PYTHON

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

import os

import statistics as st

base\_dir = os.path.normpath("/Users/celestecohen/Downloads/Biofilm\_bioinformatics")

#This would be the Code folder on google drive, where all the other required files are

st\_filename = base\_dir + os.sep + "filtering\_results.csv"

filt=pd.read\_csv(st\_filename,header=0,sep="\t")

#Make the filtering values into a ratio rather than a % to match the washing values

filt["%flocc"]=filt["%flocc"]/100

filt

#Plot filtering phenotypes

plt.rcParams["figure.figsize"] = (30,5.5)

fig, ax = plt.subplots()

segregants=['R' + str(ind) for ind in list(range(1,55))]

strain\_order=np.concatenate((["JB50","JB759"],segregants))

x=list(range(0,len(set(filt['Strain']))))

y=[filt.loc[filt["Strain"]==name] for name in strain\_order]

colours=np.concatenate((["tomato","navajowhite"],np.repeat("cornflowerblue",54)))

ax.bar(x,

height=[np.mean(yi["%flocc"]) for yi in y],

yerr=[np.std(yi["%flocc"]) for yi in y], # error bars

capsize=6,

width=0.8, # bar width,

color=colours,

zorder = 0,

tick\_label=strain\_order

)

ax.set\_xlim(-1,56)

ax.set\_ylabel('Ratio of flocculated cells',fontsize=30,fontname="Arial")

ax.set\_xlabel("Parental strains and segregants",fontsize=30,fontname="Arial")

ax.set\_title("Filtering assay",fontsize=40,fontweight="bold",fontname="Arial")

plt.axhline(y=1, color='grey', linestyle='dashed', linewidth=1)

#Scatter plot of rep1 vs rep2

plt.rcParams["figure.figsize"] = (5,4)

fig, ax = plt.subplots()

filt\_mod=filt.copy()

filt\_mod["%flocc"][2]=st.mean([filt\_mod["%flocc"][1],filt\_mod["%flocc"][2]])

filt\_mod=filt\_mod.drop([1])

x=list(filt\_mod.loc[filt\_mod["Rep"]==1]["%flocc"])

y=list(filt\_mod.loc[filt\_mod["Rep"]==2]["%flocc"])

ax.scatter(x,y,color=colours,s=20)

ax.set\_ylabel('Ratio values from Rep 2',fontsize=11)

ax.set\_xlabel("Ratio values from Rep 1",fontsize=11)

#prepare for QTL

import re

filt\_grouped=filt.groupby("Strain",as\_index=False).mean()

phenotype=filt\_grouped[["Strain","%flocc"]]

#Make all segregant names R1\_x

for p in range(0,len(phenotype["Strain"])):

if phenotype["Strain"][p][0]=="R":

phenotype["Strain"][p]=str("R1\_"+re.findall("[0-9]+",phenotype["Strain"][p])[0])

#Make parental values X968 (JB50) and Y0036 (JB759)

phenotype["Strain"][np.where(phenotype=="JB50")[0]]="X968"

phenotype["Strain"][np.where(phenotype=="JB759")[0]]="Y0036"

#phenotype.to\_csv("filtering\_phenotypes.csv")

#when you want to actually create the phenotypes csv file, remove the #