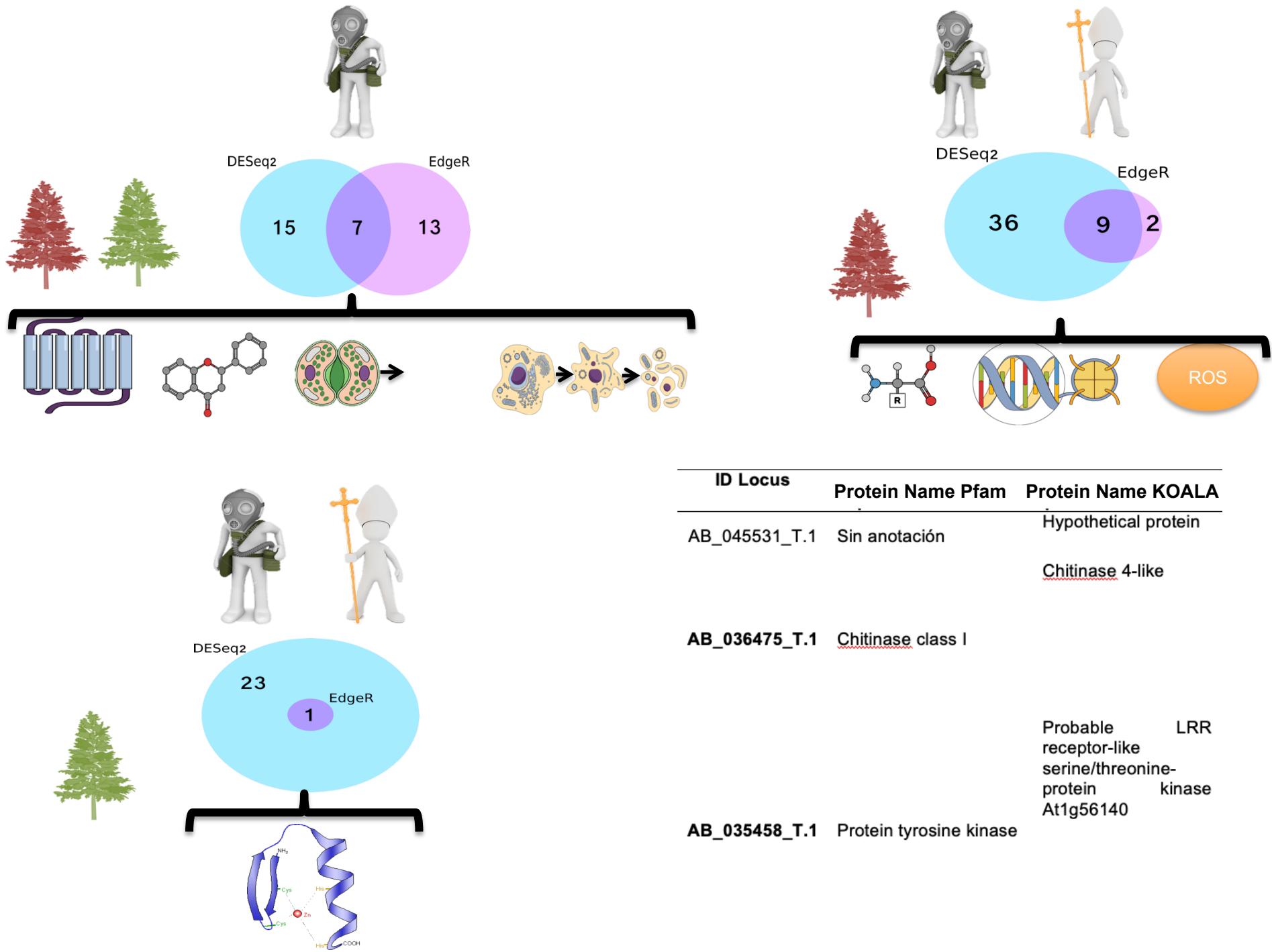
Evaluate the differential expression of **healthy** and **damaged** trees in two periods of [O₃]

Material and methods



84.5-96.7% Mapping
A. balsamea

605,147,387 total reads
33,619,299 reads per sample



Identify origins of sacred fir with tolerance to O₃.

Material and Methods

Sampling

Liquid nitrogen

DNA isolation

DNeasy® Plant Mini Kit

Secuencing

Illumina HiSeq2500 1x100.

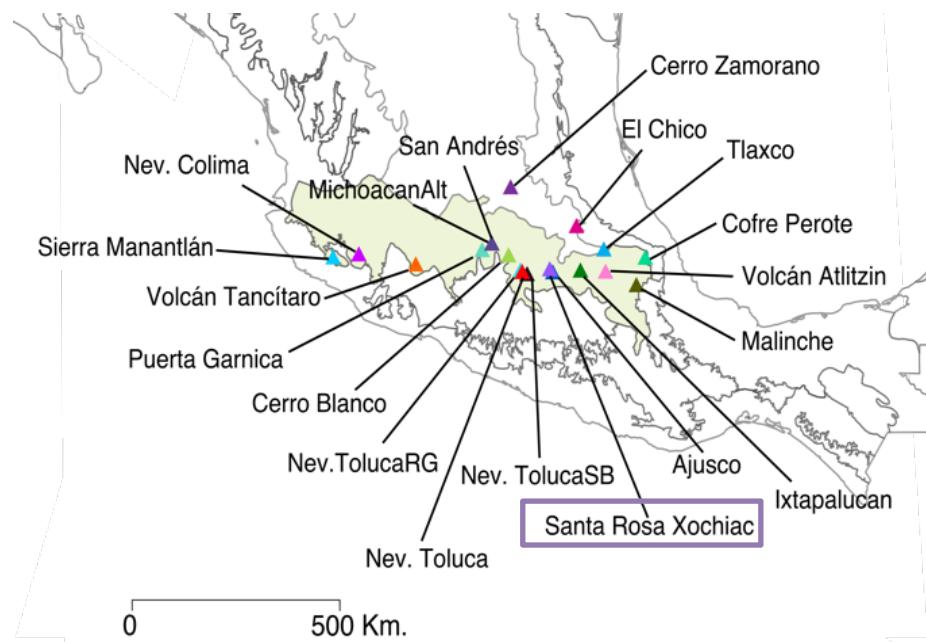
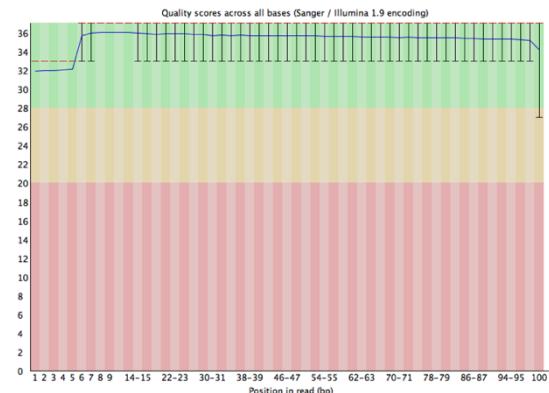
de novo assambling and filtered out

Ipyrad, FastQC, Plink y VCFTools

Detection of genetics groups

Mantel test (SNPRelate-geosphaera), relatedness (plink), PCA (SNPRelate), ADMIXTURE

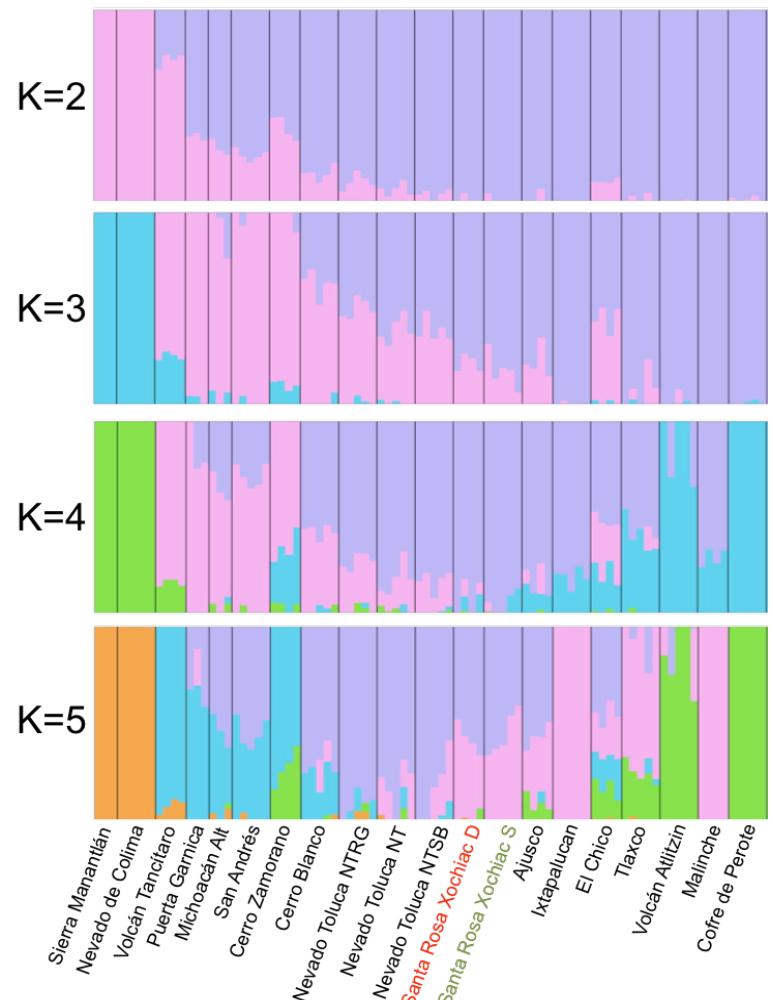
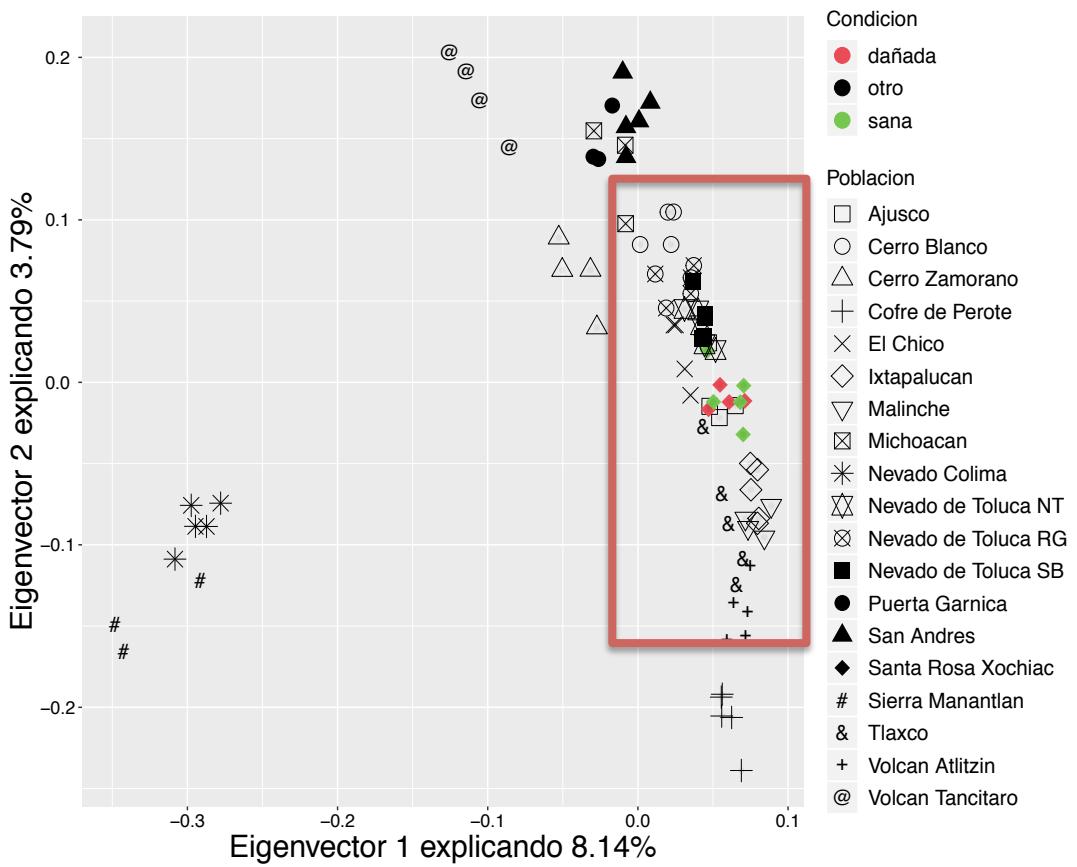
GBS: Mspl (C|CGG) y PstI (TGCA|G)



1550 SNPs, without Duplicates
SNPs
MAF > 0.05, missing data <10%



West → East



Quantify the relative abundance of secondary metabolites in **healthy** y **damaged** trees in two periods of concentration of O₃

Material and Methods

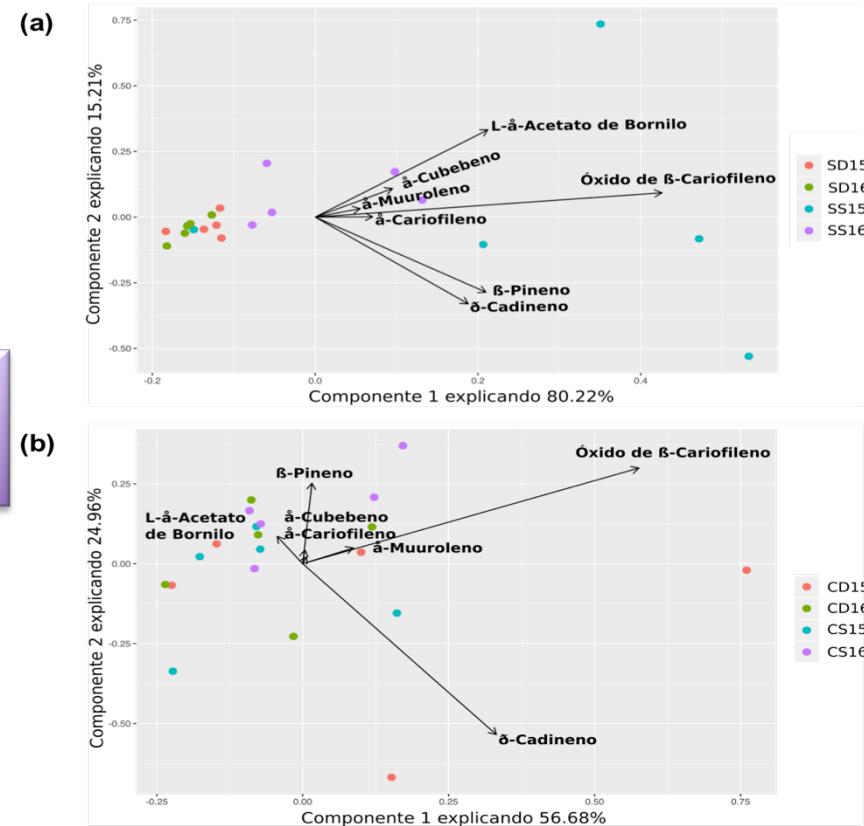
Sampling → Liquid nitrogen

Tissue extraction → Dycloromethano

Metabolite detection → Gas chromatography-mass spectrometry (GC-MS)

Compound quantification → standar: 4-Isopropylphenol

Statistical validation → Multivariate analysis

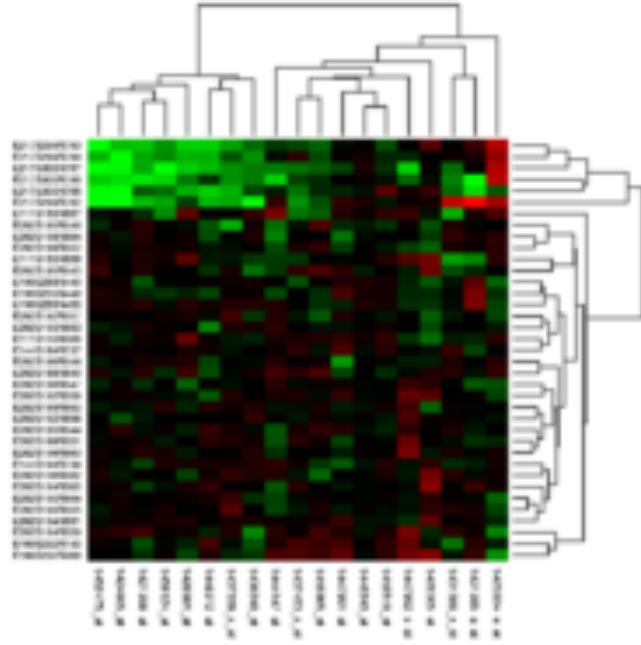


H... HELP

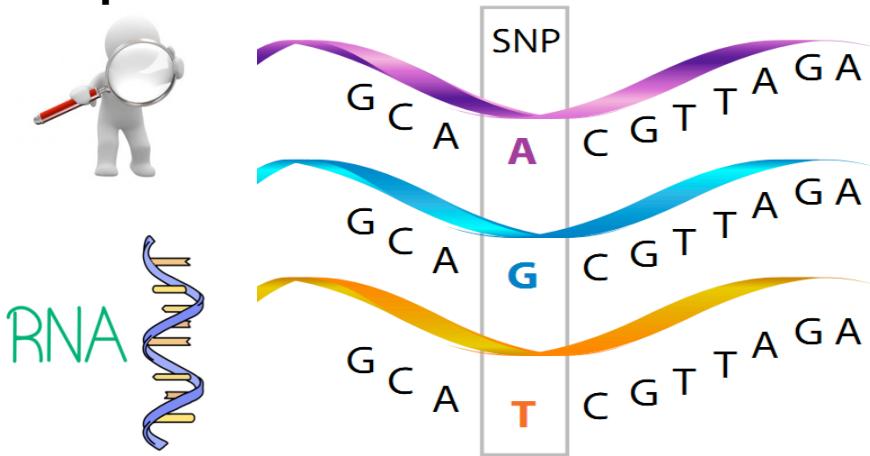
Where do I need help?



Generate my pipeline



Search for SNPs in differentially expressed transcripts



Determinate over-expressed and down-expressed genes

H... HELP

Where do I need help?



Bioinformatic Loops



```
#Calcular X2,X3,X4,X5 para cada metabolito

# beta.pinene
metabolitos2$X2_beta.pinene <- metabolitos2$x1*metabolitos$beta.pinene/metabolitos$P.Area.estandar
metabolitos2$X3_beta.pinene <- metabolitos2$X2_beta.pinene*metabolitos$microlitros.de.metabolitos/metabolitos$microlitros.de.inyeccion
metabolitos2$X4_beta.pinene <- metabolitos2$X3_beta.pinene*100/metabolitos$Peso.acicula.gramos
metabolitos2$X5_beta.pinene <- metabolitos2$X4_beta.pinene*1000

# L.alfa.bornyl.acetate
metabolitos2$X2_L.alfa.bornyl.acetate <- metabolitos2$x1*metabolitos$L.alfa.bornyl.acetate/metabolitos$P.Area.estandar
metabolitos2$X3_L.alfa.bornyl.acetate <- metabolitos2$X2_L.alfa.bornyl.acetate*metabolitos$microlitros.de.metabolitos/metabolitos$microlitros.de.inyeccion
metabolitos2$X4_L.alfa.bornyl.acetate <- metabolitos2$X3_L.alfa.bornyl.acetate*100/metabolitos$Peso.acicula.gramos
metabolitos2$X5_L.alfa.bornyl.acetate <- metabolitos2$X4_L.alfa.bornyl.acetate*1000
# beta.Caryophyllene.oxide
```

```
#!/bin/bash

#SBATCH -p keri
#SBATCH -n 5
#SBATCH --mem=10000

# Order transcripts
```

```
samtools sort ../DC01_15_sw10L50_TR.bam -l 9 -o DC01_15_sw10L50_TR_orderTrans.bam -n
samtools sort ../DC02_15_sw10L50_TR.bam -l 9 -o DC02_15_sw10L50_TR_orderTrans.bam -n
samtools sort ../DC03_15_sw10L50_TR.bam -l 9 -o DC03_15_sw10L50_TR_orderTrans.bam -n
samtools sort ../DC04_15_sw10L50_TR.bam -l 9 -o DC04_15_sw10L50_TR_orderTrans.bam -n
samtools sort ../DC05_15_sw10L50_TR.bam -l 9 -o DC05_15_sw10L50_TR_orderTrans.bam -n
```

>_

Learn more about the info I can get from each file and different ways to process them (*vcf, bcf, sam, bam, plink*)

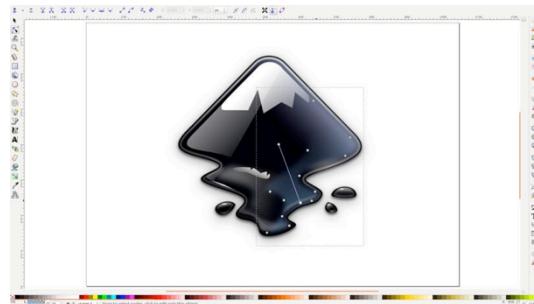
```
##fileformat=VCFv4.1
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d610ff66beb2da,species="Homo sapiens",phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0/0:48:2:0
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0/0:49:1:0
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333;0.667;AA=T;DB GT:GQ:DP:HQ 1/2:21:0:0
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0/0:54:1:0
20 1234567 microsat1 GTC G,GT,C 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:0
```

H... HELP

Where do I need help?



Improve my scripts and analyzes that I have done so far



Use editors less frequently to improve my figures



GLM to statistical analysis in R



- The genome is 15-18 Gb.
- Haven't reference genome.
- I use a transcriptome reference of *A. balsamea*



Mario Benavides
Antecedente de sitios
de reforestación



Claudio Zamora
Presidente de Bienes
Comunales de Santa
Rosa Xochiac



Ricardo Torres
Experto en ozono



Estela Sandoval
Experta en Anatomía
vegetal



César Flores
Experto en metabolitos



Lewis Spurgin
Experto en
genomas



Claudia
Martin
Experta en
genomas



David Prince
Experto en
transcriptomas



Alicia Mastretta, Juan Pablo Jaramillo
y Alejandra Moreno
Tutores y revisores