# Greenish Warbler Genomic Analysis

Darren Irwin

January 2, 2025

This page contains notes and code describing the data analysis for a manuscript on Greenish Warbler genomics. I've been working with the data for several years, and the R and then Julia code has been in development for a while. This is a Quarto notebook, which can run and display the results of Julia (or other) code blocks, along with text narration, and output in html, pdf, Word, etc.

The Julia code here is loosely based on R code written for Greenish Warbler analysis (Irwin et al. 2016), and then the North American warbler analyses (Irwin et al. 2019), and then my (unpublished) 2019 Greenish Warbler analysis. Most recently, this was adapted from the scripts called GW2022_R_analysis_script.R and IrwinLabGenomicsAnalysisScript.jl but has had a lot of opimizations since then. The SNP data here are a result of GBS reads mapped to our new 2022 Biozeron genome assembly for a greenish warbler from southern China.

## Load packages

If running this for the first time, you will need to load packages used in the script, so run what is in this section below. It will take some time to install and precompile the packages:

```
import Pkg; Pkg.add("CSV") # took less than a minute
Pkg.add("DataFrames") # took about a minute
Pkg.add("Plots") # seems to install and working more simply than Makie (but less powerful)
Pkg.add("Haversine") # for great circle (Haversine) distances
Pkg.add("Distributions") # this seemed to fix a problem installing GLMakie
Pkg.add("MultivariateStats")
Pkg.add("StatsBase")
Pkg.add("Impute")
Pkg.add("JLD2")
Pkg.add("CairoMakie")
Pkg.add("PrettyTables") # for printing nice tables to REPL
Pkg.add("GenomicDiversity")
```

Now actually load those packages into the Julia session:

```
using CSV # for reading in delimited files
using DataFrames # for storing data as type DataFrame
using Haversine # for calculating Great Circle (haversine) distances between sites
using Statistics # for mean() function
using MultivariateStats # for Principal Coordinates Analysis (multidimensional scaling)
using DelimitedFiles # for reading delimited files (the genotypic data)
using Impute # for imputing missing genotypes
using JLD2 # for saving data
using CairoMakie # for plots
using PrettyTables
CairoMakie.activate!()  # this makes CairoMakie the main package for figures (in case another already l
```

Load my custom package `GenomicDiversity`:

```
using GenomicDiversity # actually make GenomicDiversity module available with GenomicDiversity.function
# or if functions are exported from GenomicDiversity then they are available.
```

Test Julia:

```
x = 1; y = 2; z = x+y
println("z = ", z)
```

```
z = 3
```

(If Quarto is calling Julia properly, you will see `z = 3` as the output of the code block above.)

Choose working directory:

```
repoDirectory = pwd() # this gets the starting working directory, for later use
dataDirectory = "/Users/darrenirwin/Dropbox/Darren's current work/"
cd(dataDirectory)
```

### Load the chromosome (scaffold) lengths

The code below will load a fasta index file for the reference genome, and then make a dictionary (a look-up table) where the key is the scaffold name and value is the chromosome length. This will be used in the genotype-by-individual figures.

```
cd(repoDirectory)
scaffold_info_filepath = "metadata/GW2022ref.fa.fai" # a fasta index file for the reference genome
scaffold_info = DataFrame(CSV.File(scaffold_info_filepath; header=["name", "length", "offset", "linebas
scaffold_lengths = Dict(scaffold_info.name .=> scaffold_info.length)
cd(dataDirectory)
```

### Set up file directories and names

```
# choose path and filename for the 012NA files
baseName = "GW_genomics_2022_with_new_genome/GW2022_GBS_012NA_files/GW2022_all4plates.genotypes.SNPs_on
filenameTextMiddle = ".max2allele_noindel.vcf.maxmiss"
# indicate percent threshold for missing genotypes for each SNP--
# this was set by earlier filtering, and is just a record-keeper for the filenames:
missingGenotypeThreshold = 60
filenameTextEnd = ".MQ20.lowHet.tab"
tagName = ".Jan2025."   # choose a tag name for this analysis
# indicate name of metadata file, a text file with these column headings:
# ID    location    group   Fst_group   plot_order
metadataFile = "GW_genomics_2022_with_new_genome/GW_all4plates.Fst_groups.txt"
```

```
"GW_genomics_2022_with_new_genome/GW_all4plates.Fst_groups.txt"
```

### List the scaffolds to be imputed (for later inclusion in PCA)

```
chromosomes_to_process = vec(["gw2",
                              "gw1",
                              "gw3",
                              "gwZ",
                              "gw1A",
                              "gw4",
                              "gw5",
                              "gw7",
                              "gw6",
                              "gw8",
                              "gw9",
                              "gw11",
                              "gw12",
                              "gw10",
```

```
                            "gw13",
                            "gw14",
                            "gw18",
                            "gw20",
                            "gw15",
                            "gw1B",
                            "gws100",
                            "gw17",
                            "gw19",
                            "gws101",
                            "gw4A",
                            "gw21",
                            "gw26",
                            "gws102",
                            "gw23",
                            "gw25",
                            "gws103",
                            "gw22",
                            "gws104",
                            "gw28",
                            "gw27",
                            "gw24",
                            "gws105",
                            "gws106",
                            "gws107",
                            "gws108",
                            "gws109",
                            "gws110",
                            "gws112"]);
```

**Set up function to correct the names of GBS runs:**

```
function correctNames(metadataColumn)
    metadataColumn_corrected = replace(metadataColumn,
                        "GW_Armando_plate1_TTGW05_rep2" => "GW_Armando_plate1_TTGW05r2",
                        "GW_Lane5_NA3-3ul" => "GW_Lane5_NA3",
                        "GW_Armando_plate1_TTGW_15_05" => "GW_Armando_plate1_TTGW-15-05",
                        "GW_Armando_plate1_TTGW_15_07" => "GW_Armando_plate1_TTGW-15-07",
                        "GW_Armando_plate1_TTGW_15_08" => "GW_Armando_plate1_TTGW-15-08",
                        "GW_Armando_plate1_TTGW_15_09" => "GW_Armando_plate1_TTGW-15-09",
```

```
                        "GW_Armando_plate1_TTGW_15_01" => "GW_Armando_plate1_TTGW-15-01",
                        "GW_Armando_plate1_TTGW_15_02" => "GW_Armando_plate1_TTGW-15-02",
                        "GW_Armando_plate1_TTGW_15_03" => "GW_Armando_plate1_TTGW-15-03",
                        "GW_Armando_plate1_TTGW_15_04" => "GW_Armando_plate1_TTGW-15-04",
                        "GW_Armando_plate1_TTGW_15_06" => "GW_Armando_plate1_TTGW-15-06",
                        "GW_Armando_plate1_TTGW_15_10" => "GW_Armando_plate1_TTGW-15-10",
                        "GW_Armando_plate2_TTGW_15_01" => "GW_Armando_plate2_TTGW-15-01",
                        "GW_Armando_plate2_TTGW_15_02" => "GW_Armando_plate2_TTGW-15-02",
                        "GW_Armando_plate2_TTGW_15_03" => "GW_Armando_plate2_TTGW-15-03",
                        "GW_Armando_plate2_TTGW_15_04" => "GW_Armando_plate2_TTGW-15-04",
                        "GW_Armando_plate2_TTGW_15_06" => "GW_Armando_plate2_TTGW-15-06",
                        "GW_Armando_plate2_TTGW_15_10" => "GW_Armando_plate2_TTGW-15-10")
end
```

```
correctNames (generic function with 1 method)
```

**The above should be run before any of the other Quarto pages are run**

### *OK, let's load the genomic data!*

```
# load metadata
metadata = DataFrame(CSV.File(metadataFile)) # the CSV.File function interprets the correct delimiter
num_metadata_cols = ncol(metadata)
num_individuals = nrow(metadata)
# read in individual names for this dataset
individuals_file_name = string(baseName, filenameTextMiddle, missingGenotypeThreshold, filenameTextEnd,
ind = DataFrame(CSV.File(individuals_file_name; header=["ind"], types=[String]))
indNum = size(ind, 1) # number of individuals
if num_individuals != indNum
    println("WARNING: number of rows in metadata file different than number of individuals in .indv fil
end
# read in position data for this dataset
position_file_name = string(baseName, filenameTextMiddle, missingGenotypeThreshold, filenameTextEnd, ".
pos_whole_genome = DataFrame(CSV.File(position_file_name; header=["chrom", "position"], types=[String,
# read in genotype data
genotype_file_name = string(baseName, filenameTextMiddle, missingGenotypeThreshold, filenameTextEnd, ".
@time if 1 <= indNum <= 127
    geno = readdlm(genotype_file_name, '\t', Int8, '\n'); # this has been sped up dramatically, by firs
elseif 128 <= indNum <= 32767
    geno = readdlm(genotype_file_name, '\t', Int16, '\n'); # this needed for first column, which is num
```

```
else
    print("Error: Number of individuals in .indv appears outside of range from 1 to 32767")
end
loci_count = size(geno, 2) - 1   # because the first column is not a SNP (just a count from zero)
print(string("Read in genotypic data at ", loci_count," loci for ", indNum, " individuals. \n"))
```

```
 82.744376 seconds (3.67 M allocations: 15.966 GiB, 8.43% gc time)
Read in genotypic data at 2431709 loci for 310 individuals.
```

### Check that individuals are same in genotype data and metadata

```
ind_with_metadata = hcat(ind, metadata)
println(ind_with_metadata)
println()  # prints a line break
if isequal(ind_with_metadata.ind, ind_with_metadata.ID)
    println("GOOD NEWS: names of individuals in metadata file and genotype ind file match perfectly.")
else
    println("WARNING: names of individuals in metadata file and genotype ind file do not completely mat
end
```

```
310×6 DataFrame
 Row │ ind                    ID                     location  group       Fst_group     plot_order
     │ String                 String31               String7   String15    String15      Float64
─────┼──────────────────────────────────────────────────────────────────────────────────────────────
   1 │ GW_Armando_plate1_AB1  GW_Armando_plate1_AB1  AB        vir         vir             20.01
   2 │ GW_Armando_plate1_JF07G02  GW_Armando_plate1_JF07G02  ST  plumb  plumb  108.0
   3 │ GW_Armando_plate1_JF07G03  GW_Armando_plate1_JF07G03  ST  plumb  plumb  109.0
   4 │ GW_Armando_plate1_JF07G04  GW_Armando_plate1_JF07G04  ST  plumb  plumb  110.0
   5 │ GW_Armando_plate1_JF08G02  GW_Armando_plate1_JF08G02  ST  plumb  plumb  111.0
   6 │ GW_Armando_plate1_JF09G01  GW_Armando_plate1_JF09G01  ST  plumb  plumb  112.0
   7 │ GW_Armando_plate1_JF09G02  GW_Armando_plate1_JF09G02  ST  plumb  plumb  113.0
   8 │ GW_Armando_plate1_JF10G03  GW_Armando_plate1_JF10G03  ST  plumb_vir  plumb_vir  17
   9 │ GW_Armando_plate1_JF11G01  GW_Armando_plate1_JF11G01  ST  plumb  plumb  114.0
  10 │ GW_Armando_plate1_JF12G01  GW_Armando_plate1_JF12G01  ST  plumb  plumb  115.
  11 │ GW_Armando_plate1_JF12G02  GW_Armando_plate1_JF12G02  ST  plumb  plumb  116.
  12 │ GW_Armando_plate1_JF12G04  GW_Armando_plate1_JF12G04  ST_vi  vir  vir  24.0
  13 │ GW_Armando_plate1_JF13G01  GW_Armando_plate1_JF13G01  ST  plumb  plumb  117.
```

```
 14 | GW_Armando_plate1_JF15G03    GW_Armando_plate1_JF15G03    DV     plumb       plumb         103.
 15 | GW_Armando_plate1_JF16G01    GW_Armando_plate1_JF16G01    DV_vi   plumb_vir    vir           24
 16 | GW_Armando_plate1_JF20G01    GW_Armando_plate1_JF20G01    MB     plumb       plumb          94.0
 17 | GW_Armando_plate1_JF22G01    GW_Armando_plate1_JF22G01    MB     plumb       plumb          95.0
 18 | GW_Armando_plate1_JF23G01    GW_Armando_plate1_JF23G01    VB     plumb       plumb          98.0
 19 | GW_Armando_plate1_JF23G02    GW_Armando_plate1_JF23G02    VB     plumb       plumb          99.0
 20 | GW_Armando_plate1_JF24G02    GW_Armando_plate1_JF24G02    VB     plumb       plumb         100.
 21 | GW_Armando_plate1_JF26G01    GW_Armando_plate1_JF26G01    ST     plumb       plumb         118.
 22 | GW_Armando_plate1_JF27G01    GW_Armando_plate1_JF27G01    ST     plumb       plumb         119.
 23 | GW_Armando_plate1_JF29G01    GW_Armando_plate1_JF29G01    ST     plumb       plumb         120.
 24 | GW_Armando_plate1_JF29G02    GW_Armando_plate1_JF29G02    ST     plumb       plumb         121.
 25 | GW_Armando_plate1_JF29G03    GW_Armando_plate1_JF29G03    ST     plumb       plumb         122.
 26 | GW_Armando_plate1_JG02G02    GW_Armando_plate1_JG02G02    PR     plumb       plumb         145.
 27 | GW_Armando_plate1_JG02G04    GW_Armando_plate1_JG02G04    PR     plumb       plumb         146.
 28 | GW_Armando_plate1_JG08G01    GW_Armando_plate1_JG08G01    ST     plumb       plumb         123.
 29 | GW_Armando_plate1_JG08G02    GW_Armando_plate1_JG08G02    ST     plumb       plumb         124.
 30 | GW_Armando_plate1_JG10G01    GW_Armando_plate1_JG10G01    ST     plumb       plumb         125.
 31 | GW_Armando_plate1_JG12G01    GW_Armando_plate1_JG12G01    ST     plumb       plumb         126.
 32 | GW_Armando_plate1_JG17G01    GW_Armando_plate1_JG17G01    ST     plumb_vir    plumb          127
 33 | GW_Armando_plate1_NO_BC_TTGW05 GW_Armando_plate1_NO_BC_TTGW05 blank   blank       blank
99.0
 34 | GW_Armando_plate1_NO_DNA     GW_Armando_plate1_NO_DNA     blank   blank       blank         -
99.0
 35 | GW_Armando_plate1_RF20G01    GW_Armando_plate1_RF20G01    BJ     obs_plumb    plumb_BJ        7
 36 | GW_Armando_plate1_RF29G02    GW_Armando_plate1_RF29G02    BJ     obs_plumb    plumb_BJ        7
 37 | GW_Armando_plate1_TL3       GW_Armando_plate1_TL3       TL     vir         vir           11.01
 38 | GW_Armando_plate1_TTGW01     GW_Armando_plate1_TTGW01     MN     troch_MN     troch_west       53
 39 | GW_Armando_plate1_TTGW05_rep1  GW_Armando_plate1_TTGW05_rep1  MN_rep   troch_MN_rep   troch_west_r
 40 | GW_Armando_plate1_TTGW05_rep2  GW_Armando_plate1_TTGW05_rep2  MN     troch_MN     troch_west
 41 | GW_Armando_plate1_TTGW06     GW_Armando_plate1_TTGW06     SU     lud_Sukhto    lud_central      4
 42 | GW_Armando_plate1_TTGW07     GW_Armando_plate1_TTGW07     SU     lud_Sukhto    lud_central      4
 43 | GW_Armando_plate1_TTGW10     GW_Armando_plate1_TTGW10     SU     lud_Sukhto    lud_central      4
 44 | GW_Armando_plate1_TTGW11     GW_Armando_plate1_TTGW11     SU     lud_Sukhto    lud_central      4
 45 | GW_Armando_plate1_TTGW13     GW_Armando_plate1_TTGW13     TH     lud_Thallighar lud_central
 46 | GW_Armando_plate1_TTGW17     GW_Armando_plate1_TTGW17     TH     lud_Thallighar lud_central
 47 | GW_Armando_plate1_TTGW19     GW_Armando_plate1_TTGW19     TH     lud_Thallighar lud_central
 48 | GW_Armando_plate1_TTGW21     GW_Armando_plate1_TTGW21     SR     lud_Sural     lud_central      4
 49 | GW_Armando_plate1_TTGW22     GW_Armando_plate1_TTGW22     SR     lud_Sural     lud_central      4
```

| | | | | | | |
|---|---|---|---|---|---|---|
| 50 | GW_Armando_plate1_TTGW23 | GW_Armando_plate1_TTGW23 | SR | lud_Sural | lud_central | 4 |
| 51 | GW_Armando_plate1_TTGW29 | GW_Armando_plate1_TTGW29 | SR | lud_Sural | lud_central | 4 |
| 52 | GW_Armando_plate1_TTGW52 | GW_Armando_plate1_TTGW52 | NG | lud_Nainaghar | lud_central | |
| 53 | GW_Armando_plate1_TTGW53 | GW_Armando_plate1_TTGW53 | NG | lud_Nainaghar | lud_central | |
| 54 | GW_Armando_plate1_TTGW55 | GW_Armando_plate1_TTGW55 | NG | lud_Nainaghar | lud_central | |
| 55 | GW_Armando_plate1_TTGW57 | GW_Armando_plate1_TTGW57 | NG | lud_Nainaghar | lud_central | |
| 56 | GW_Armando_plate1_TTGW58 | GW_Armando_plate1_TTGW58 | NG | lud_Nainaghar | lud_central | |
| 57 | GW_Armando_plate1_TTGW59 | GW_Armando_plate1_TTGW59 | NG | lud_Nainaghar | lud_central | |
| 58 | GW_Armando_plate1_TTGW63 | GW_Armando_plate1_TTGW63 | SP | lud_Spiti | troch_west | 5 |
| 59 | GW_Armando_plate1_TTGW64 | GW_Armando_plate1_TTGW64 | SP | lud_Spiti | troch_west | 5 |
| 60 | GW_Armando_plate1_TTGW65 | GW_Armando_plate1_TTGW65 | SP | lud_Spiti | troch_west | 5 |
| 61 | GW_Armando_plate1_TTGW66 | GW_Armando_plate1_TTGW66 | SP | lud_Spiti | troch_west | 5 |
| 62 | GW_Armando_plate1_TTGW68 | GW_Armando_plate1_TTGW68 | SP | lud_Spiti | troch_west | 5 |
| 63 | GW_Armando_plate1_TTGW70 | GW_Armando_plate1_TTGW70 | SA | lud_Sathrundi | lud_Sath | 4 |
| 64 | GW_Armando_plate1_TTGW71 | GW_Armando_plate1_TTGW71 | SA | lud_Sathrundi | lud_Sath | 4 |
| 65 | GW_Armando_plate1_TTGW72 | GW_Armando_plate1_TTGW72 | SA | lud_Sathrundi | lud_Sath | 4 |
| 66 | GW_Armando_plate1_TTGW74 | GW_Armando_plate1_TTGW74 | SA | lud_Sathrundi | lud_Sath | 4 |
| 67 | GW_Armando_plate1_TTGW78 | GW_Armando_plate1_TTGW78 | SA | lud_Sathrundi | lud_Sath | 4 |
| 68 | GW_Armando_plate1_TTGW_15_05 | GW_Armando_plate1_TTGW_15_05 | SR | lud_Sural | lud_central | |
| 69 | GW_Armando_plate1_TTGW_15_07 | GW_Armando_plate1_TTGW_15_07 | SR | lud_Sural | lud_central | |
| 70 | GW_Armando_plate1_TTGW_15_08 | GW_Armando_plate1_TTGW_15_08 | SR | lud_Sural | lud_central | |
| 71 | GW_Armando_plate1_TTGW_15_09 | GW_Armando_plate1_TTGW_15_09 | SR | lud_Sural | lud_central | |
| 72 | GW_Armando_plate1_UY1 | GW_Armando_plate1_UY1 | UY | plumb | plumb | 87.0 |
| 73 | GW_Armando_plate2_IL2 | GW_Armando_plate2_IL2 | IL_rep | plumb_rep | plumb_rep | 84 |
| 74 | GW_Armando_plate2_JE31G01 | GW_Armando_plate2_JE31G01 | VB_vi | vir_misID | vir | 24 |
| 75 | GW_Armando_plate2_JF03G01 | GW_Armando_plate2_JF03G01 | ST_vi | vir_misID | vir | 24 |
| 76 | GW_Armando_plate2_JF03G02 | GW_Armando_plate2_JF03G02 | VB_vi | vir_misID | vir | 24 |
| 77 | GW_Armando_plate2_JF07G01 | GW_Armando_plate2_JF07G01 | ST | plumb | plumb | 128. |
| 78 | GW_Armando_plate2_JF08G04 | GW_Armando_plate2_JF08G04 | ST | plumb | plumb | 129. |
| 79 | GW_Armando_plate2_JF10G02 | GW_Armando_plate2_JF10G02 | ST | plumb | plumb | 130. |
| 80 | GW_Armando_plate2_JF11G02 | GW_Armando_plate2_JF11G02 | ST | plumb | plumb | 131. |
| 81 | GW_Armando_plate2_JF12G03 | GW_Armando_plate2_JF12G03 | ST | plumb | plumb | 132. |
| 82 | GW_Armando_plate2_JF12G05 | GW_Armando_plate2_JF12G05 | ST | plumb | plumb | 133. |
| 83 | GW_Armando_plate2_JF13G02 | GW_Armando_plate2_JF13G02 | ST | plumb | plumb | 134. |
| 84 | GW_Armando_plate2_JF14G01 | GW_Armando_plate2_JF14G01 | DV | plumb | plumb | 104. |
| 85 | GW_Armando_plate2_JF14G02 | GW_Armando_plate2_JF14G02 | DV | plumb | plumb | 105. |
| 86 | GW_Armando_plate2_JF15G01 | GW_Armando_plate2_JF15G01 | DV | plumb | plumb | 106. |

| | | | | | | |
|---|---|---|---|---|---|---|
| 87 | GW_Armando_plate2_JF15G02 | GW_Armando_plate2_JF15G02 | DV | plumb | plumb | 107. |
| 88 | GW_Armando_plate2_JF16G02 | GW_Armando_plate2_JF16G02 | DV_vi | plumb_vir | vir | 24 |
| 89 | GW_Armando_plate2_JF19G01 | GW_Armando_plate2_JF19G01 | MB | plumb | plumb | 96.0 |
| 90 | GW_Armando_plate2_JF20G02 | GW_Armando_plate2_JF20G02 | MB | plumb | plumb | 97.0 |
| 91 | GW_Armando_plate2_JF24G01 | GW_Armando_plate2_JF24G01 | VB | plumb | plumb | 101. |
| 92 | GW_Armando_plate2_JF24G03 | GW_Armando_plate2_JF24G03 | ST | plumb | plumb | 135. |
| 93 | GW_Armando_plate2_JF25G01 | GW_Armando_plate2_JF25G01 | VB | plumb | plumb | 102. |
| 94 | GW_Armando_plate2_JF26G02 | GW_Armando_plate2_JF26G02 | ST | plumb | plumb | 136. |
| 95 | GW_Armando_plate2_JF27G02 | GW_Armando_plate2_JF27G02 | ST | plumb | plumb | 137. |
| 96 | GW_Armando_plate2_JF30G01 | GW_Armando_plate2_JF30G01 | ST_vi | vir_misID | vir | 24 |
| 97 | GW_Armando_plate2_JG01G01 | GW_Armando_plate2_JG01G01 | PR | plumb | plumb | 147. |
| 98 | GW_Armando_plate2_JG02G01 | GW_Armando_plate2_JG02G01 | PR | plumb | plumb | 148. |
| 99 | GW_Armando_plate2_JG02G03 | GW_Armando_plate2_JG02G03 | PR | plumb | plumb | 149. |
| 100 | GW_Armando_plate2_JG10G02 | GW_Armando_plate2_JG10G02 | ST | plumb | plumb | 138. |
| 101 | GW_Armando_plate2_JG10G03 | GW_Armando_plate2_JG10G03 | ST | plumb | plumb | 139. |
| 102 | GW_Armando_plate2_JG12G02 | GW_Armando_plate2_JG12G02 | ST | plumb | plumb | 140. |
| 103 | GW_Armando_plate2_JG12G03 | GW_Armando_plate2_JG12G03 | ST | plumb | plumb | 141. |
| 104 | GW_Armando_plate2_LN11 | GW_Armando_plate2_LN11 | LN_rep | troch_LN_rep | troch_LN_rep | |
| 105 | GW_Armando_plate2_LN2 | GW_Armando_plate2_LN2 | LN | troch_LN | troch_LN | 58.0 |
| 106 | GW_Armando_plate2_NO_BC_TTGW05 | GW_Armando_plate2_NO_BC_TTGW05 | blank | blank | blank | |
| 99.0 | | | | | | |
| 107 | GW_Armando_plate2_NO_DNA | GW_Armando_plate2_NO_DNA | blank | blank | blank | - |
| 99.0 | | | | | | |
| 108 | GW_Armando_plate2_RF29G01 | GW_Armando_plate2_RF29G01 | BJ | obs_plumb | plumb_BJ | 7 |
| 109 | GW_Armando_plate2_TTGW02 | GW_Armando_plate2_TTGW02 | MN | troch_MN | troch_west | 5 |
| 110 | GW_Armando_plate2_TTGW03 | GW_Armando_plate2_TTGW03 | MN | troch_MN | troch_west | 5 |
| 111 | GW_Armando_plate2_TTGW05_rep3 | GW_Armando_plate2_TTGW05_rep3 | MN_rep | troch_MN_rep | troch_west_r | |
| 112 | GW_Armando_plate2_TTGW05_rep4 | GW_Armando_plate2_TTGW05_rep4 | MN_rep | troch_MN_rep | troch_west_r | |
| 113 | GW_Armando_plate2_TTGW08 | GW_Armando_plate2_TTGW08 | SU | lud_Sukhto | lud_central | |
| 114 | GW_Armando_plate2_TTGW09 | GW_Armando_plate2_TTGW09 | SU | lud_Sukhto | lud_central | |
| 115 | GW_Armando_plate2_TTGW12 | GW_Armando_plate2_TTGW12 | TH | lud_Thallighar | lud_central | |
| 116 | GW_Armando_plate2_TTGW14 | GW_Armando_plate2_TTGW14 | TH | lud_Thallighar | lud_central | |
| 117 | GW_Armando_plate2_TTGW15 | GW_Armando_plate2_TTGW15 | TH | lud_Thallighar | lud_central | |
| 118 | GW_Armando_plate2_TTGW16 | GW_Armando_plate2_TTGW16 | TH | lud_Thallighar | lud_central | |
| 119 | GW_Armando_plate2_TTGW18 | GW_Armando_plate2_TTGW18 | TH | lud_Thallighar | lud_central | |
| 120 | GW_Armando_plate2_TTGW20 | GW_Armando_plate2_TTGW20 | SR | lud_Sural | lud_central | 4 |
| 121 | GW_Armando_plate2_TTGW24 | GW_Armando_plate2_TTGW24 | SR | lud_Sural | lud_central | 4 |
| 122 | GW_Armando_plate2_TTGW25 | GW_Armando_plate2_TTGW25 | SR | lud_Sural | lud_central | 4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 123 | GW_Armando_plate2_TTGW27 | GW_Armando_plate2_TTGW27 | SR | lud_Sural | lud_central | 4 |
| 124 | GW_Armando_plate2_TTGW28 | GW_Armando_plate2_TTGW28 | SR | lud_Sural | lud_central | 4 |
| 125 | GW_Armando_plate2_TTGW50 | GW_Armando_plate2_TTGW50 | NG | lud_Nainaghar | lud_central | |
| 126 | GW_Armando_plate2_TTGW51 | GW_Armando_plate2_TTGW51 | NG | lud_Nainaghar | lud_central | |
| 127 | GW_Armando_plate2_TTGW54 | GW_Armando_plate2_TTGW54 | NG | lud_Nainaghar | lud_central | |
| 128 | GW_Armando_plate2_TTGW56 | GW_Armando_plate2_TTGW56 | NG | lud_Nainaghar | lud_central | |
| 129 | GW_Armando_plate2_TTGW60 | GW_Armando_plate2_TTGW60 | SP | lud_Spiti | troch_west | 5 |
| 130 | GW_Armando_plate2_TTGW61 | GW_Armando_plate2_TTGW61 | SP | lud_Spiti | troch_west | 5 |
| 131 | GW_Armando_plate2_TTGW62 | GW_Armando_plate2_TTGW62 | SP | lud_Spiti | troch_west | 5 |
| 132 | GW_Armando_plate2_TTGW67 | GW_Armando_plate2_TTGW67 | SP | lud_Spiti | troch_west | 5 |
| 133 | GW_Armando_plate2_TTGW69 | GW_Armando_plate2_TTGW69 | SP | lud_Spiti | troch_west | 5 |
| 134 | GW_Armando_plate2_TTGW73 | GW_Armando_plate2_TTGW73 | SA | lud_Sathrundi | lud_Sath | |
| 135 | GW_Armando_plate2_TTGW75 | GW_Armando_plate2_TTGW75 | SA | lud_Sathrundi | lud_Sath | |
| 136 | GW_Armando_plate2_TTGW77 | GW_Armando_plate2_TTGW77 | SA | lud_Sathrundi | lud_Sath | |
| 137 | GW_Armando_plate2_TTGW79 | GW_Armando_plate2_TTGW79 | SA | lud_Sathrundi | lud_Sath | |
| 138 | GW_Armando_plate2_TTGW80 | GW_Armando_plate2_TTGW80 | SA | lud_Sathrundi | lud_Sath | |
| 139 | GW_Armando_plate2_TTGW_15_01 | GW_Armando_plate2_TTGW_15_01 | SR | lud_Sural | lud_central | |
| 140 | GW_Armando_plate2_TTGW_15_02 | GW_Armando_plate2_TTGW_15_02 | SR | lud_Sural | lud_central | |
| 141 | GW_Armando_plate2_TTGW_15_03 | GW_Armando_plate2_TTGW_15_03 | SR | lud_Sural | lud_central | |
| 142 | GW_Armando_plate2_TTGW_15_04 | GW_Armando_plate2_TTGW_15_04 | SR | lud_Sural | lud_central | |
| 143 | GW_Armando_plate2_TTGW_15_06 | GW_Armando_plate2_TTGW_15_06 | SR | lud_Sural | lud_central | |
| 144 | GW_Armando_plate2_TTGW_15_10 | GW_Armando_plate2_TTGW_15_10 | SR | lud_Sural | lud_central | |
| 145 | GW_Lane5_AA1 | GW_Lane5_AA1 | AA | vir_S | vir_S | 25.0 |
| 146 | GW_Lane5_AA10 | GW_Lane5_AA10 | AA | vir_S | vir_S | 33.0 |
| 147 | GW_Lane5_AA11 | GW_Lane5_AA11 | AA | vir_S | vir_S | 34.0 |
| 148 | GW_Lane5_AA3 | GW_Lane5_AA3 | AA | vir_S | vir_S | 26.0 |
| 149 | GW_Lane5_AA4 | GW_Lane5_AA4 | AA | vir_S | vir_S | 27.0 |
| 150 | GW_Lane5_AA5 | GW_Lane5_AA5 | AA | vir_S | vir_S | 28.0 |
| 151 | GW_Lane5_AA6 | GW_Lane5_AA6 | AA | vir_S | vir_S | 29.0 |
| 152 | GW_Lane5_AA7 | GW_Lane5_AA7 | AA | vir_S | vir_S | 30.0 |
| 153 | GW_Lane5_AA8 | GW_Lane5_AA8 | AA | vir_S | vir_S | 31.0 |
| 154 | GW_Lane5_AA9 | GW_Lane5_AA9 | AA | vir_S | vir_S | 32.0 |
| 155 | GW_Lane5_AB1 | GW_Lane5_AB1 | AB_rep | vir_rep | vir_rep | 20.0 |
| 156 | GW_Lane5_AB2 | GW_Lane5_AB2 | AB | vir | vir | 21.0 |
| 157 | GW_Lane5_AN1 | GW_Lane5_AN1 | AN | plumb | plumb | 80.0 |
| 158 | GW_Lane5_AN2 | GW_Lane5_AN2 | AN | plumb | plumb | 81.0 |
| 159 | GW_Lane5_BK2 | GW_Lane5_BK2 | BK | plumb | plumb | 78.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 160 | GW_Lane5_BK3 | GW_Lane5_BK3 | BK | plumb | plumb | 79.0 |
| 161 | GW_Lane5_DA2 | GW_Lane5_DA2 | XN | obs | obs | 73.0 |
| 162 | GW_Lane5_DA3 | GW_Lane5_DA3 | XN | obs | obs | 74.0 |
| 163 | GW_Lane5_DA4 | GW_Lane5_DA4 | XN | obs | obs | 75.0 |
| 164 | GW_Lane5_DA6 | GW_Lane5_DA6 | XN | obs | low_reads | 76.0 |
| 165 | GW_Lane5_DA7 | GW_Lane5_DA7 | XN | obs | obs | 77.0 |
| 166 | GW_Lane5_EM1 | GW_Lane5_EM1 | EM | troch_EM | troch_EM | 72.0 |
| 167 | GW_Lane5_IL1 | GW_Lane5_IL1 | IL | plumb | plumb | 82.0 |
| 168 | GW_Lane5_IL2 | GW_Lane5_IL2 | IL_rep | plumb_rep | plumb_rep | 85.0 |
| 169 | GW_Lane5_IL4 | GW_Lane5_IL4 | IL | plumb | plumb | 83.0 |
| 170 | GW_Lane5_KS1 | GW_Lane5_KS1 | OV | lud_KS | lud_KS | 40.0 |
| 171 | GW_Lane5_KS2 | GW_Lane5_KS2 | OV | lud_KS | lud_KS | 40.0 |
| 172 | GW_Lane5_LN1 | GW_Lane5_LN1 | LN | troch_LN | troch_LN | 57.0 |
| 173 | GW_Lane5_LN10 | GW_Lane5_LN10 | LN | troch_LN | troch_LN | 64.0 |
| 174 | GW_Lane5_LN11 | GW_Lane5_LN11 | LN | troch_LN | troch_LN | 65.0 |
| 175 | GW_Lane5_LN12 | GW_Lane5_LN12 | LN | troch_LN | troch_LN | 66.0 |
| 176 | GW_Lane5_LN14 | GW_Lane5_LN14 | LN | troch_LN | troch_LN | 67.0 |
| 177 | GW_Lane5_LN16 | GW_Lane5_LN16 | LN | troch_LN | troch_LN | 68.0 |
| 178 | GW_Lane5_LN18 | GW_Lane5_LN18 | LN | troch_LN | troch_LN | 69.0 |
| 179 | GW_Lane5_LN19 | GW_Lane5_LN19 | LN | troch_LN | troch_LN | 70.0 |
| 180 | GW_Lane5_LN2 | GW_Lane5_LN2 | LN_rep | troch_LN_rep | troch_LN_rep | 58.0 |
| 181 | GW_Lane5_LN20 | GW_Lane5_LN20 | LN | troch_LN | troch_LN | 71.0 |
| 182 | GW_Lane5_LN3 | GW_Lane5_LN3 | LN | troch_LN | troch_LN | 59.0 |
| 183 | GW_Lane5_LN4 | GW_Lane5_LN4 | LN | troch_LN | troch_LN | 60.0 |
| 184 | GW_Lane5_LN6 | GW_Lane5_LN6 | LN | troch_LN | troch_LN | 61.0 |
| 185 | GW_Lane5_LN7 | GW_Lane5_LN7 | LN | troch_LN | troch_LN | 62.0 |
| 186 | GW_Lane5_LN8 | GW_Lane5_LN8 | LN | troch_LN | troch_LN | 63.0 |
| 187 | GW_Lane5_MN1 | GW_Lane5_MN1 | MN | troch_MN | troch_west | 51.0 |
| 188 | GW_Lane5_MN12 | GW_Lane5_MN12 | MN | troch_MN | troch_west | 56.0 |
| 189 | GW_Lane5_MN3 | GW_Lane5_MN3 | MN | troch_MN | troch_west | 52.0 |
| 190 | GW_Lane5_MN5 | GW_Lane5_MN5 | MN | troch_MN | troch_west | 53.0 |
| 191 | GW_Lane5_MN8 | GW_Lane5_MN8 | MN | troch_MN | troch_west | 54.0 |
| 192 | GW_Lane5_MN9 | GW_Lane5_MN9 | MN | troch_MN | troch_west | 55.0 |
| 193 | GW_Lane5_NA1 | GW_Lane5_NA1 | NR | lud_PK | lud_PK | 39.2 |
| 194 | GW_Lane5_NA3-3ul | GW_Lane5_NA3-3ul | NR | lud_PK | lud_PK | 39.2 |
| 195 | GW_Lane5_PT11 | GW_Lane5_PT11 | KL | lud_KL | lud_central | 42.0 |
| 196 | GW_Lane5_PT12 | GW_Lane5_PT12 | KL | lud_KL | lud_central | 42.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 197 | GW_Lane5_PT2 | GW_Lane5_PT2 | ML | lud_ML | lud_ML | 51.0 |
| 198 | GW_Lane5_PT3 | GW_Lane5_PT3 | PA | lud_PA | lud_central | 46.0 |
| 199 | GW_Lane5_PT4 | GW_Lane5_PT4 | PA | lud_PA | lud_central | 46.0 |
| 200 | GW_Lane5_PT6 | GW_Lane5_PT6 | KL | lud_KL | lud_central | 42.0 |
| 201 | GW_Lane5_SH1 | GW_Lane5_SH1 | SH | lud_PK | lud_PK | 39.1 |
| 202 | GW_Lane5_SH2 | GW_Lane5_SH2 | SH | lud_PK | lud_PK | 39.1 |
| 203 | GW_Lane5_SH4 | GW_Lane5_SH4 | SH | lud_PK | lud_PK | 39.1 |
| 204 | GW_Lane5_SH5 | GW_Lane5_SH5 | SH | lud_PK | lud_PK | 39.1 |
| 205 | GW_Lane5_SL1 | GW_Lane5_SL1 | SL | plumb | plumb | 150.0 |
| 206 | GW_Lane5_SL2 | GW_Lane5_SL2 | SL | plumb | plumb | 151.0 |
| 207 | GW_Lane5_ST1 | GW_Lane5_ST1 | ST | plumb | plumb | 142.0 |
| 208 | GW_Lane5_ST12 | GW_Lane5_ST12 | ST | plumb | plumb | 144.0 |
| 209 | GW_Lane5_ST3 | GW_Lane5_ST3 | ST | plumb | plumb | 143.0 |
| 210 | GW_Lane5_STvi1 | GW_Lane5_STvi1 | ST_vi | vir | vir | 22.0 |
| 211 | GW_Lane5_STvi2 | GW_Lane5_STvi2 | ST_vi | vir | vir | 23.0 |
| 212 | GW_Lane5_STvi3 | GW_Lane5_STvi3 | ST_vi | vir | vir | 24.0 |
| 213 | GW_Lane5_TA1 | GW_Lane5_TA1 | TA | plumb | plumb | 86.0 |
| 214 | GW_Lane5_TL1 | GW_Lane5_TL1 | TL | vir | vir | 9.0 |
| 215 | GW_Lane5_TL10 | GW_Lane5_TL10 | TL | vir | vir | 17.0 |
| 216 | GW_Lane5_TL11 | GW_Lane5_TL11 | TL | vir | vir | 18.0 |
| 217 | GW_Lane5_TL12 | GW_Lane5_TL12 | TL | vir | vir | 19.0 |
| 218 | GW_Lane5_TL2 | GW_Lane5_TL2 | TL | vir | vir | 10.0 |
| 219 | GW_Lane5_TL3 | GW_Lane5_TL3 | TL_rep | vir_rep | vir_rep | 11.0 |
| 220 | GW_Lane5_TL4 | GW_Lane5_TL4 | TL | vir | vir | 12.0 |
| 221 | GW_Lane5_TL5 | GW_Lane5_TL5 | TL | vir | vir | 13.0 |
| 222 | GW_Lane5_TL7 | GW_Lane5_TL7 | TL | vir | vir | 14.0 |
| 223 | GW_Lane5_TL8 | GW_Lane5_TL8 | TL | vir | vir | 15.0 |
| 224 | GW_Lane5_TL9 | GW_Lane5_TL9 | TL | vir | vir | 16.0 |
| 225 | GW_Lane5_TU1 | GW_Lane5_TU1 | TU | nit | nit | 35.0 |
| 226 | GW_Lane5_TU2 | GW_Lane5_TU2 | TU | nit | nit | 36.0 |
| 227 | GW_Lane5_UY1 | GW_Lane5_UY1 | UY_rep | plumb_rep | plumb_rep | 93.0 |
| 228 | GW_Lane5_UY2 | GW_Lane5_UY2 | UY | plumb | plumb | 88.0 |
| 229 | GW_Lane5_UY3 | GW_Lane5_UY3 | UY | plumb | plumb | 89.0 |
| 230 | GW_Lane5_UY4 | GW_Lane5_UY4 | UY | plumb | plumb | 90.0 |
| 231 | GW_Lane5_UY5 | GW_Lane5_UY5 | UY | plumb | plumb | 91.0 |
| 232 | GW_Lane5_UY6 | GW_Lane5_UY6 | UY | plumb | plumb | 92.0 |
| 233 | GW_Lane5_YK1 | GW_Lane5_YK1 | YK | vir | vir | 1.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 234 | GW_Lane5_YK11 | GW_Lane5_YK11 | YK | vir | vir | 8.0 |
| 235 | GW_Lane5_YK3 | GW_Lane5_YK3 | YK | vir | vir | 2.0 |
| 236 | GW_Lane5_YK4 | GW_Lane5_YK4 | YK | vir | vir | 3.0 |
| 237 | GW_Lane5_YK5 | GW_Lane5_YK5 | YK | vir | vir | 4.0 |
| 238 | GW_Lane5_YK6 | GW_Lane5_YK6 | YK | vir | vir | 5.0 |
| 239 | GW_Lane5_YK7 | GW_Lane5_YK7 | YK | vir | vir | 6.0 |
| 240 | GW_Lane5_YK9 | GW_Lane5_YK9 | YK | vir | vir | 7.0 |
| 241 | GW_Liz_GBS_Liz10045 | GW_Liz_GBS_Liz10045 | ML | lud | lud_ML | 51.01 |
| 242 | GW_Liz_GBS_Liz10094 | GW_Liz_GBS_Liz10094 | ML | lud | lud_ML | 51.02 |
| 243 | GW_Liz_GBS_Liz5101 | GW_Liz_GBS_Liz5101 | ML | lud | lud_ML | 51.03 |
| 244 | GW_Liz_GBS_Liz5101_R | GW_Liz_GBS_Liz5101_R | ML_rep | lud_rep | lud_ML_rep | 51. |
| 245 | GW_Liz_GBS_Liz5118 | GW_Liz_GBS_Liz5118 | ML | lud | lud_ML | 51.05 |
| 246 | GW_Liz_GBS_Liz5139 | GW_Liz_GBS_Liz5139 | ML | lud | lud_ML | 51.06 |
| 247 | GW_Liz_GBS_Liz5142 | GW_Liz_GBS_Liz5142 | ML | lud | lud_ML | 51.07 |
| 248 | GW_Liz_GBS_Liz5144 | GW_Liz_GBS_Liz5144 | ML | lud | lud_ML | 51.08 |
| 249 | GW_Liz_GBS_Liz5150 | GW_Liz_GBS_Liz5150 | ML | lud | lud_ML | 51.09 |
| 250 | GW_Liz_GBS_Liz5159 | GW_Liz_GBS_Liz5159 | ML | lud_chick | lud_ML | 51.1 |
| 251 | GW_Liz_GBS_Liz5162 | GW_Liz_GBS_Liz5162 | ML | lud_chick | lud_ML | 51.11 |
| 252 | GW_Liz_GBS_Liz5163 | GW_Liz_GBS_Liz5163 | ML | lud_chick | lud_ML | 51.12 |
| 253 | GW_Liz_GBS_Liz5164 | GW_Liz_GBS_Liz5164 | ML | lud_chick | lud_ML | 51.13 |
| 254 | GW_Liz_GBS_Liz5165 | GW_Liz_GBS_Liz5165 | ML | lud | lud_ML | 51.14 |
| 255 | GW_Liz_GBS_Liz5167 | GW_Liz_GBS_Liz5167 | ML | lud_chick | lud_ML | 51.15 |
| 256 | GW_Liz_GBS_Liz5168 | GW_Liz_GBS_Liz5168 | ML | lud_chick | lud_ML | 51.16 |
| 257 | GW_Liz_GBS_Liz5169 | GW_Liz_GBS_Liz5169 | ML | lud_chick | lud_ML | 51.17 |
| 258 | GW_Liz_GBS_Liz5171 | GW_Liz_GBS_Liz5171 | ML | lud | lud_ML | 51.18 |
| 259 | GW_Liz_GBS_Liz5172 | GW_Liz_GBS_Liz5172 | ML | lud_chick | lud_ML | 51.19 |
| 260 | GW_Liz_GBS_Liz5173 | GW_Liz_GBS_Liz5173 | ML | lud_chick | lud_ML | 51.2 |
| 261 | GW_Liz_GBS_Liz5174 | GW_Liz_GBS_Liz5174 | ML | lud | lud_ML | 51.21 |
| 262 | GW_Liz_GBS_Liz5175 | GW_Liz_GBS_Liz5175 | ML | lud | lud_ML | 51.22 |
| 263 | GW_Liz_GBS_Liz5176 | GW_Liz_GBS_Liz5176 | ML | lud | lud_ML | 51.23 |
| 264 | GW_Liz_GBS_Liz5177 | GW_Liz_GBS_Liz5177 | ML | lud_chick | lud_ML | 51.24 |
| 265 | GW_Liz_GBS_Liz5178 | GW_Liz_GBS_Liz5178 | ML | lud_chick | lud_ML | 51.25 |
| 266 | GW_Liz_GBS_Liz5179 | GW_Liz_GBS_Liz5179 | ML | lud_chick | lud_ML | 51.26 |
| 267 | GW_Liz_GBS_Liz5180 | GW_Liz_GBS_Liz5180 | ML | lud | lud_ML | 51.27 |
| 268 | GW_Liz_GBS_Liz5182 | GW_Liz_GBS_Liz5182 | ML | lud_chick | lud_ML | 51.28 |
| 269 | GW_Liz_GBS_Liz5184 | GW_Liz_GBS_Liz5184 | ML | lud_chick | lud_ML | 51.29 |
| 270 | GW_Liz_GBS_Liz5185 | GW_Liz_GBS_Liz5185 | ML | lud | lud_ML | 51.3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 271 | GW_Liz_GBS_Liz5186 | GW_Liz_GBS_Liz5186 | ML | lud_chick | lud_ML | 51.31 |
| 272 | GW_Liz_GBS_Liz5187 | GW_Liz_GBS_Liz5187 | ML | lud_chick | lud_ML | 51.32 |
| 273 | GW_Liz_GBS_Liz5188 | GW_Liz_GBS_Liz5188 | ML | lud | lud_ML | 51.33 |
| 274 | GW_Liz_GBS_Liz5189 | GW_Liz_GBS_Liz5189 | ML | lud_chick | lud_ML | 51.34 |
| 275 | GW_Liz_GBS_Liz5190 | GW_Liz_GBS_Liz5190 | ML | lud_chick | lud_ML | 51.35 |
| 276 | GW_Liz_GBS_Liz5191 | GW_Liz_GBS_Liz5191 | ML | lud_chick | lud_ML | 51.36 |
| 277 | GW_Liz_GBS_Liz5192 | GW_Liz_GBS_Liz5192 | ML | lud_chick | lud_ML | 51.37 |
| 278 | GW_Liz_GBS_Liz5193 | GW_Liz_GBS_Liz5193 | ML | lud_chick | lud_ML | 51.38 |
| 279 | GW_Liz_GBS_Liz5194 | GW_Liz_GBS_Liz5194 | ML | lud_chick | lud_ML | 51.39 |
| 280 | GW_Liz_GBS_Liz5195 | GW_Liz_GBS_Liz5195 | ML | lud | lud_ML | 51.4 |
| 281 | GW_Liz_GBS_Liz5197 | GW_Liz_GBS_Liz5197 | ML | lud | lud_ML | 51.41 |
| 282 | GW_Liz_GBS_Liz5199 | GW_Liz_GBS_Liz5199 | ML | lud_chick | lud_ML | 51.42 |
| 283 | GW_Liz_GBS_Liz6002 | GW_Liz_GBS_Liz6002 | ML | lud | lud_ML | 51.43 |
| 284 | GW_Liz_GBS_Liz6006 | GW_Liz_GBS_Liz6006 | ML | lud | lud_ML | 51.44 |
| 285 | GW_Liz_GBS_Liz6008 | GW_Liz_GBS_Liz6008 | ML | lud | lud_ML | 51.45 |
| 286 | GW_Liz_GBS_Liz6009 | GW_Liz_GBS_Liz6009 | ML | lud | lud_ML | 51.46 |
| 287 | GW_Liz_GBS_Liz6010 | GW_Liz_GBS_Liz6010 | ML | lud | lud_ML | 51.47 |
| 288 | GW_Liz_GBS_Liz6012 | GW_Liz_GBS_Liz6012 | ML | lud | lud_ML | 51.48 |
| 289 | GW_Liz_GBS_Liz6014 | GW_Liz_GBS_Liz6014 | ML | lud | lud_ML | 51.49 |
| 290 | GW_Liz_GBS_Liz6055 | GW_Liz_GBS_Liz6055 | ML | lud | lud_ML | 51.5 |
| 291 | GW_Liz_GBS_Liz6057 | GW_Liz_GBS_Liz6057 | ML | lud | lud_ML | 51.51 |
| 292 | GW_Liz_GBS_Liz6060 | GW_Liz_GBS_Liz6060 | ML | lud | lud_ML | 51.52 |
| 293 | GW_Liz_GBS_Liz6062 | GW_Liz_GBS_Liz6062 | ML | lud | lud_ML | 51.53 |
| 294 | GW_Liz_GBS_Liz6063 | GW_Liz_GBS_Liz6063 | ML | lud | lud_ML | 51.54 |
| 295 | GW_Liz_GBS_Liz6066 | GW_Liz_GBS_Liz6066 | ML | lud | lud_ML | 51.55 |
| 296 | GW_Liz_GBS_Liz6072 | GW_Liz_GBS_Liz6072 | ML | lud | lud_ML | 51.56 |
| 297 | GW_Liz_GBS_Liz6079 | GW_Liz_GBS_Liz6079 | ML | lud | lud_ML | 51.57 |
| 298 | GW_Liz_GBS_Liz6203 | GW_Liz_GBS_Liz6203 | ML | lud_chick | lud_ML | 51.58 |
| 299 | GW_Liz_GBS_Liz6204 | GW_Liz_GBS_Liz6204 | ML | lud_chick | lud_ML | 51.59 |
| 300 | GW_Liz_GBS_Liz6461 | GW_Liz_GBS_Liz6461 | ML | lud | lud_ML | 51.6 |
| 301 | GW_Liz_GBS_Liz6472 | GW_Liz_GBS_Liz6472 | ML | lud | lud_ML | 51.61 |
| 302 | GW_Liz_GBS_Liz6478 | GW_Liz_GBS_Liz6478 | ML | lud | lud_ML | 51.62 |
| 303 | GW_Liz_GBS_Liz6766 | GW_Liz_GBS_Liz6766 | ML | lud | lud_ML | 51.63 |
| 304 | GW_Liz_GBS_Liz6776 | GW_Liz_GBS_Liz6776 | ML | lud | lud_ML | 51.64 |
| 305 | GW_Liz_GBS_Liz6794 | GW_Liz_GBS_Liz6794 | ML | lud | lud_ML | 51.65 |
| 306 | GW_Liz_GBS_P_fusc | GW_Liz_GBS_P_fusc | fusc | fusc | fusc | 201.0 |
| 307 | GW_Liz_GBS_P_h_man | GW_Liz_GBS_P_h_man | hmand | hmand | hmand | 202.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 308 | GW_Liz_GBS_P_humei | GW_Liz_GBS_P_humei | hume | hume | hume | 203.0 |
| 309 | GW_Liz_GBS_P_inor | GW_Liz_GBS_P_inor | inor | inor | inor | 204.0 |
| 310 | GW_Liz_GBS_S_burk | GW_Liz_GBS_S_burk | burk | burk | burk | 205.0 |

GOOD NEWS: names of individuals in metadata file and genotype ind file match perfectly.

## Filtering

### Filter out duplicate runs (indicated with `_rep` in `Fst_group` column)

```
    selection = occursin.("_rep", ind_with_metadata.Fst_group)
    println("""Filtering out these runs because they are duplicates of another,
    according to having "rep" in Fst_group: """)
    display(ind_with_metadata.ind[selection])
    ind_with_metadata_indFiltered = ind_with_metadata[Not(selection), :];
    geno_indFiltered = view(geno, Not(selection), :);  # use of view() avoids copying large memory obje
```

```
Filtering out these runs because they are duplicates of another,
according to having "rep" in Fst_group:
```

```
11-element Vector{String}:
 "GW_Armando_plate1_TTGW05_rep1"
 "GW_Armando_plate2_IL2"
 "GW_Armando_plate2_LN11"
 "GW_Armando_plate2_TTGW05_rep3"
 "GW_Armando_plate2_TTGW05_rep4"
 "GW_Lane5_AB1"
 "GW_Lane5_IL2"
 "GW_Lane5_LN2"
 "GW_Lane5_TL3"
 "GW_Lane5_UY1"
 "GW_Liz_GBS_Liz5101_R"
```

### Filter specific individuals

If there are certain individuals that we want to filter out prior to any additional analysis, we can do so here by setting filter to `true` and specifying the individual row numbers in `filter_out_inds`:

```
filter = true
# Specify individuals to filter out:
filter_out_inds = ["GW_Liz_GBS_P_fusc", "GW_Liz_GBS_P_h_man", "GW_Liz_GBS_P_humei", "GW_Liz_GBS_P_inor"
if filter
    selection = map(in(filter_out_inds), ind_with_metadata_indFiltered.ind)
    filtered_out = ind_with_metadata_indFiltered.ind[selection]
    ind_with_metadata_indFiltered = ind_with_metadata_indFiltered[Not(selection), :]
    geno_indFiltered = view(geno_indFiltered, Not(selection), :)
    println("Specific individuals filtered out as requested: ")
    display(filtered_out)
else
    println("No specific individuals filtered (because filter not true)")
end
```

```
Specific individuals filtered out as requested:


5-element Vector{String}:
 "GW_Liz_GBS_P_fusc"
 "GW_Liz_GBS_P_h_man"
 "GW_Liz_GBS_P_humei"
 "GW_Liz_GBS_P_inor"
 "GW_Liz_GBS_S_burk"
```

### Filter individuals based on missing genotypes

Here we determine number of missing SNPs per individual (40% for this round), and filter out those individual datasets with more than a certain percent of missing SNPs:

```
SNPmissing_percent_allowed_per_ind = 40   # this is the percentage threshold
threshold_missing = loci_count * SNPmissing_percent_allowed_per_ind/100
numMissings = sum(geno_indFiltered .== -1, dims=2)
ind_with_metadata_indFiltered.numMissings .= numMissings
selection = vec(numMissings .<= threshold_missing) # the vec command converts to BitVector rather than
println("Filtering out these individuals based on too many missing genotypes: ")
filtered_inds = ind_with_metadata_indFiltered.ind[selection.==false]
println(DataFrame(filtered_inds = filtered_inds)) # did this to print all lines
ind_with_metadata_indFiltered = ind_with_metadata_indFiltered[selection, :]
geno_indFiltered = view(geno_indFiltered, selection, :);
println()
println("Here are the remaining individuals: ")
println(DataFrame(ind_with_metadata_indFiltered))
```

Filtering out these individuals based on too many missing genotypes:

33×1 DataFrame

| Row | filtered_inds |
|---|---|
| | String |
| 1 | GW_Armando_plate1_JG08G02 |
| 2 | GW_Armando_plate1_JG10G01 |
| 3 | GW_Armando_plate1_NO_BC_TTGW05 |
| 4 | GW_Armando_plate1_NO_DNA |
| 5 | GW_Armando_plate1_TTGW21 |
| 6 | GW_Armando_plate1_TTGW71 |
| 7 | GW_Armando_plate2_NO_BC_TTGW05 |
| 8 | GW_Armando_plate2_NO_DNA |
| 9 | GW_Armando_plate2_TTGW15 |
| 10 | GW_Lane5_AA10 |
| 11 | GW_Lane5_DA6 |
| 12 | GW_Lane5_LN11 |
| 13 | GW_Liz_GBS_Liz5101 |
| 14 | GW_Liz_GBS_Liz5118 |
| 15 | GW_Liz_GBS_Liz5139 |
| 16 | GW_Liz_GBS_Liz5142 |
| 17 | GW_Liz_GBS_Liz5150 |
| 18 | GW_Liz_GBS_Liz5159 |
| 19 | GW_Liz_GBS_Liz5162 |
| 20 | GW_Liz_GBS_Liz5169 |
| 21 | GW_Liz_GBS_Liz5171 |
| 22 | GW_Liz_GBS_Liz5172 |
| 23 | GW_Liz_GBS_Liz5174 |
| 24 | GW_Liz_GBS_Liz5176 |
| 25 | GW_Liz_GBS_Liz5177 |
| 26 | GW_Liz_GBS_Liz5180 |
| 27 | GW_Liz_GBS_Liz5186 |
| 28 | GW_Liz_GBS_Liz5187 |
| 29 | GW_Liz_GBS_Liz5192 |
| 30 | GW_Liz_GBS_Liz5195 |
| 31 | GW_Liz_GBS_Liz6012 |
| 32 | GW_Liz_GBS_Liz6203 |

```
 33 │ GW_Liz_GBS_Liz6766
```

Here are the remaining individuals:

261×7 DataFrame

| Row | ind | ID | location | group | Fst_group | plot_order | numMissing |
|---|---|---|---|---|---|---|---|
| | String | String31 | String7 | String15 | String15 | Float64 | Int64 |
| 1 | GW_Armando_plate1_AB1 | GW_Armando_plate1_AB1 | AB | vir | vir | 20.01 | |
| 2 | GW_Armando_plate1_JF07G02 | GW_Armando_plate1_JF07G02 | ST | plumb | plumb | 108.0 | |
| 3 | GW_Armando_plate1_JF07G03 | GW_Armando_plate1_JF07G03 | ST | plumb | plumb | 109.0 | |
| 4 | GW_Armando_plate1_JF07G04 | GW_Armando_plate1_JF07G04 | ST | plumb | plumb | 110.0 | |
| 5 | GW_Armando_plate1_JF08G02 | GW_Armando_plate1_JF08G02 | ST | plumb | plumb | 111.0 | |
| 6 | GW_Armando_plate1_JF09G01 | GW_Armando_plate1_JF09G01 | ST | plumb | plumb | 112.0 | |
| 7 | GW_Armando_plate1_JF09G02 | GW_Armando_plate1_JF09G02 | ST | plumb | plumb | 113.0 | |
| 8 | GW_Armando_plate1_JF10G03 | GW_Armando_plate1_JF10G03 | ST | plumb_vir | plumb_vir | 170.0 | |
| 9 | GW_Armando_plate1_JF11G01 | GW_Armando_plate1_JF11G01 | ST | plumb | plumb | 114.0 | |
| 10 | GW_Armando_plate1_JF12G01 | GW_Armando_plate1_JF12G01 | ST | plumb | plumb | 115.0 | |
| 11 | GW_Armando_plate1_JF12G02 | GW_Armando_plate1_JF12G02 | ST | plumb | plumb | 116.0 | |
| 12 | GW_Armando_plate1_JF12G04 | GW_Armando_plate1_JF12G04 | ST_vi | vir | vir | 24.001 | |
| 13 | GW_Armando_plate1_JF13G01 | GW_Armando_plate1_JF13G01 | ST | plumb | plumb | 117.0 | |
| 14 | GW_Armando_plate1_JF15G03 | GW_Armando_plate1_JF15G03 | DV | plumb | plumb | 103.0 | |
| 15 | GW_Armando_plate1_JF16G01 | GW_Armando_plate1_JF16G01 | DV_vi | plumb_vir | vir | 24.04 | |
| 16 | GW_Armando_plate1_JF20G01 | GW_Armando_plate1_JF20G01 | MB | plumb | plumb | 94.0 | |
| 17 | GW_Armando_plate1_JF22G01 | GW_Armando_plate1_JF22G01 | MB | plumb | plumb | 95.0 | |
| 18 | GW_Armando_plate1_JF23G01 | GW_Armando_plate1_JF23G01 | VB | plumb | plumb | 98.0 | |
| 19 | GW_Armando_plate1_JF23G02 | GW_Armando_plate1_JF23G02 | VB | plumb | plumb | 99.0 | |
| 20 | GW_Armando_plate1_JF24G02 | GW_Armando_plate1_JF24G02 | VB | plumb | plumb | 100.0 | |
| 21 | GW_Armando_plate1_JF26G01 | GW_Armando_plate1_JF26G01 | ST | plumb | plumb | 118.0 | |
| 22 | GW_Armando_plate1_JF27G01 | GW_Armando_plate1_JF27G01 | ST | plumb | plumb | 119.0 | |
| 23 | GW_Armando_plate1_JF29G01 | GW_Armando_plate1_JF29G01 | ST | plumb | plumb | 120.0 | |
| 24 | GW_Armando_plate1_JF29G02 | GW_Armando_plate1_JF29G02 | ST | plumb | plumb | 121.0 | |
| 25 | GW_Armando_plate1_JF29G03 | GW_Armando_plate1_JF29G03 | ST | plumb | plumb | 122.0 | |
| 26 | GW_Armando_plate1_JG02G02 | GW_Armando_plate1_JG02G02 | PR | plumb | plumb | 145.0 | |
| 27 | GW_Armando_plate1_JG02G04 | GW_Armando_plate1_JG02G04 | PR | plumb | plumb | 146.0 | |
| 28 | GW_Armando_plate1_JG08G01 | GW_Armando_plate1_JG08G01 | ST | plumb | plumb | 123.0 | |
| 29 | GW_Armando_plate1_JG12G01 | GW_Armando_plate1_JG12G01 | ST | plumb | plumb | 126.0 | |
| 30 | GW_Armando_plate1_JG17G01 | GW_Armando_plate1_JG17G01 | ST | plumb_vir | plumb | 127.0 | |
| 31 | GW_Armando_plate1_RF20G01 | GW_Armando_plate1_RF20G01 | BJ | obs_plumb | plumb_BJ | 77.50 | |

```
GW_Armando_plate1_RF29G02    GW_Armando_plate1_RF29G02    BJ     obs_plumb     plumb_BJ      77.50
GW_Armando_plate1_TL3        GW_Armando_plate1_TL3        TL     vir           vir           11.01     1
GW_Armando_plate1_TTGW01     GW_Armando_plate1_TTGW01     MN     troch_MN      troch_west    53.0
GW_Armando_plate1_TTGW05_rep2 GW_Armando_plate1_TTGW05_rep2 MN     troch_MN      troch_west    53
GW_Armando_plate1_TTGW06     GW_Armando_plate1_TTGW06     SU     lud_Sukhto    lud_central   47.0
GW_Armando_plate1_TTGW07     GW_Armando_plate1_TTGW07     SU     lud_Sukhto    lud_central   47.0
GW_Armando_plate1_TTGW10     GW_Armando_plate1_TTGW10     SU     lud_Sukhto    lud_central   47.0
GW_Armando_plate1_TTGW11     GW_Armando_plate1_TTGW11     SU     lud_Sukhto    lud_central   47.0
GW_Armando_plate1_TTGW13     GW_Armando_plate1_TTGW13     TH     lud_Thallighar lud_central   43.
GW_Armando_plate1_TTGW17     GW_Armando_plate1_TTGW17     TH     lud_Thallighar lud_central   43.
GW_Armando_plate1_TTGW19     GW_Armando_plate1_TTGW19     TH     lud_Thallighar lud_central   43.
GW_Armando_plate1_TTGW22     GW_Armando_plate1_TTGW22     SR     lud_Sural     lud_central   45.0
GW_Armando_plate1_TTGW23     GW_Armando_plate1_TTGW23     SR     lud_Sural     lud_central   45.0
GW_Armando_plate1_TTGW29     GW_Armando_plate1_TTGW29     SR     lud_Sural     lud_central   45.0
GW_Armando_plate1_TTGW52     GW_Armando_plate1_TTGW52     NG     lud_Nainaghar  lud_central   49.
GW_Armando_plate1_TTGW53     GW_Armando_plate1_TTGW53     NG     lud_Nainaghar  lud_central   49.
GW_Armando_plate1_TTGW55     GW_Armando_plate1_TTGW55     NG     lud_Nainaghar  lud_central   49.
GW_Armando_plate1_TTGW57     GW_Armando_plate1_TTGW57     NG     lud_Nainaghar  lud_central   49.
GW_Armando_plate1_TTGW58     GW_Armando_plate1_TTGW58     NG     lud_Nainaghar  lud_central   49.
GW_Armando_plate1_TTGW59     GW_Armando_plate1_TTGW59     NG     lud_Nainaghar  lud_central   49.
GW_Armando_plate1_TTGW63     GW_Armando_plate1_TTGW63     SP     lud_Spiti     troch_west    55.0
GW_Armando_plate1_TTGW64     GW_Armando_plate1_TTGW64     SP     lud_Spiti     troch_west    55.0
GW_Armando_plate1_TTGW65     GW_Armando_plate1_TTGW65     SP     lud_Spiti     troch_west    55.0
GW_Armando_plate1_TTGW66     GW_Armando_plate1_TTGW66     SP     lud_Spiti     troch_west    55.0
GW_Armando_plate1_TTGW68     GW_Armando_plate1_TTGW68     SP     lud_Spiti     troch_west    55.0
GW_Armando_plate1_TTGW70     GW_Armando_plate1_TTGW70     SA     lud_Sathrundi  lud_Sath     41.0
GW_Armando_plate1_TTGW72     GW_Armando_plate1_TTGW72     SA     lud_Sathrundi  lud_Sath     41.0
GW_Armando_plate1_TTGW74     GW_Armando_plate1_TTGW74     SA     lud_Sathrundi  lud_Sath     41.0
GW_Armando_plate1_TTGW78     GW_Armando_plate1_TTGW78     SA     lud_Sathrundi  lud_Sath     41.0
GW_Armando_plate1_TTGW_15_05  GW_Armando_plate1_TTGW_15_05  SR     lud_Sural     lud_central   45
GW_Armando_plate1_TTGW_15_07  GW_Armando_plate1_TTGW_15_07  SR     lud_Sural     lud_central   45
GW_Armando_plate1_TTGW_15_08  GW_Armando_plate1_TTGW_15_08  SR     lud_Sural     lud_central   45
GW_Armando_plate1_TTGW_15_09  GW_Armando_plate1_TTGW_15_09  SR     lud_Sural     lud_central   45
GW_Armando_plate1_UY1        GW_Armando_plate1_UY1        UY     plumb         plumb         87.0
GW_Armando_plate2_JE31G01    GW_Armando_plate2_JE31G01    VB_vi  vir_misID     vir           24.00
GW_Armando_plate2_JF03G01    GW_Armando_plate2_JF03G01    ST_vi  vir_misID     vir           24.00
GW_Armando_plate2_JF03G02    GW_Armando_plate2_JF03G02    VB_vi  vir_misID     vir           24.00
```

| | | | | | | |
|---|---|---|---|---|---|---|
| 69 | GW_Armando_plate2_JF07G01 | GW_Armando_plate2_JF07G01 | ST | plumb | plumb | 128.0 |
| 70 | GW_Armando_plate2_JF08G04 | GW_Armando_plate2_JF08G04 | ST | plumb | plumb | 129.0 |
| 71 | GW_Armando_plate2_JF10G02 | GW_Armando_plate2_JF10G02 | ST | plumb | plumb | 130.0 |
| 72 | GW_Armando_plate2_JF11G02 | GW_Armando_plate2_JF11G02 | ST | plumb | plumb | 131.0 |
| 73 | GW_Armando_plate2_JF12G03 | GW_Armando_plate2_JF12G03 | ST | plumb | plumb | 132.0 |
| 74 | GW_Armando_plate2_JF12G05 | GW_Armando_plate2_JF12G05 | ST | plumb | plumb | 133.0 |
| 75 | GW_Armando_plate2_JF13G02 | GW_Armando_plate2_JF13G02 | ST | plumb | plumb | 134.0 |
| 76 | GW_Armando_plate2_JF14G01 | GW_Armando_plate2_JF14G01 | DV | plumb | plumb | 104.0 |
| 77 | GW_Armando_plate2_JF14G02 | GW_Armando_plate2_JF14G02 | DV | plumb | plumb | 105.0 |
| 78 | GW_Armando_plate2_JF15G01 | GW_Armando_plate2_JF15G01 | DV | plumb | plumb | 106.0 |
| 79 | GW_Armando_plate2_JF15G02 | GW_Armando_plate2_JF15G02 | DV | plumb | plumb | 107.0 |
| 80 | GW_Armando_plate2_JF16G02 | GW_Armando_plate2_JF16G02 | DV_vi | plumb_vir | vir | 24.04 |
| 81 | GW_Armando_plate2_JF19G01 | GW_Armando_plate2_JF19G01 | MB | plumb | plumb | 96.0 |
| 82 | GW_Armando_plate2_JF20G02 | GW_Armando_plate2_JF20G02 | MB | plumb | plumb | 97.0 |
| 83 | GW_Armando_plate2_JF24G01 | GW_Armando_plate2_JF24G01 | VB | plumb | plumb | 101.0 |
| 84 | GW_Armando_plate2_JF24G03 | GW_Armando_plate2_JF24G03 | ST | plumb | plumb | 135.0 |
| 85 | GW_Armando_plate2_JF25G01 | GW_Armando_plate2_JF25G01 | VB | plumb | plumb | 102.0 |
| 86 | GW_Armando_plate2_JF26G02 | GW_Armando_plate2_JF26G02 | ST | plumb | plumb | 136.0 |
| 87 | GW_Armando_plate2_JF27G02 | GW_Armando_plate2_JF27G02 | ST | plumb | plumb | 137.0 |
| 88 | GW_Armando_plate2_JF30G01 | GW_Armando_plate2_JF30G01 | ST_vi | vir_misID | vir | 24.00 |
| 89 | GW_Armando_plate2_JG01G01 | GW_Armando_plate2_JG01G01 | PR | plumb | plumb | 147.0 |
| 90 | GW_Armando_plate2_JG02G01 | GW_Armando_plate2_JG02G01 | PR | plumb | plumb | 148.0 |
| 91 | GW_Armando_plate2_JG02G03 | GW_Armando_plate2_JG02G03 | PR | plumb | plumb | 149.0 |
| 92 | GW_Armando_plate2_JG10G02 | GW_Armando_plate2_JG10G02 | ST | plumb | plumb | 138.0 |
| 93 | GW_Armando_plate2_JG10G03 | GW_Armando_plate2_JG10G03 | ST | plumb | plumb | 139.0 |
| 94 | GW_Armando_plate2_JG12G02 | GW_Armando_plate2_JG12G02 | ST | plumb | plumb | 140.0 |
| 95 | GW_Armando_plate2_JG12G03 | GW_Armando_plate2_JG12G03 | ST | plumb | plumb | 141.0 |
| 96 | GW_Armando_plate2_LN2 | GW_Armando_plate2_LN2 | LN | troch_LN | troch_LN | 58.01 |
| 97 | GW_Armando_plate2_RF29G01 | GW_Armando_plate2_RF29G01 | BJ | obs_plumb | plumb_BJ | 77.50 |
| 98 | GW_Armando_plate2_TTGW02 | GW_Armando_plate2_TTGW02 | MN | troch_MN | troch_west | 53.0 |
| 99 | GW_Armando_plate2_TTGW03 | GW_Armando_plate2_TTGW03 | MN | troch_MN | troch_west | 53.0 |
| 100 | GW_Armando_plate2_TTGW08 | GW_Armando_plate2_TTGW08 | SU | lud_Sukhto | lud_central | 47.0 |
| 101 | GW_Armando_plate2_TTGW09 | GW_Armando_plate2_TTGW09 | SU | lud_Sukhto | lud_central | 47.0 |
| 102 | GW_Armando_plate2_TTGW12 | GW_Armando_plate2_TTGW12 | TH | lud_Thallighar | lud_central | 43 |
| 103 | GW_Armando_plate2_TTGW14 | GW_Armando_plate2_TTGW14 | TH | lud_Thallighar | lud_central | 43 |
| 104 | GW_Armando_plate2_TTGW16 | GW_Armando_plate2_TTGW16 | TH | lud_Thallighar | lud_central | 43 |
| 105 | GW_Armando_plate2_TTGW18 | GW_Armando_plate2_TTGW18 | TH | lud_Thallighar | lud_central | 43 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 106 | GW_Armando_plate2_TTGW20 | GW_Armando_plate2_TTGW20 | SR | lud_Sural | lud_central | 45.0 | |
| 107 | GW_Armando_plate2_TTGW24 | GW_Armando_plate2_TTGW24 | SR | lud_Sural | lud_central | 45.0 | |
| 108 | GW_Armando_plate2_TTGW25 | GW_Armando_plate2_TTGW25 | SR | lud_Sural | lud_central | 45.0 | |
| 109 | GW_Armando_plate2_TTGW27 | GW_Armando_plate2_TTGW27 | SR | lud_Sural | lud_central | 45.0 | |
| 110 | GW_Armando_plate2_TTGW28 | GW_Armando_plate2_TTGW28 | SR | lud_Sural | lud_central | 45.0 | |
| 111 | GW_Armando_plate2_TTGW50 | GW_Armando_plate2_TTGW50 | NG | lud_Nainaghar | lud_central | 49. | |
| 112 | GW_Armando_plate2_TTGW51 | GW_Armando_plate2_TTGW51 | NG | lud_Nainaghar | lud_central | 49. | |
| 113 | GW_Armando_plate2_TTGW54 | GW_Armando_plate2_TTGW54 | NG | lud_Nainaghar | lud_central | 49. | |
| 114 | GW_Armando_plate2_TTGW56 | GW_Armando_plate2_TTGW56 | NG | lud_Nainaghar | lud_central | 49. | |
| 115 | GW_Armando_plate2_TTGW60 | GW_Armando_plate2_TTGW60 | SP | lud_Spiti | troch_west | 55.0 | |
| 116 | GW_Armando_plate2_TTGW61 | GW_Armando_plate2_TTGW61 | SP | lud_Spiti | troch_west | 55.0 | |
| 117 | GW_Armando_plate2_TTGW62 | GW_Armando_plate2_TTGW62 | SP | lud_Spiti | troch_west | 55.0 | |
| 118 | GW_Armando_plate2_TTGW67 | GW_Armando_plate2_TTGW67 | SP | lud_Spiti | troch_west | 55.0 | |
| 119 | GW_Armando_plate2_TTGW69 | GW_Armando_plate2_TTGW69 | SP | lud_Spiti | troch_west | 55.0 | |
| 120 | GW_Armando_plate2_TTGW73 | GW_Armando_plate2_TTGW73 | SA | lud_Sathrundi | lud_Sath | 41.0 | |
| 121 | GW_Armando_plate2_TTGW75 | GW_Armando_plate2_TTGW75 | SA | lud_Sathrundi | lud_Sath | 41.0 | |
| 122 | GW_Armando_plate2_TTGW77 | GW_Armando_plate2_TTGW77 | SA | lud_Sathrundi | lud_Sath | 41.0 | |
| 123 | GW_Armando_plate2_TTGW79 | GW_Armando_plate2_TTGW79 | SA | lud_Sathrundi | lud_Sath | 41.0 | |
| 124 | GW_Armando_plate2_TTGW80 | GW_Armando_plate2_TTGW80 | SA | lud_Sathrundi | lud_Sath | 41.0 | |
| 125 | GW_Armando_plate2_TTGW_15_01 | GW_Armando_plate2_TTGW_15_01 | SR | lud_Sural | lud_central | 4 | |
| 126 | GW_Armando_plate2_TTGW_15_02 | GW_Armando_plate2_TTGW_15_02 | SR | lud_Sural | lud_central | 4 | |
| 127 | GW_Armando_plate2_TTGW_15_03 | GW_Armando_plate2_TTGW_15_03 | SR | lud_Sural | lud_central | 4 | |
| 128 | GW_Armando_plate2_TTGW_15_04 | GW_Armando_plate2_TTGW_15_04 | SR | lud_Sural | lud_central | 4 | |
| 129 | GW_Armando_plate2_TTGW_15_06 | GW_Armando_plate2_TTGW_15_06 | SR | lud_Sural | lud_central | 4 | |
| 130 | GW_Armando_plate2_TTGW_15_10 | GW_Armando_plate2_TTGW_15_10 | SR | lud_Sural | lud_central | 4 | |
| 131 | GW_Lane5_AA1 | GW_Lane5_AA1 | AA | vir_S | vir_S | 25.0 | 71767 |
| 132 | GW_Lane5_AA11 | GW_Lane5_AA11 | AA | vir_S | vir_S | 34.0 | 73608 |
| 133 | GW_Lane5_AA3 | GW_Lane5_AA3 | AA | vir_S | vir_S | 26.0 | 66255 |
| 134 | GW_Lane5_AA4 | GW_Lane5_AA4 | AA | vir_S | vir_S | 27.0 | 66596 |
| 135 | GW_Lane5_AA5 | GW_Lane5_AA5 | AA | vir_S | vir_S | 28.0 | 71861 |
| 136 | GW_Lane5_AA6 | GW_Lane5_AA6 | AA | vir_S | vir_S | 29.0 | 77882 |
| 137 | GW_Lane5_AA7 | GW_Lane5_AA7 | AA | vir_S | vir_S | 30.0 | 72471 |
| 138 | GW_Lane5_AA8 | GW_Lane5_AA8 | AA | vir_S | vir_S | 31.0 | 92231 |
| 139 | GW_Lane5_AA9 | GW_Lane5_AA9 | AA | vir_S | vir_S | 32.0 | 79151 |
| 140 | GW_Lane5_AB2 | GW_Lane5_AB2 | AB | vir | vir | 21.0 | 915622 |
| 141 | GW_Lane5_AN1 | GW_Lane5_AN1 | AN | plumb | plumb | 80.0 | 71904 |
| 142 | GW_Lane5_AN2 | GW_Lane5_AN2 | AN | plumb | plumb | 81.0 | 68434 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 143 | GW_Lane5_BK2 | GW_Lane5_BK2 | BK | plumb | plumb | 78.0 | 66601 |
| 144 | GW_Lane5_BK3 | GW_Lane5_BK3 | BK | plumb | plumb | 79.0 | 66509 |
| 145 | GW_Lane5_DA2 | GW_Lane5_DA2 | XN | obs | obs | 73.0 | 655980 |
| 146 | GW_Lane5_DA3 | GW_Lane5_DA3 | XN | obs | obs | 74.0 | 696722 |
| 147 | GW_Lane5_DA4 | GW_Lane5_DA4 | XN | obs | obs | 75.0 | 634940 |
| 148 | GW_Lane5_DA7 | GW_Lane5_DA7 | XN | obs | obs | 77.0 | 733664 |
| 149 | GW_Lane5_EM1 | GW_Lane5_EM1 | EM | troch_EM | troch_EM | 72.0 | 711 |
| 150 | GW_Lane5_IL1 | GW_Lane5_IL1 | IL | plumb | plumb | 82.0 | 69803 |
| 151 | GW_Lane5_IL4 | GW_Lane5_IL4 | IL | plumb | plumb | 83.0 | 72605 |
| 152 | GW_Lane5_KS1 | GW_Lane5_KS1 | OV | lud_KS | lud_KS | 40.0 | 75164 |
| 153 | GW_Lane5_KS2 | GW_Lane5_KS2 | OV | lud_KS | lud_KS | 40.0 | 82024 |
| 154 | GW_Lane5_LN1 | GW_Lane5_LN1 | LN | troch_LN | troch_LN | 57.0 | 707 |
| 155 | GW_Lane5_LN10 | GW_Lane5_LN10 | LN | troch_LN | troch_LN | 64.0 | 783 |
| 156 | GW_Lane5_LN12 | GW_Lane5_LN12 | LN | troch_LN | troch_LN | 66.0 | 884 |
| 157 | GW_Lane5_LN14 | GW_Lane5_LN14 | LN | troch_LN | troch_LN | 67.0 | 753 |
| 158 | GW_Lane5_LN16 | GW_Lane5_LN16 | LN | troch_LN | troch_LN | 68.0 | 738 |
| 159 | GW_Lane5_LN18 | GW_Lane5_LN18 | LN | troch_LN | troch_LN | 69.0 | 711 |
| 160 | GW_Lane5_LN19 | GW_Lane5_LN19 | LN | troch_LN | troch_LN | 70.0 | 731 |
| 161 | GW_Lane5_LN20 | GW_Lane5_LN20 | LN | troch_LN | troch_LN | 71.0 | 738 |
| 162 | GW_Lane5_LN3 | GW_Lane5_LN3 | LN | troch_LN | troch_LN | 59.0 | 702 |
| 163 | GW_Lane5_LN4 | GW_Lane5_LN4 | LN | troch_LN | troch_LN | 60.0 | 683 |
| 164 | GW_Lane5_LN6 | GW_Lane5_LN6 | LN | troch_LN | troch_LN | 61.0 | 690 |
| 165 | GW_Lane5_LN7 | GW_Lane5_LN7 | LN | troch_LN | troch_LN | 62.0 | 758 |
| 166 | GW_Lane5_LN8 | GW_Lane5_LN8 | LN | troch_LN | troch_LN | 63.0 | 662 |
| 167 | GW_Lane5_MN1 | GW_Lane5_MN1 | MN | troch_MN | troch_west | 51.0 | 943 |
| 168 | GW_Lane5_MN12 | GW_Lane5_MN12 | MN | troch_MN | troch_west | 56.0 | 67 |
| 169 | GW_Lane5_MN3 | GW_Lane5_MN3 | MN | troch_MN | troch_west | 52.0 | 940 |
| 170 | GW_Lane5_MN5 | GW_Lane5_MN5 | MN | troch_MN | troch_west | 53.0 | 752 |
| 171 | GW_Lane5_MN8 | GW_Lane5_MN8 | MN | troch_MN | troch_west | 54.0 | 771 |
| 172 | GW_Lane5_MN9 | GW_Lane5_MN9 | MN | troch_MN | troch_west | 55.0 | 897 |
| 173 | GW_Lane5_NA1 | GW_Lane5_NA1 | NR | lud_PK | lud_PK | 39.2 | 91921 |
| 174 | GW_Lane5_NA3-3ul | GW_Lane5_NA3-3ul | NR | lud_PK | lud_PK | 39.2 | 79 |
| 175 | GW_Lane5_PT11 | GW_Lane5_PT11 | KL | lud_KL | lud_central | 42.0 | 77 |
| 176 | GW_Lane5_PT12 | GW_Lane5_PT12 | KL | lud_KL | lud_central | 42.0 | 79 |
| 177 | GW_Lane5_PT2 | GW_Lane5_PT2 | ML | lud_ML | lud_ML | 51.0 | 76034 |
| 178 | GW_Lane5_PT3 | GW_Lane5_PT3 | PA | lud_PA | lud_central | 46.0 | 722 |
| 179 | GW_Lane5_PT4 | GW_Lane5_PT4 | PA | lud_PA | lud_central | 46.0 | 705 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 180 | GW_Lane5_PT6 | GW_Lane5_PT6 | KL | lud_KL | lud_central | 42.0 | 763 |
| 181 | GW_Lane5_SH1 | GW_Lane5_SH1 | SH | lud_PK | lud_PK | 39.1 | 96676 |
| 182 | GW_Lane5_SH2 | GW_Lane5_SH2 | SH | lud_PK | lud_PK | 39.1 | 76864 |
| 183 | GW_Lane5_SH4 | GW_Lane5_SH4 | SH | lud_PK | lud_PK | 39.1 | 84964 |
| 184 | GW_Lane5_SH5 | GW_Lane5_SH5 | SH | lud_PK | lud_PK | 39.1 | 92939 |
| 185 | GW_Lane5_SL1 | GW_Lane5_SL1 | SL | plumb | plumb | 150.0 | 64888 |
| 186 | GW_Lane5_SL2 | GW_Lane5_SL2 | SL | plumb | plumb | 151.0 | 65473 |
| 187 | GW_Lane5_ST1 | GW_Lane5_ST1 | ST | plumb | plumb | 142.0 | 60624 |
| 188 | GW_Lane5_ST12 | GW_Lane5_ST12 | ST | plumb | plumb | 144.0 | 6912 |
| 189 | GW_Lane5_ST3 | GW_Lane5_ST3 | ST | plumb | plumb | 143.0 | 69699 |
| 190 | GW_Lane5_STvi1 | GW_Lane5_STvi1 | ST_vi | vir | vir | 22.0 | 6900 |
| 191 | GW_Lane5_STvi2 | GW_Lane5_STvi2 | ST_vi | vir | vir | 23.0 | 8973 |
| 192 | GW_Lane5_STvi3 | GW_Lane5_STvi3 | ST_vi | vir | vir | 24.0 | 7681 |
| 193 | GW_Lane5_TA1 | GW_Lane5_TA1 | TA | plumb | plumb | 86.0 | 71190 |
| 194 | GW_Lane5_TL1 | GW_Lane5_TL1 | TL | vir | vir | 9.0 | 743509 |
| 195 | GW_Lane5_TL10 | GW_Lane5_TL10 | TL | vir | vir | 17.0 | 669934 |
| 196 | GW_Lane5_TL11 | GW_Lane5_TL11 | TL | vir | vir | 18.0 | 638402 |
| 197 | GW_Lane5_TL12 | GW_Lane5_TL12 | TL | vir | vir | 19.0 | 585697 |
| 198 | GW_Lane5_TL2 | GW_Lane5_TL2 | TL | vir | vir | 10.0 | 770857 |
| 199 | GW_Lane5_TL4 | GW_Lane5_TL4 | TL | vir | vir | 12.0 | 758037 |
| 200 | GW_Lane5_TL5 | GW_Lane5_TL5 | TL | vir | vir | 13.0 | 867165 |
| 201 | GW_Lane5_TL7 | GW_Lane5_TL7 | TL | vir | vir | 14.0 | 803407 |
| 202 | GW_Lane5_TL8 | GW_Lane5_TL8 | TL | vir | vir | 15.0 | 698745 |
| 203 | GW_Lane5_TL9 | GW_Lane5_TL9 | TL | vir | vir | 16.0 | 606969 |
| 204 | GW_Lane5_TU1 | GW_Lane5_TU1 | TU | nit | nit | 35.0 | 793640 |
| 205 | GW_Lane5_TU2 | GW_Lane5_TU2 | TU | nit | nit | 36.0 | 736785 |
| 206 | GW_Lane5_UY2 | GW_Lane5_UY2 | UY | plumb | plumb | 88.0 | 72900 |
| 207 | GW_Lane5_UY3 | GW_Lane5_UY3 | UY | plumb | plumb | 89.0 | 67752 |
| 208 | GW_Lane5_UY4 | GW_Lane5_UY4 | UY | plumb | plumb | 90.0 | 74984 |
| 209 | GW_Lane5_UY5 | GW_Lane5_UY5 | UY | plumb | plumb | 91.0 | 71837 |
| 210 | GW_Lane5_UY6 | GW_Lane5_UY6 | UY | plumb | plumb | 92.0 | 71367 |
| 211 | GW_Lane5_YK1 | GW_Lane5_YK1 | YK | vir | vir | 1.0 | 831245 |
| 212 | GW_Lane5_YK11 | GW_Lane5_YK11 | YK | vir | vir | 8.0 | 730798 |
| 213 | GW_Lane5_YK3 | GW_Lane5_YK3 | YK | vir | vir | 2.0 | 731944 |
| 214 | GW_Lane5_YK4 | GW_Lane5_YK4 | YK | vir | vir | 3.0 | 740051 |
| 215 | GW_Lane5_YK5 | GW_Lane5_YK5 | YK | vir | vir | 4.0 | 738740 |
| 216 | GW_Lane5_YK6 | GW_Lane5_YK6 | YK | vir | vir | 5.0 | 697420 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 217 | GW_Lane5_YK7 | GW_Lane5_YK7 | YK | vir | vir | 6.0 | 692052 | |
| 218 | GW_Lane5_YK9 | GW_Lane5_YK9 | YK | vir | vir | 7.0 | 768722 | |
| 219 | GW_Liz_GBS_Liz10045 | GW_Liz_GBS_Liz10045 | ML | lud | lud_ML | 51.01 | 8 | |
| 220 | GW_Liz_GBS_Liz10094 | GW_Liz_GBS_Liz10094 | ML | lud | lud_ML | 51.02 | 9 | |
| 221 | GW_Liz_GBS_Liz5144 | GW_Liz_GBS_Liz5144 | ML | lud | lud_ML | 51.08 | 9 | |
| 222 | GW_Liz_GBS_Liz5163 | GW_Liz_GBS_Liz5163 | ML | lud_chick | lud_ML | 51.12 | | |
| 223 | GW_Liz_GBS_Liz5164 | GW_Liz_GBS_Liz5164 | ML | lud_chick | lud_ML | 51.13 | | |
| 224 | GW_Liz_GBS_Liz5165 | GW_Liz_GBS_Liz5165 | ML | lud | lud_ML | 51.14 | 9 | |
| 225 | GW_Liz_GBS_Liz5167 | GW_Liz_GBS_Liz5167 | ML | lud_chick | lud_ML | 51.15 | | |
| 226 | GW_Liz_GBS_Liz5168 | GW_Liz_GBS_Liz5168 | ML | lud_chick | lud_ML | 51.16 | | |
| 227 | GW_Liz_GBS_Liz5173 | GW_Liz_GBS_Liz5173 | ML | lud_chick | lud_ML | 51.2 | | |
| 228 | GW_Liz_GBS_Liz5175 | GW_Liz_GBS_Liz5175 | ML | lud | lud_ML | 51.22 | 9 | |
| 229 | GW_Liz_GBS_Liz5178 | GW_Liz_GBS_Liz5178 | ML | lud_chick | lud_ML | 51.25 | | |
| 230 | GW_Liz_GBS_Liz5179 | GW_Liz_GBS_Liz5179 | ML | lud_chick | lud_ML | 51.26 | | |
| 231 | GW_Liz_GBS_Liz5182 | GW_Liz_GBS_Liz5182 | ML | lud_chick | lud_ML | 51.28 | | |
| 232 | GW_Liz_GBS_Liz5184 | GW_Liz_GBS_Liz5184 | ML | lud_chick | lud_ML | 51.29 | | |
| 233 | GW_Liz_GBS_Liz5185 | GW_Liz_GBS_Liz5185 | ML | lud | lud_ML | 51.3 | 94 | |
| 234 | GW_Liz_GBS_Liz5188 | GW_Liz_GBS_Liz5188 | ML | lud | lud_ML | 51.33 | 8 | |
| 235 | GW_Liz_GBS_Liz5189 | GW_Liz_GBS_Liz5189 | ML | lud_chick | lud_ML | 51.34 | | |
| 236 | GW_Liz_GBS_Liz5190 | GW_Liz_GBS_Liz5190 | ML | lud_chick | lud_ML | 51.35 | | |
| 237 | GW_Liz_GBS_Liz5191 | GW_Liz_GBS_Liz5191 | ML | lud_chick | lud_ML | 51.36 | | |
| 238 | GW_Liz_GBS_Liz5193 | GW_Liz_GBS_Liz5193 | ML | lud_chick | lud_ML | 51.38 | | |
| 239 | GW_Liz_GBS_Liz5194 | GW_Liz_GBS_Liz5194 | ML | lud_chick | lud_ML | 51.39 | | |
| 240 | GW_Liz_GBS_Liz5197 | GW_Liz_GBS_Liz5197 | ML | lud | lud_ML | 51.41 | 8 | |
| 241 | GW_Liz_GBS_Liz5199 | GW_Liz_GBS_Liz5199 | ML | lud_chick | lud_ML | 51.42 | | |
| 242 | GW_Liz_GBS_Liz6002 | GW_Liz_GBS_Liz6002 | ML | lud | lud_ML | 51.43 | 9 | |
| 243 | GW_Liz_GBS_Liz6006 | GW_Liz_GBS_Liz6006 | ML | lud | lud_ML | 51.44 | 8 | |
| 244 | GW_Liz_GBS_Liz6008 | GW_Liz_GBS_Liz6008 | ML | lud | lud_ML | 51.45 | 9 | |
| 245 | GW_Liz_GBS_Liz6009 | GW_Liz_GBS_Liz6009 | ML | lud | lud_ML | 51.46 | 9 | |
| 246 | GW_Liz_GBS_Liz6010 | GW_Liz_GBS_Liz6010 | ML | lud | lud_ML | 51.47 | 8 | |
| 247 | GW_Liz_GBS_Liz6014 | GW_Liz_GBS_Liz6014 | ML | lud | lud_ML | 51.49 | 8 | |
| 248 | GW_Liz_GBS_Liz6055 | GW_Liz_GBS_Liz6055 | ML | lud | lud_ML | 51.5 | 87 | |
| 249 | GW_Liz_GBS_Liz6057 | GW_Liz_GBS_Liz6057 | ML | lud | lud_ML | 51.51 | 8 | |
| 250 | GW_Liz_GBS_Liz6060 | GW_Liz_GBS_Liz6060 | ML | lud | lud_ML | 51.52 | 8 | |
| 251 | GW_Liz_GBS_Liz6062 | GW_Liz_GBS_Liz6062 | ML | lud | lud_ML | 51.53 | 8 | |
| 252 | GW_Liz_GBS_Liz6063 | GW_Liz_GBS_Liz6063 | ML | lud | lud_ML | 51.54 | 8 | |
| 253 | GW_Liz_GBS_Liz6066 | GW_Liz_GBS_Liz6066 | ML | lud | lud_ML | 51.55 | 8 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 254 | GW_Liz_GBS_Liz6072 | GW_Liz_GBS_Liz6072 | ML | lud | lud_ML | 51.56 | 8 |
| 255 | GW_Liz_GBS_Liz6079 | GW_Liz_GBS_Liz6079 | ML | lud | lud_ML | 51.57 | 8 |
| 256 | GW_Liz_GBS_Liz6204 | GW_Liz_GBS_Liz6204 | ML | lud_chick | lud_ML | 51.59 | |
| 257 | GW_Liz_GBS_Liz6461 | GW_Liz_GBS_Liz6461 | ML | lud | lud_ML | 51.6 | 81 |
| 258 | GW_Liz_GBS_Liz6472 | GW_Liz_GBS_Liz6472 | ML | lud | lud_ML | 51.61 | 9 |
| 259 | GW_Liz_GBS_Liz6478 | GW_Liz_GBS_Liz6478 | ML | lud | lud_ML | 51.62 | 7 |
| 260 | GW_Liz_GBS_Liz6776 | GW_Liz_GBS_Liz6776 | ML | lud | lud_ML | 51.64 | 9 |
| 261 | GW_Liz_GBS_Liz6794 | GW_Liz_GBS_Liz6794 | ML | lud | lud_ML | 51.65 | 8 |

### Filter SNPs with too many missing genotypes:

```
# (remember that first column is arbitrary row number in input file)
missing_genotypes_per_SNP = sum(geno_indFiltered .== -1, dims=1)
missing_genotypes_percent_allowed_per_site = 5   # this is the percentage threshold
threshold_genotypes_missing = size(geno_indFiltered)[1] * missing_genotypes_percent_allowed_per_site/10
selection = vec(missing_genotypes_per_SNP .<= threshold_genotypes_missing)
geno_ind_SNP_filtered = geno_indFiltered[:, selection]
pos_SNP_filtered = pos_whole_genome[selection[Not(1)],:]  # the Not(1) is needed because first column i
println("Started with ", size(geno_indFiltered, 2)-1, " SNPs.
After filtering SNPs for no more than ", missing_genotypes_percent_allowed_per_site, "% missing genotyp
```

```
Started with 2431709 SNPs.
After filtering SNPs for no more than 5% missing genotypes, 1017581 SNPs remain.
```

### 2nd round of filtering individuals

I added this in August 2023, to improve accuracy of imputation-based PCA, because I noticed outliers tended to have more missing data. Now I only allow up to 10% missing SNPs per individual.

```
SNPmissing_percent_allowed_per_ind_round2 = 10   # this is the percentage threshold
threshold_missing = (size(geno_ind_SNP_filtered, 2) - 1) * SNPmissing_percent_allowed_per_ind_round2/10
numMissings = sum(geno_ind_SNP_filtered .== -1, dims=2)
selection = vec(numMissings .<= threshold_missing) # the vec command converts to BitVector rather than
geno_ind_SNP_ind_filtered = geno_ind_SNP_filtered[selection, :]
println("Filtering out these individuals based on too many missing genotypes: ")
filtered_inds = ind_with_metadata_indFiltered.ind[selection.==false]
println(DataFrame(filtered_inds = filtered_inds)) # did this to print all lines
ind_with_metadata_indFiltered = ind_with_metadata_indFiltered[selection, :]
println("This leaves ", size(geno_ind_SNP_ind_filtered, 1), " individuals and ", size(geno_ind_SNP_ind_
```

```
with no individuals missing more than ", SNPmissing_percent_allowed_per_ind_round2, "% of genotypes
and no loci missing in more than ", missing_genotypes_percent_allowed_per_site, "% of individuals.")
```

```
Filtering out these individuals based on too many missing genotypes:
4×1 DataFrame
 Row │ filtered_inds
     │ String
─────┼──────────────────────────
   1 │ GW_Armando_plate1_TTGW74
   2 │ GW_Armando_plate2_TTGW54
   3 │ GW_Lane5_AA8
   4 │ GW_Lane5_YK1
This leaves 257 individuals and 1017581 loci,
with no individuals missing more than 10% of genotypes
and no loci missing in more than 5% of individuals.
```

Remove the first column of the genotype matrix (which was an initial row number):

```
genosOnly_ind_SNP_ind_filtered = geno_ind_SNP_ind_filtered[:, Not(1)]
```

```
257×1017581 Matrix{Int16}:
 0  0  0  0  0  1  0   0  0  0  0  0  …  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0  …  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   1  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0  …  0  0  0  0  0  0  0  0  0  0  0  1
 0  0  0  0  0  0  0   0  0  0  0  1     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 ⋮              ⋮              ⋮     ⋱     ⋮              ⋮              ⋮
 0  0  0  0  0  0  0   0  0  0  0  0  …  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  1  0     1  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
```

```
 0  0  0  0  0  0  0  -1  0  0  0  0  …  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0  …  0  0  0  0  0  0  0  0  0  0  0  0
 0  0  0  0  0  0  0   0  0  0  0  0     0  0  0  0  0  0  0  0  0  0  0  0
```

**Polish a few individual names (for more readable graphs):**

```
ind_with_metadata_indFiltered.ind = correctNames(ind_with_metadata_indFiltered.ind)

ind_with_metadata_indFiltered.ID = correctNames(ind_with_metadata_indFiltered.ID)
```

```
257-element Vector{String}:
 "GW_Armando_plate1_AB1"
 "GW_Armando_plate1_JF07G02"
 "GW_Armando_plate1_JF07G03"
 "GW_Armando_plate1_JF07G04"
 "GW_Armando_plate1_JF08G02"
 "GW_Armando_plate1_JF09G01"
 "GW_Armando_plate1_JF09G02"
 "GW_Armando_plate1_JF10G03"
 "GW_Armando_plate1_JF11G01"
 "GW_Armando_plate1_JF12G01"
 "GW_Armando_plate1_JF12G02"
 "GW_Armando_plate1_JF12G04"
 "GW_Armando_plate1_JF13G01"
 ⋮
 "GW_Liz_GBS_Liz6060"
 "GW_Liz_GBS_Liz6062"
 "GW_Liz_GBS_Liz6063"
 "GW_Liz_GBS_Liz6066"
 "GW_Liz_GBS_Liz6072"
 "GW_Liz_GBS_Liz6079"
 "GW_Liz_GBS_Liz6204"
 "GW_Liz_GBS_Liz6461"
 "GW_Liz_GBS_Liz6472"
 "GW_Liz_GBS_Liz6478"
 "GW_Liz_GBS_Liz6776"
 "GW_Liz_GBS_Liz6794"
```

## Calculate distances around ring

The locations around the ring (assuming barrier in North) can be graphed against genomic PC1 (or other variables).

### Load lat/long data

```
cd(repoDirectory)
latlong_filepath = "metadata/GW_locations_LatLong_2023.txt"
latlongs = DataFrame(CSV.File(latlong_filepath))
print(latlongs)
```

37×5 DataFrame

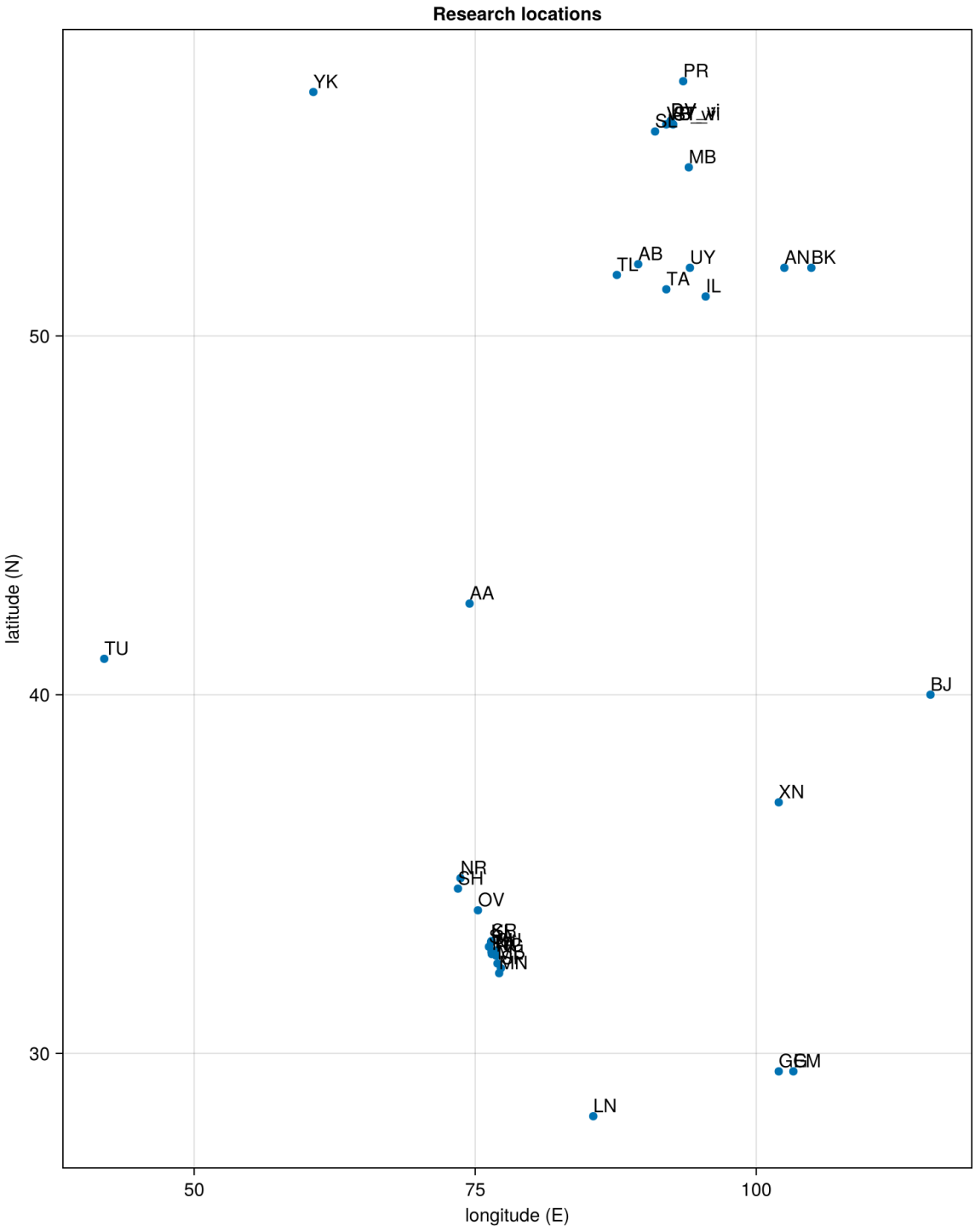
| Row | Location_name | location_short | lat_N | long_E | subspecies |
|---|---|---|---|---|---|
| | String31 | String7 | Float64 | Float64 | String15 |
| 1 | Yekaterinburg | YK | 56.8 | 60.6 | viridanus |
| 2 | Abakan | AB | 52.0 | 89.5 | viridanus |
| 3 | Teletsk | TL | 51.7 | 87.6 | viridanus |
| 4 | Verkh_Biryusa_vir | VB_vi | 55.9 | 92.0 | viridanus |
| 5 | Divnogorsk_vir | DV_vi | 56.0 | 92.4 | viridanus |
| 6 | Stolbi_vir | ST_vi | 55.9 | 92.6 | viridanus |
| 7 | Turkey | TU | 41.0 | 42.0 | nitidus |
| 8 | Ala_Archa | AA | 42.54 | 74.5 | viridanus |
| 9 | Naran_Pakistan | NR | 34.884 | 73.691 | ludlowi |
| 10 | Shogran_Pakistan | SH | 34.594 | 73.466 | ludlowi |
| 11 | Overa_Kashmir | OV | 33.991 | 75.243 | ludlowi |
| 12 | Satharundhi_ChambaDistrict | SA | 32.974 | 76.222 | ludlowi |
| 13 | KL_Killar_HP | KL | 33.106 | 76.409 | ludlowi |
| 14 | Thalighar | TH | 32.828 | 76.45 | ludlowi |
| 15 | Sural | SR | 33.134 | 76.455 | ludlowi |
| 16 | PA_Tindi_HP | PA | 32.771 | 76.472 | ludlowi |
| 17 | Sukhto | SU | 32.868 | 76.855 | ludlowi |
| 18 | Nainaghar | NG | 32.728 | 76.8594 | ludlowi |
| 19 | Mooling_and_Keylong | ML | 32.508 | 76.981 | ludlowi |

| | | | | | |
|---|---|---|---|---|---|
| 20 | Manali | MN | 32.237 | 77.13 | ludlowi |
| 21 | Spiti | SP | 32.377 | 77.281 | ludlowi |
| 22 | Langtang | LN | 28.25 | 85.5 | trochiloides |
| 23 | Gongga | GG | 29.5 | 102.0 | obscuratus |
| 24 | Emeishan | EM | 29.5 | 103.3 | obscuratus |
| 25 | Xining | XN | 37.0 | 102.0 | obscuratus |
| 26 | Beijing | BJ | 40.0 | 115.5 | plumbeitarsus |
| 27 | Baikal | BK | 51.9 | 104.9 | plumbeitarsus |
| 28 | Arshan | AN | 51.9 | 102.5 | plumbeitarsus |
| 29 | Ilinka | IL | 51.1 | 95.5 | plumbeitarsus |
| 30 | Tuva | TA | 51.3 | 92.0 | plumbeitarsus |
| 31 | Uyukski | UY | 51.9 | 94.1 | plumbeitarsus |
| 32 | Manskoe_Belogorie | MB | 54.7 | 94.0 | plumbeitarsus |
| 33 | Verkh_Biryusa | VB | 55.9 | 92.0 | plumbeitarsus |
| 34 | Divnogorsk | DV | 56.0 | 92.4 | plumbeitarsus |
| 35 | Stolbi | ST | 55.9 | 92.6 | plumbeitarsus |
| 36 | Predivinsk | PR | 57.1 | 93.5 | plumbeitarsus |
| 37 | Solgonski | SL | 55.7 | 91.0 | plumbeitarsus |

**Make a quick plot to inspect latlong data:**

```
f = CairoMakie.Figure()
ax = Axis(f[1, 1],
    title = "Research locations",
    xlabel = "longitude (E)",
    ylabel = "latitude (N)"
)
scatter!(latlongs.long_E, latlongs.lat_N)
text!(latlongs.long_E, latlongs.lat_N; text = latlongs.location_short)
f
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```


Research locations
YK
PR
MB
TL
AB
UY
ANBK
TA
IL
AA
TU
BJ
XN
NR
SH
OV
LN
50
40
30
50
75
100
latitude (N)
longitude (E)

### remove "Green warbler" *nitidus*

*Phylloscopus [t.] nitidus* is outside of the main ring, so remove these samples from this analysis:

```
latlongs2 = latlongs[Not(latlongs.subspecies .== "nitidus"), :];
print(latlongs2)
```

```
36×5 DataFrame
 Row │ Location_name               location_short  lat_N    long_E    subspecies
     │ String31                    String7         Float64  Float64   String15
─────┼──────────────────────────────────────────────────────────────────────────
   1 │ Yekaterinburg               YK               56.8     60.6     viridanus
   2 │ Abakan                      AB               52.0     89.5     viridanus
   3 │ Teletsk                     TL               51.7     87.6     viridanus
   4 │ Verkh_Biryusa_vir           VB_vi            55.9     92.0     viridanus
   5 │ Divnogorsk_vir              DV_vi            56.0     92.4     viridanus
   6 │ Stolbi_vir                  ST_vi            55.9     92.6     viridanus
   7 │ Ala_Archa                   AA               42.54    74.5     viridanus
   8 │ Naran_Pakistan              NR               34.884   73.691   ludlowi
   9 │ Shogran_Pakistan            SH               34.594   73.466   ludlowi
  10 │ Overa_Kashmir               OV               33.991   75.243   ludlowi
  11 │ Satharundhi_ChambaDistrict  SA               32.974   76.222   ludlowi
  12 │ KL_Killar_HP                KL               33.106   76.409   ludlowi
  13 │ Thalighar                   TH               32.828   76.45    ludlowi
  14 │ Sural                       SR               33.134   76.455   ludlowi
  15 │ PA_Tindi_HP                 PA               32.771   76.472   ludlowi
  16 │ Sukhto                      SU               32.868   76.855   ludlowi
  17 │ Nainaghar                   NG               32.728   76.8594  ludlowi
  18 │ Mooling_and_Keylong         ML               32.508   76.981   ludlowi
  19 │ Manali                      MN               32.237   77.13    ludlowi
  20 │ Spiti                       SP               32.377   77.281   ludlowi
  21 │ Langtang                    LN               28.25    85.5     trochiloides
  22 │ Gongga                      GG               29.5    102.0     obscuratus
  23 │ Emeishan                    EM               29.5    103.3     obscuratus
  24 │ Xining                      XN               37.0    102.0     obscuratus
  25 │ Beijing                     BJ               40.0    115.5     plumbeitarsus
  26 │ Baikal                      BK               51.9    104.9     plumbeitarsus
```

| | | | | | |
|---|---|---|---|---|---|
| 27 | Arshan | AN | 51.9 | 102.5 | plumbeitarsus |
| 28 | Ilinka | IL | 51.1 | 95.5 | plumbeitarsus |
| 29 | Tuva | TA | 51.3 | 92.0 | plumbeitarsus |
| 30 | Uyukski | UY | 51.9 | 94.1 | plumbeitarsus |
| 31 | Manskoe_Belogorie | MB | 54.7 | 94.0 | plumbeitarsus |
| 32 | Verkh_Biryusa | VB | 55.9 | 92.0 | plumbeitarsus |
| 33 | Divnogorsk | DV | 56.0 | 92.4 | plumbeitarsus |
| 34 | Stolbi | ST | 55.9 | 92.6 | plumbeitarsus |
| 35 | Predivinsk | PR | 57.1 | 93.5 | plumbeitarsus |
| 36 | Solgonski | SL | 55.7 | 91.0 | plumbeitarsus |

### Make a matrix of great circle distances

These are Haversine distances, assuming spherical Earth which is really close:

```
geoPoints = GeoLocation.(latlongs2.long_E, latlongs2.lat_N)
# this next line is so neat--uses list comprehension to make a matrix of pairwise calculations
distances = [(HaversineDistance(geoPoints[i], geoPoints[j])/1000) for i in eachindex(geoPoints), j in e
```

```
36×36 Matrix{Float64}:
    0.0   1929.03   1829.5    1920.28    …  1956.21    1976.01   1866.47
 1929.03     0.0     134.698   463.414       478.643    622.753   422.996
 1829.5    134.698     0.0     548.935       570.583    711.032   497.776
 1920.28   463.414   548.935     0.0          37.404    162.101    66.3387
 1941.17   483.376   572.19     27.2735       16.6941   139.661    93.5371
 1956.21   478.643   570.583    37.404    …     0.0      144.411   102.442
 1866.88  1539.57   1417.12   1943.75       1971.61    2101.68   1883.0
 2629.03  2280.66   2174.21   2720.63       2744.17    2882.26   2664.81
 2653.46  2318.81   2212.24   2758.56       2782.15    2920.16   2702.67
 2768.57  2304.58   2204.89   2753.66       2775.3     2915.75   2700.28
 2905.38  2370.69   2276.33   2825.06    …  2845.24    2987.08   2773.58
 2897.44  2350.4    2256.5    2805.15       2825.21    2967.15   2753.85
 2927.82  2377.48   2284.21   2832.75       2852.64    2994.71   2781.67
    ⋮                                   ⋱                       ⋮
 4317.27  2390.71   2499.26   2464.31       2434.21    2474.86   2502.82
 2869.29  1053.52   1187.02    953.003   …   918.603    934.047  1003.57
 2724.34   889.835  1023.03    817.872       785.302    818.533   863.881
 2341.58   426.63    551.809   581.592       567.027    679.716   591.859
 2117.8    189.217   307.752   511.497       513.021    652.226   493.694
 2214.04   315.404   447.369   465.516       455.478    579.504   468.914
 2082.09   423.338   540.944   183.922   …   160.175    268.68    220.448
```

```
 1920.28   463.414   548.935     0.0          37.404    162.101    66.3387
 1941.17   483.376   572.19     27.2735       16.6941   139.661    93.5371
 1956.21   478.643   570.583    37.404         0.0      144.411   102.442
 1976.01   622.753   711.032   162.101       144.411      0.0     218.833
 1866.47   422.996   497.776    66.3387  …   102.442    218.833     0.0
```

**Now adjust distances to assume no gene flow through centre of ring.**

```
# get some key distances
function getIndex(name; nameVector = latlongs2.Location_name)
    findfirst(isequal(name), nameVector)
end

index_AA = getIndex("Ala_Archa")
index_NR = getIndex("Naran_Pakistan")
index_LN = getIndex("Langtang")
index_GG = getIndex("Gongga")
index_XN = getIndex("Xining")
index_BJ = getIndex("Beijing")
index_last = nrow(latlongs2)

dist_NR_to_LN = distances[index_NR, index_LN]
dist_LN_to_GG = distances[index_LN, index_GG]
dist_GG_to_BJ = distances[index_GG, index_BJ]

# This next part will assume locations in the input file are arranged in order around ring:
distsAroundRing = Matrix{Float32}(undef, size(distances)[1], size(distances)[2])
# accept all distances within viridanus:

# function for accepting straight-line great circle dists as distances between sets of sites
acceptDists = function(straightGreatCircleDists, start, finish, distsAroundRing)
    distsAroundRing[start:finish, start:finish] = straightGreatCircleDists[start:finish, start:finish]
    return(distsAroundRing)
end

# accept all distances within viridanus:
distsAroundRing = acceptDists(distances, 1, index_AA, distsAroundRing)

# accept dist from AA to NR:
distsAroundRing = acceptDists(distances, index_AA, index_NR, distsAroundRing)

# accept all distances from NR to LN:
```

```
distsAroundRing = acceptDists(distances, index_NR, index_LN, distsAroundRing)

# accept dist from LN to GG:
distsAroundRing = acceptDists(distances, index_LN, index_GG, distsAroundRing)

# accept dists between GG, EM, XN, BJ:
distsAroundRing = acceptDists(distances, index_GG, index_BJ, distsAroundRing)

# accept all distances within plumbeitarsus:
distsAroundRing = acceptDists(distances, index_BJ, index_last, distsAroundRing)

# function for adding up distances measured through certain sites:
addDists = function(set1start, set1end, set2start, set2end, distsAroundRing)
    firstDists = repeat(distsAroundRing[set1start:(set1end-1), set1end], 1, set2end-set2start+1)
    secondDists = repeat(transpose(distsAroundRing[set1end, set2start:set2end]), set1end-set1start, 1)
    totalDists = firstDists + secondDists
    distsAroundRing[set1start:(set1end-1), set2start:set2end] = totalDists
    distsAroundRing[set2start:set2end, set1start:(set1end-1)] = transpose(totalDists)
    return(distsAroundRing)
end

# dists from viridanus to NR are sum of dists to AA plus AA to NR:
distsAroundRing = addDists(1, index_AA, index_NR, index_NR, distsAroundRing)

# dists from "northwest of PK" to Himalayas are sum of ringdists to NR plus NR to locations up to LN:
distsAroundRing = addDists(1, index_NR, index_NR+1, index_LN, distsAroundRing)

# dists from "west / northwest of LN" to GG are sum of dists to LN plus LN to EM:
distsAroundRing = addDists(1, index_LN, index_GG, index_GG, distsAroundRing)

# dists from "west / northwest of EM" to China are sum of dists to GG plus GG to (EM, XN, BJ):
distsAroundRing = addDists(1, index_GG, index_GG, index_BJ, distsAroundRing)

# dists from "west of BJ" to east Siberia are sum of dists to BJ plus BJ to other plumbeitarsus:
distsAroundRing = addDists(1, index_BJ, index_BJ+1, index_last, distsAroundRing);
```

### Do Principal Coordinates Analysis on the distances around the ring

This produces a single location axis around ring, going from west Siberia south, then east, then north to east Siberia.

```
PCO_around_ring = fit(MDS, distsAroundRing; distances=true, maxoutdim=1)
# add this as a column to the data frame:
latlongs2.LocationAroundRing = vec(-predict(PCO_around_ring))
# another way:
# latlongs2[:, :LocationAroundRing] = vec(-predict(PCO_around_ring))
latlongs2[:, [:location_short, :LocationAroundRing]]
println(latlongs2[:, [:location_short, :LocationAroundRing]])
```

```
36×2 DataFrame
 Row │ location_short  LocationAroundRing
     │ String7         Float32
─────┼──────────────────────────────────
   1 │ YK                        -4799.38
   2 │ AB                        -4548.43
   3 │ TL                        -4428.78
   4 │ VB_vi                     -4953.25
   5 │ DV_vi                     -4978.83
   6 │ ST_vi                     -4980.5
   7 │ AA                        -3027.02
   8 │ NR                        -2171.62
   9 │ SH                        -2166.63
  10 │ OV                        -1998.24
  11 │ SA                        -1861.95
  12 │ KL                        -1854.67
  13 │ TH                        -1835.41
  14 │ SR                        -1852.63
  15 │ PA                        -1830.4
  16 │ SU                        -1805.38
  17 │ NG                        -1796.96
  18 │ ML                        -1774.64
  19 │ MN                        -1747.34
  20 │ SP                        -1743.06
  21 │ LN                         -830.651
  22 │ GG                          783.671
  23 │ EM                          890.416
  24 │ XN                         1481.8
  25 │ BJ                         2480.12
  26 │ BK                         4027.02
```

| | | |
|---|---|---|
| 27 | AN | 4134.32 |
| 28 | IL | 4451.78 |
| 29 | TA | 4673.03 |
| 30 | UY | 4581.05 |
| 31 | MB | 4762.52 |
| 32 | VB | 4943.39 |
| 33 | DV | 4929.77 |
| 34 | ST | 4913.15 |
| 35 | PR | 4951.66 |
| 36 | SL | 4982.05 |

Add these ring locations to the metadata table:

```
ind_with_metadata_indFiltered.ring_km .= NaN  # pre-allocate the column
for i in axes(latlongs2, 1)
    match_indices = findall(ind_with_metadata_indFiltered.location .== latlongs2.location_short[i])
    ind_with_metadata_indFiltered.ring_km[match_indices] .= latlongs2.LocationAroundRing[i];
end
```

## Make a table of locations and groups of individuals included after filtering

```
gdf = groupby(ind_with_metadata_indFiltered, [:location, :Fst_group])
sample_origin_summary = combine(gdf, nrow)
println(sample_origin_summary)
# to save as comma-delimited text:
# CSV.write("GW2023_sample_origin_summary.txt", sample_origin_summary; delim='\t')
```

```
37×3 DataFrame
 Row │ location  Fst_group    nrow
     │ String7   String15     Int64
─────┼──────────────────────────────
   1 │ AB        vir              2
   2 │ ST        plumb           35
   3 │ ST        plumb_vir        1
   4 │ ST_vi     vir              6
   5 │ DV        plumb            5
```

| | | | |
|---|---|---|---|
| 6 | DV_vi | vir | 2 |
| 7 | MB | plumb | 4 |
| 8 | VB | plumb | 5 |
| 9 | PR | plumb | 5 |
| 10 | BJ | plumb_BJ | 3 |
| 11 | TL | vir | 11 |
| 12 | MN | troch_west | 10 |
| 13 | SU | lud_central | 6 |
| 14 | TH | lud_central | 7 |
| 15 | SR | lud_central | 18 |
| 16 | NG | lud_central | 9 |
| 17 | SP | troch_west | 10 |
| 18 | SA | lud_Sath | 8 |
| 19 | UY | plumb | 6 |
| 20 | VB_vi | vir | 2 |
| 21 | LN | troch_LN | 14 |
| 22 | AA | vir_S | 8 |
| 23 | AN | plumb | 2 |
| 24 | BK | plumb | 2 |
| 25 | XN | obs | 4 |
| 26 | EM | troch_EM | 1 |
| 27 | IL | plumb | 2 |
| 28 | OV | lud_KS | 2 |
| 29 | NR | lud_PK | 2 |
| 30 | KL | lud_central | 3 |
| 31 | ML | lud_ML | 44 |
| 32 | PA | lud_central | 2 |
| 33 | SH | lud_PK | 4 |
| 34 | SL | plumb | 2 |
| 35 | TA | plumb | 1 |
| 36 | TU | nit | 2 |
| 37 | YK | vir | 7 |

## Save the filtered dataset

```
cd(dataDirectory)
filename = string(baseName, tagName, "ind_SNP_ind_filtered.jld2")
jldsave(filename; genosOnly_ind_SNP_ind_filtered, ind_with_metadata_indFiltered, pos_SNP_filtered)
println("Saved the filtered data.")
```

```
Saved the filtered data.
```

HERE WILL PASS THE DATA OFF TO OTHER QUARTO PAGES, TO LOAD DATA AS BELOW:

### Load the filtered dataset

```
filename = string(baseName, tagName, "ind_SNP_ind_filtered.jld2")
genosOnly_ind_SNP_ind_filtered = load(filename, "genosOnly_ind_SNP_ind_filtered")
ind_with_metadata_indFiltered = load(filename, "ind_with_metadata_indFiltered")
pos_SNP_filtered = load(filename, "pos_SNP_filtered")
println("Loaded the filtered data.")
```

### Prepare data for Genotype-by-individual plots and PCA

For missing genotypes, change our code of `-1` to `missing`:

```
genosOnly_with_missing = Matrix{Union{Missing, Int16}}(genosOnly_ind_SNP_ind_filtered)
genosOnly_with_missing[genosOnly_with_missing .== -1] .= missing;
```

### Genotype-by-individual plots

Now, show individual genotypes for subsets of the dataset. Can choose individuals and genomic regions to plot, along with an Fst cutoff (only show SNPs with greater Fst than the cutoff).

```
groups = ["vir","plumb"]
plotGroups = ["vir","plumb_vir","plumb"]
plotGroupColors = ["blue","purple","red"]
numIndsToPlot = [100,100,100] # maximum number of individuals to plot from each group
group1 = "vir"   # these groups will determine the color used in the graph
group2 = "plumb"
groupsToCompare = "vir_plumb"
Fst_cutoff =  0.8
missingFractionAllowed = 0.2  # only show SNPs with less than this fraction of missing data among indiv
```

```
0.2
```

### Calculate allele freqs and sample sizes (use column Fst_group)

```
freqs, sampleSizes = getFreqsAndSampleSizes(genosOnly_with_missing, ind_with_metadata_indFiltered.Fst_g
println("Calculated population allele frequencies and sample sizes")
```

```
Calculated population allele frequencies and sample sizes
```

### calculate Fst

```
Fst, FstNumerator, FstDenominator, pairwiseNamesFst = getFst(freqs, sampleSizes, groups; among=true)  #
println("Calculated Fst values")
```

```
Calculated Fst values
```

### limit the individuals to include in plot

```
# For this figure only, filter out individuals with lots of missing genotypes
numMissings_threshold = 800_000
selection = ind_with_metadata_indFiltered.numMissings .< numMissings_threshold
genosOnly_with_missing_selected = view(genosOnly_with_missing, selection, :)
ind_with_metadata_indFiltered_selected = view(ind_with_metadata_indFiltered, selection, :)
# now limit each group to specified numbers
genosOnly_included, ind_with_metadata_included = limitIndsToPlot(plotGroups, numIndsToPlot, genosOnly_w
```

### Choose the scaffold and region to show

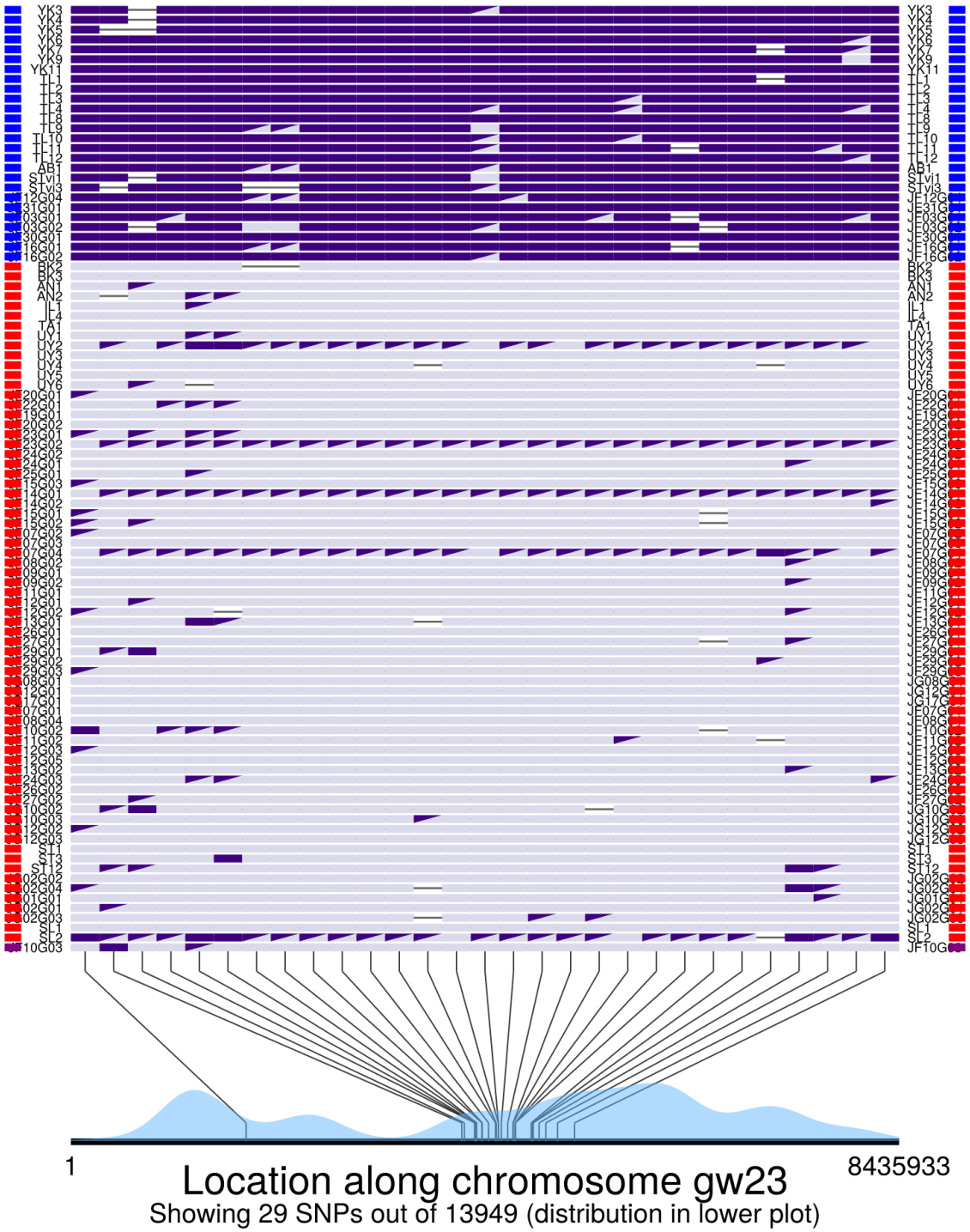
```
chr = "gw23"
regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr]) #
```

```
("gw23", 1, 8435933, "chr gw23 1 to 8435933")
```

### Now actually make the plot

```
plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff, missingFractionAllowed,
    regionInfo, pos_SNP_filtered, Fst, pairwiseNamesFst,
    genosOnly_included, ind_with_metadata_included, freqs, plotGroups, plotGroupColors; plotTitle = "")
# plotInfo contains a tuple with: (f, plottedGenotype, locations, plottedMetadata)

if false  # set to true to save plot
    save("Figure4_upperleft_gw23GBIplot_from_Julia.png", plotInfo[1], px_per_unit = 2.0)
end
```

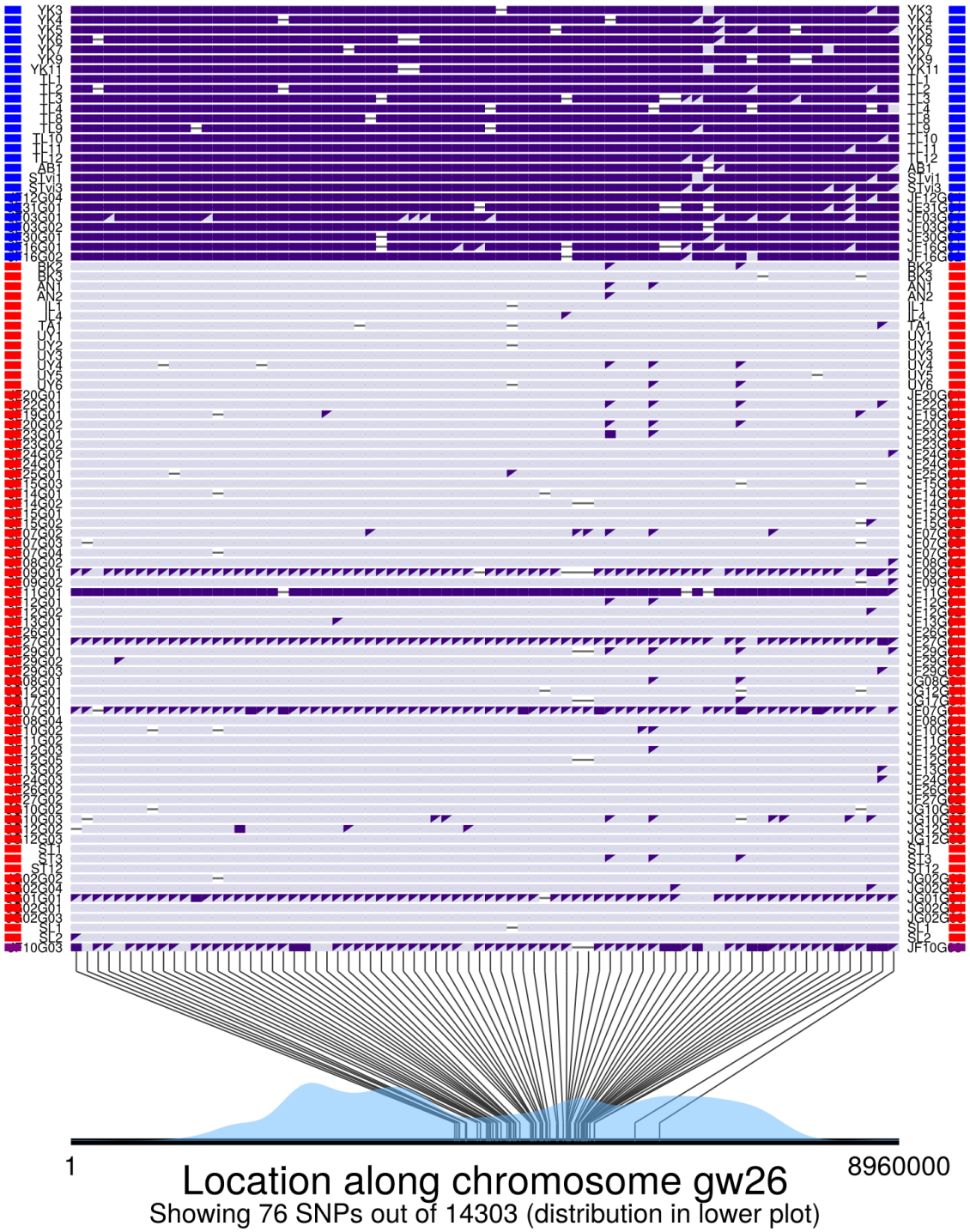
```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```


1
8435933
Location along chromosome gw23
Showing 29 SNPs out of 13949 (distribution in lower plot)

**choose another chromosome, and plot similarly to above**

```
chr = "gw26"
regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr]) #
plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff, missingFractionAllowed,
    regionInfo, pos_SNP_filtered, Fst, pairwiseNamesFst,
    genosOnly_included, ind_with_metadata_included, freqs, plotGroups, plotGroupColors; plotTitle = "")
if false  # set to true to save plot
    save("Figure4_upperright_gw26GBIplot_from_Julia.png", plotInfo[1], px_per_unit = 2.0)
end
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```

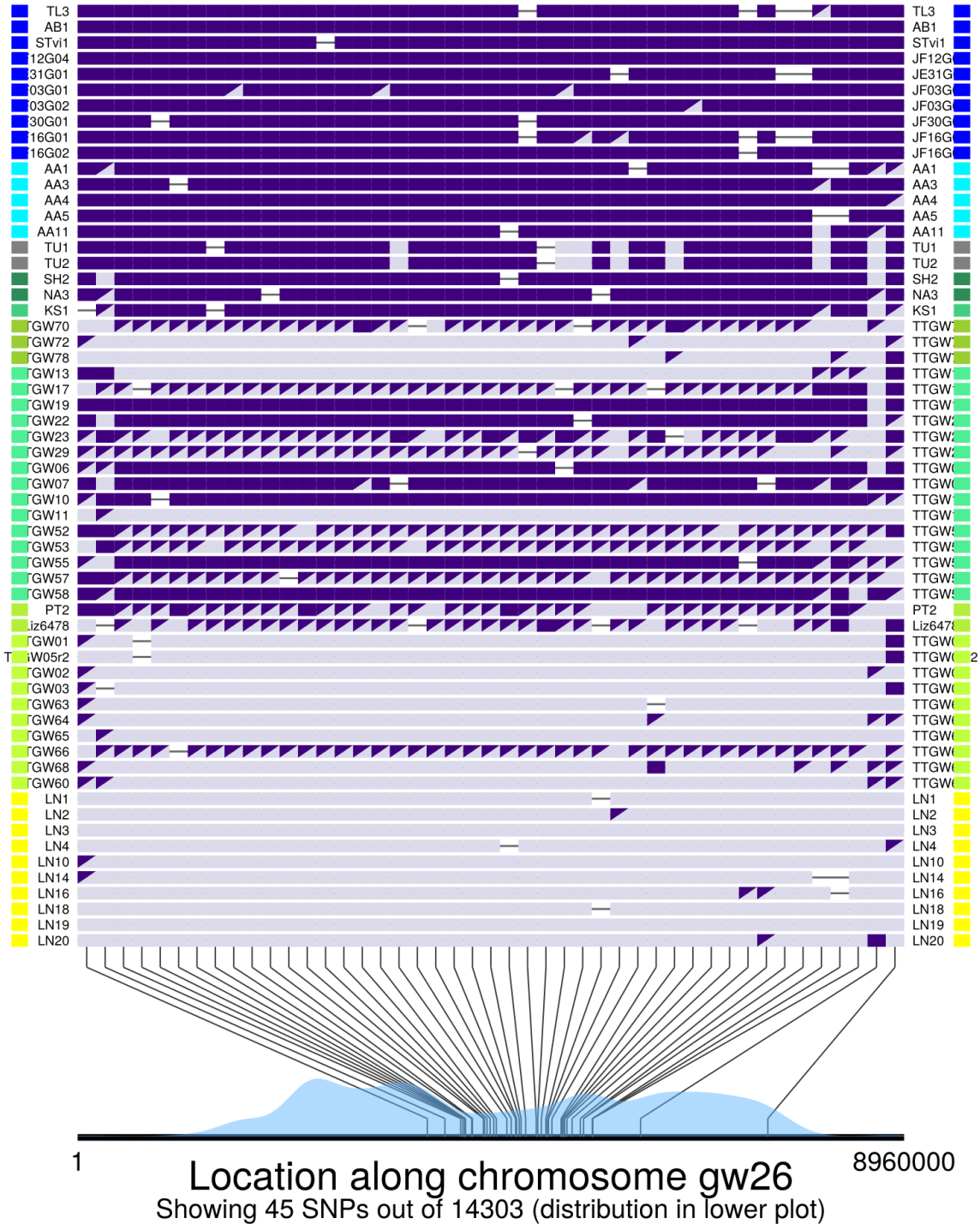

1
8960000
Location along chromosome gw26
Showing 76 SNPs out of 14303 (distribution in lower plot)

### Make a GBI plot to illustrate variation along west side of ring

```
groups = ["vir","troch_LN"] # for purpose of calculating pairwise Fst and Fst_group (to determine SNPs)
plotGroups = ["vir","vir_S","nit", "lud_PK", "lud_KS", "lud_central", "lud_Sath", "lud_ML","troch_west"
plotGroupColors = ["blue","turquoise1","grey","seagreen4","seagreen3","seagreen2","olivedrab3","olivedr
numIndsToPlot = [10, 5, 2, 3, 5, 15, 3, 5, 10, 10] # maximum number of individuals to plot from each gr
group1 = "vir"   # these groups will determine the color used in the graph
group2 = "troch_LN"
groupsToCompare = "vir_troch_LN" # "Fst_among"
Fst_cutoff = 0.9
missingFractionAllowed = 0.2  # only show SNPs with less than this fraction of missing data among indiv

# Calculate allele freqs and sample sizes (use column Fst_group)
freqs, sampleSizes = getFreqsAndSampleSizes(genosOnly_with_missing, ind_with_metadata_indFiltered.Fst_g
println("Calculated population allele frequencies and sample sizes")

# calculate Fst
Fst, FstNumerator, FstDenominator, pairwiseNamesFst = getFst(freqs, sampleSizes, groups; among=true)  #
println("Calculated Fst values")

# limit the individuals to include in plot
# For this figure only, filter out individuals with lots of missing genotypes
numMissings_threshold = 800_000
selection = ind_with_metadata_indFiltered.numMissings .< numMissings_threshold
genosOnly_with_missing_selected = view(genosOnly_with_missing, selection, :)
ind_with_metadata_indFiltered_selected = view(ind_with_metadata_indFiltered, selection, :)
# now limit each group to specified numbers
genosOnly_included, ind_with_metadata_included = limitIndsToPlot(plotGroups, numIndsToPlot, genosOnly_w

# choose the scaffold and region to show
chr = "gw26"
regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr])

#### Now actually make the plot
plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff, missingFractionAllowed,
    regionInfo, pos_SNP_filtered, Fst, pairwiseNamesFst,
    genosOnly_included, ind_with_metadata_included, freqs, plotGroups, plotGroupColors; plotTitle = "")
if false  # set to true to save plot
    save("Figure5_left_gw26GBIplotWest_from_Julia.png", plotInfo[1], px_per_unit = 2.0)
end
```

```
Calculated population allele frequencies and sample sizes
Calculated Fst values
```

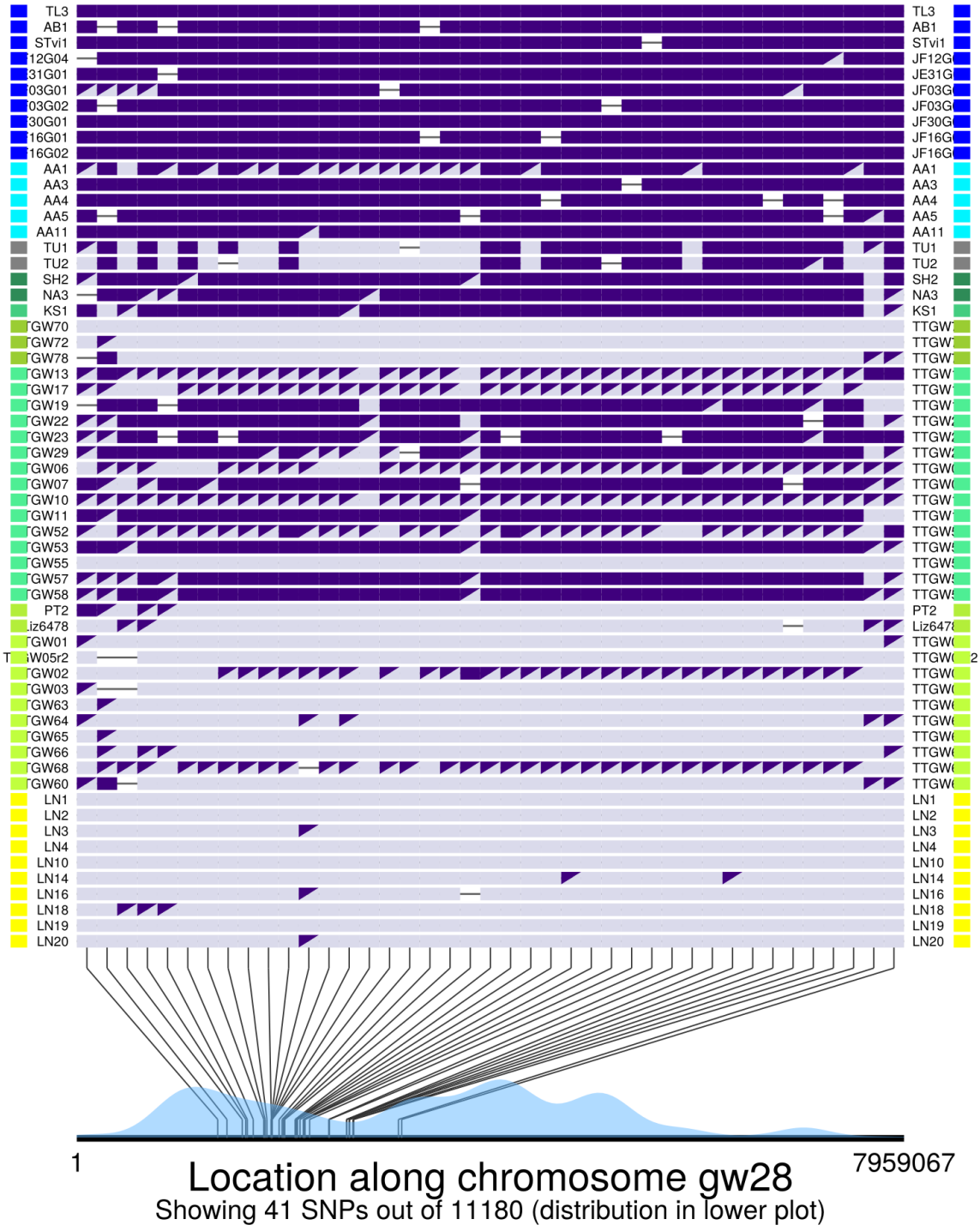
```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```




Show another chromosome along west side of ring:

```
# choose the scaffold and region to show
chr = "gw28"
regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr])

#### Now actually make the plot
plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff, missingFractionAllowed,
    regionInfo, pos_SNP_filtered, Fst, pairwiseNamesFst,
    genosOnly_included, ind_with_metadata_included, freqs, plotGroups, plotGroupColors; plotTitle = "")
if false  # set to true to save plot
    save("Figure5_right_gw28GBIplotWest_from_Julia.png", plotInfo[1], px_per_unit = 2.0)
end
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```


TL3
AB1
STvi1
12G04
31G01
03G01
03G02
30G01
16G01
16G02
AA1
AA3
AA4
AA5
AA11
TU1
TU2
SH2
NA3
KS1
GW70
GW72
GW78
GW13
GW17
GW19
GW22
GW23
GW29
GW06
GW07
GW10
GW11
GW52
GW53
GW55
GW57
GW58
PT2
iz6478
GW01
W05r2
GW02
GW03
GW63
GW64
GW65
GW66
GW68
GW60
LN1
LN2
LN3
LN4
LN10
LN14
LN16
LN18
LN19
LN20
1
7959067
Location along chromosome gw28
Showing 41 SNPs out of 11180 (distribution in lower plot)

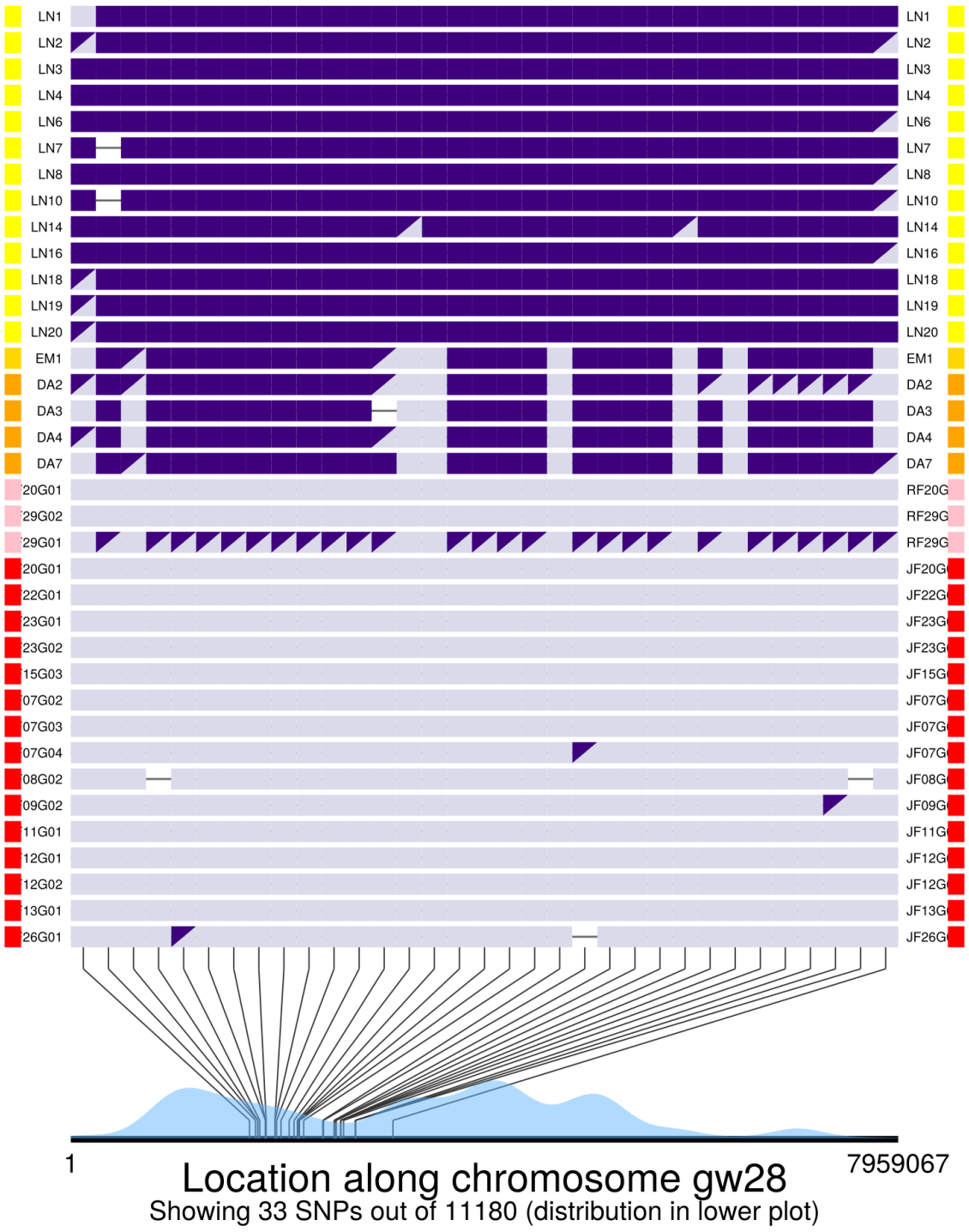
### Make a GBI plot to illustrate variation along east side of ring

```
groups = ["troch_LN","obs","plumb"] # for purpose of calculating pairwise Fst and Fst_group (to determi
plotGroups = ["troch_LN","troch_EM","obs","plumb_BJ","plumb"]
plotGroupColors = ["yellow","gold","orange","pink","red"]
numIndsToPlot = [15, 15, 15, 15, 17] # maximum number of individuals to plot from each group
group1 = "troch_LN"   # these groups will determine the color used in the graph
group2 = "plumb"
groupsToCompare = "troch_LN_plumb" #"troch_LN_plumb" #"Fst_among"
Fst_cutoff = 0.9
missingFractionAllowed = 0.2  # only show SNPs with less than this fraction of missing data among indiv

# Calculate allele freqs and sample sizes (use column Fst_group)
freqs, sampleSizes = getFreqsAndSampleSizes(genosOnly_with_missing, ind_with_metadata_indFiltered.Fst_g
println("Calculated population allele frequencies and sample sizes")

# calculate Fst
Fst, FstNumerator, FstDenominator, pairwiseNamesFst = getFst(freqs, sampleSizes, groups; among=true)  #
println("Calculated Fst values")

# limit the individuals to include in plot
# For this figure only, filter out individuals with lots of missing genotypes
numMissings_threshold = 800_000
selection = ind_with_metadata_indFiltered.numMissings .< numMissings_threshold
genosOnly_with_missing_selected = view(genosOnly_with_missing, selection, :)
ind_with_metadata_indFiltered_selected = view(ind_with_metadata_indFiltered, selection, :)

# now limit each group to specified numbers
genosOnly_included, ind_with_metadata_included = limitIndsToPlot(plotGroups, numIndsToPlot, genosOnly_w

# choose the scaffold and region to show
chr = "gw28"
regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr]) #

# need to remove two plumbeitarsus individuals that have introgression from viridanus, as that would ot
selection = Not(ind_with_metadata_included.ind .∈ Ref(["GW_Armando_plate1_JF09G01", "GW_Armando_plate1_
ind_with_metadata_included = ind_with_metadata_included[selection,:]
genosOnly_included = genosOnly_included[selection,:]

#### Now actually make the plot
plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff, missingFractionAllowed,
    regionInfo, pos_SNP_filtered, Fst, pairwiseNamesFst,
    genosOnly_included, ind_with_metadata_included, freqs, plotGroups, plotGroupColors; plotTitle = "")
```

```julia
if false  # set to true to save plot
    save("Figure6_top_gw28GBIplotEast_from_Julia.png", plotInfo[1], px_per_unit = 2.0)
end
```

```
Calculated population allele frequencies and sample sizes
Calculated Fst values
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```


LN1
LN2
LN3
LN4
LN6
LN7
LN8
LN10
LN14
LN16
LN18
LN19
LN20
EM1
DA2
DA3
DA4
DA7
20G01
29G02
29G01
20G01
22G01
23G01
23G02
15G03
07G02
07G03
07G04
08G02
09G02
11G01
12G01
12G02
13G01
26G01
RF20G
RF29G
RF29G
JF20G
JF22G
JF23G
JF23G
JF15G
JF07G
JF07G
JF07G
JF08G
JF09G
JF11G
JF12G
JF12G
JF13G
JF26G
1
7959067
Location along chromosome gw28
Showing 33 SNPs out of 11180 (distribution in lower plot)

### Make GBI plots showing all individuals in the study

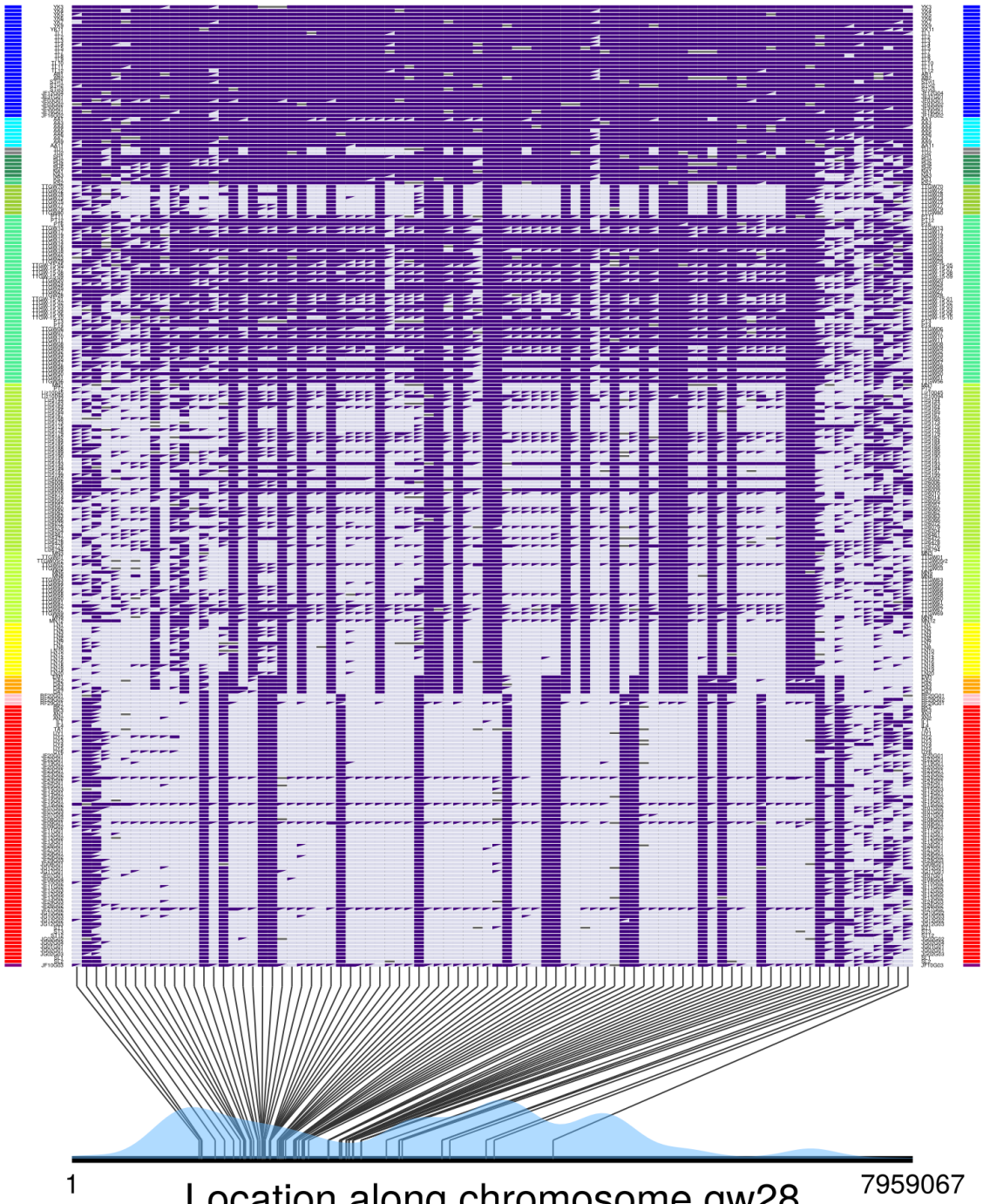
```
groups_to_plot_all = ["vir","vir_S","nit", "lud_PK", "lud_KS", "lud_central", "lud_Sath", "lud_ML","tro
group_colors_all = ["blue","turquoise1","grey","seagreen4","seagreen3","seagreen2","olivedrab3","olived
```

This will use Fst among multiple groups to determine SNPs to plot. For chr 28:

```
groups = ["vir","lud_PK","troch_LN","obs","plumb"] # for purpose of calculating pairwise Fst and Fst_gr
plotGroups = groups_to_plot_all
plotGroupColors = group_colors_all
group1 = "vir"   # these groups will determine the color used in the graph
group2 = "plumb"
groupsToCompare = ["vir_plumb", "vir_troch_LN", "troch_LN_plumb"]     # "Fst_among" #"vir_plumb"
Fst_cutoff = 0.8
missingFractionAllowed = 0.2  # only show SNPs with less than this fraction of missing data among indiv

# Calculate allele freqs and sample sizes (use column Fst_group)
freqs, sampleSizes = getFreqsAndSampleSizes(genosOnly_with_missing, ind_with_metadata_indFiltered.Fst_g
println("Calculated population allele frequencies and sample sizes")

Fst, FstNumerator, FstDenominator, pairwiseNamesFst = getFst(freqs, sampleSizes, groups; among=true)  #
println("Calculated Fst values")

chr = "gw28"
regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr]) #

plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff,
    missingFractionAllowed, regionInfo,
    pos_SNP_filtered, Fst, pairwiseNamesFst,
    genosOnly_with_missing, ind_with_metadata_indFiltered, freqs,
    plotGroups, plotGroupColors;
    indFontSize=4, figureSize=(1200,1600),
    plotTitle = "");
# plotInfo contains a tuple with: (f, plottedGenotype, locations, plottedMetadata)
```

```
Calculated population allele frequencies and sample sizes
Calculated Fst values
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```


1
7959067
Location along chromosome gw28
Showing 86 SNPs out of 11180 (distribution in lower plot)

## Make list of scaffolds to plot according to above:

```
scaffolds_to_plot = "gw" .* string.(vcat(28:-1:17, 15:-1:1))
push!(scaffolds_to_plot, "gw1A", "gw4A")  # add two other scaffolds
```

```
29-element Vector{String}:
 "gw28"
 "gw27"
 "gw26"
 "gw25"
 "gw24"
 "gw23"
 "gw22"
 "gw21"
 "gw20"
 "gw19"
 "gw18"
 "gw17"
 "gw15"
 ⋮
 "gw10"
 "gw9"
 "gw8"
 "gw7"
 "gw6"
 "gw5"
 "gw4"
 "gw3"
 "gw2"
 "gw1"
 "gw1A"
 "gw4A"
```

Loop through each scaffold and produce GBI plot. (Making inactive because already saved figs and takes long.)

```
for i in 1:length(scaffolds_to_plot)
    chr = scaffolds_to_plot[i]
    regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr
    plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff,
```

```
        missingFractionAllowed, regionInfo,
        pos_SNP_filtered, Fst, pairwiseNamesFst,
        genosOnly_with_missing, ind_with_metadata_indFiltered, freqs,
        plotGroups, plotGroupColors;
        indFontSize=4, figureSize=(1200,1600), plotTitle = "");
    println("Completed the figure for ", chr, ".")
    if true  # set to true to save plot
        filename = string("Figure_", chr, "_Fst3groups_GBI_allInds_from_Julia.png")
        save(filename, plotInfo[1], px_per_unit = 2.0)
        println("Saved ", filename)
    end
end
```

## Estimate relationships of individuals using PCA

Our goal is to produce plots showing individuals in genotype space, using Principal Components Analysis.

### Impute and save genotypes for each scaffold

PCA requires imputation of missing genotypes. I did imputation for each scaffold above a certain size threshold. These are listed in `chromosomes_to_process` .

Imputation can take several minutes per scaffold, so I ran this imputation step separately from this Quarto notebook (otherwise render would take long) and saved the genotype data for each scaffold for loading in the next step. Below is the code I initially used for imputing, BUT IT USES THE SVD ALGORITHM AND I LATER DECIDED THAT KNN IS BETTER (BUT IN OCT2024 DECIDED SVD IS BEST HA HA, but it takes a lot longer–now changing my mind back, that K=1 and KNN is best because nitidus has only 2 inds):

```
julia  # DON'T RUN THIS AS IT USES SVD IMPUTING (NOT KNN--SEE BELOW FOR THAT) #=
for i in eachindex(chromosomes_to_process)        chrom = chromosomes_to_process[i]
regionText = string("chr", chrom)         loci_selection = (pos_SNP_filtered.chrom
.== chrom)            pos_SNP_filtered_region = pos_SNP_filtered[loci_selection,:]
genosOnly_region_for_imputing = Matrix{Union{Missing, Float32}}(genosOnly_with_missing[:,loci_selection]
@time imputed_genos = Impute.svd(genosOnly_region_for_imputing)           filename
= string(baseName, tagName, regionText, ".SVDimputedMissing.jld2")           jld-
save(filename; imputed_genos, ind_with_metadata_indFiltered, pos_SNP_filtered_region)
println(string("Chromosome ", chrom, ": Saved real and imputed genotypes for ",
size(pos_SNP_filtered_region, 1)," SNPs and ", size(genosOnly_region_for_imputing, 1),"
filtered individuals.")) end =#
```

Now we can cycle through a set of chromosomes and plot a PCA for each. We need to first specify some groups to include in the plot, and their colors:

```
groups_to_plot_PCA = ["vir","vir_S","nit", "lud_PK", "lud_KS", "lud_central", "lud_Sath", "lud_ML","tro
group_colors_PCA = ["blue","turquoise1","grey","seagreen4","seagreen3","seagreen2","olivedrab3","olived
```

Now we'll actually do the PCA and make the plot for each scaffold. The code block below is what I initially used, based on the SVD method of imputing. I've now decided that KNN is better (see further below)–CHANGED MIND AGAIN 1NOV2024–changed back to KNN and K=1 in Jan2025.

MAKING INACTIVE FOR NOW, BECAUSE MAKES A LOT OF PLOTS:

```
#= for i in eachindex(chromosomes_to_process)
    chrom = chromosomes_to_process[i]
    regionText = string("chr", chrom)
    filename = string(baseName, tagName, regionText, ".SVDimputedMissing.jld2")
    imputed_genos = load(filename, "imputed_genos")
    ind_with_metadata_indFiltered = load(filename, "ind_with_metadata_indFiltered")
    pos_SNP_filtered_region = load(filename, "pos_SNP_filtered_region")
    println(string("Loaded ", filename))
    println(string(regionText, ": ", size(imputed_genos, 2), " SNPs from ", size(imputed_genos, 1), " i
    flipPC1 = true
    flipPC2 = true
    PCAmodel = plotPCA(imputed_genos, ind_with_metadata_indFiltered,
        groups_to_plot_PCA, group_colors_PCA;
        sampleSet="greenish warblers", regionText=regionText,
        flip1=flipPC1, flip2=flipPC2,
        showPlot=false)
    # add position of reference genome
    refGenomePCAposition = predict(PCAmodel.model, zeros(size(imputed_genos, 2)))
    flipPC1 && (refGenomePCAposition[1] *= -1)  # this flips PC1 if flipPC1 = true
    flipPC2 && (refGenomePCAposition[2] *= -1)  # same for PC2
    CairoMakie.scatter!(refGenomePCAposition[1], refGenomePCAposition[2], marker=:diamond, color="black
    try
        display(PCAmodel.PCAfig)
    catch
        println("NOTICE: Figure for ", regionText, " could not be shown due to an unknown error.")
    end
end =#
```

### Imputation using KNN

Troyanskaya *et al.* (2001) recommend imputation using K-nearest neighbors approach as being better than SVD, both of which are better than other methods for DNA genotyping. They also recommend using Euclidian distance. So I will try KNN with Euclidian distance, which like SVD is provided by Impute.jl. I am going with setting `dims` to `:rows`, as that seems to run much faster and produces PCAs that make a lot of sense. I've already run this next code cell, which does the imputing and saves the imputed data matrix for each scaffold. This actually runs quite quickly–at most a couple minutes per scaffold. :)

```
for i in eachindex(chromosomes_to_process)
    chrom = chromosomes_to_process[i]
    regionText = string("chr", chrom)
    loci_selection = (pos_SNP_filtered.chrom .== chrom)
    pos_SNP_filtered_region = pos_SNP_filtered[loci_selection,:]
    genosOnly_region_for_imputing = Matrix{Union{Missing, Float32}}(genosOnly_with_missing[:,loci_selec
    @time imputed_genos = Impute.knn(genosOnly_region_for_imputing; dims = :rows) # much faster with th
    filename = string(baseName, tagName, regionText, ".KNNimputedMissing.jld2")
    jldsave(filename; imputed_genos, ind_with_metadata_indFiltered, pos_SNP_filtered_region)
    println(string("Chromosome ", chrom, ": Saved real and imputed genotypes for ", size(pos_SNP_filter
end
```

Now do the KNN PCA:

MAKING INACTIVE FOR NOW, BECAUSE MAKES A LOT OF PLOTS:

```
for i in eachindex(chromosomes_to_process)
    scaffold = chromosomes_to_process[i]
    regionText = string("chr", scaffold)
    filename = string(baseName, tagName, regionText, ".KNNimputedMissing.jld2")
    imputed_genos = load(filename, "imputed_genos")
    ind_with_metadata_indFiltered = load(filename, "ind_with_metadata_indFiltered")
    pos_SNP_filtered_region = load(filename, "pos_SNP_filtered_region")
    println(string("Loaded ",filename))
    println(string(regionText, ": ", size(imputed_genos,2), " SNPs from ", size(imputed_genos,1), " ind
    flipPC1 = true
    flipPC2 = true
    PCAmodel = plotPCA(imputed_genos, ind_with_metadata_indFiltered,
            groups_to_plot_PCA, group_colors_PCA;
            sampleSet = "greenish warblers", regionText=regionText,
            flip1 = flipPC1, flip2 = flipPC2,
            lineOpacity = 0.7, fillOpacity = 0.6,
            symbolSize = 14, showTitle = true,
            xLabelText = string("Chromosome ", scaffold," PC1"), yLabelText = string("Chromosome ", sca
            showPlot = false)
```

```
    # add position of reference genome
    refGenomePCAposition = predict(PCAmodel.model, zeros(size(imputed_genos, 2)))
    flipPC1 && (refGenomePCAposition[1] *= -1)  # this flips PC1 if flipPC1 = true
    flipPC2 && (refGenomePCAposition[2] *= -1)  # same for PC2
    CairoMakie.scatter!(refGenomePCAposition[1], refGenomePCAposition[2], marker = :diamond, color="bla
    try
        display(PCAmodel.PCAfig)
    catch
        println("NOTICE: Figure for ", regionText, " could not be shown due to an unknown error.")
    end
end
```

## Make GBI plots showing just west individuals:

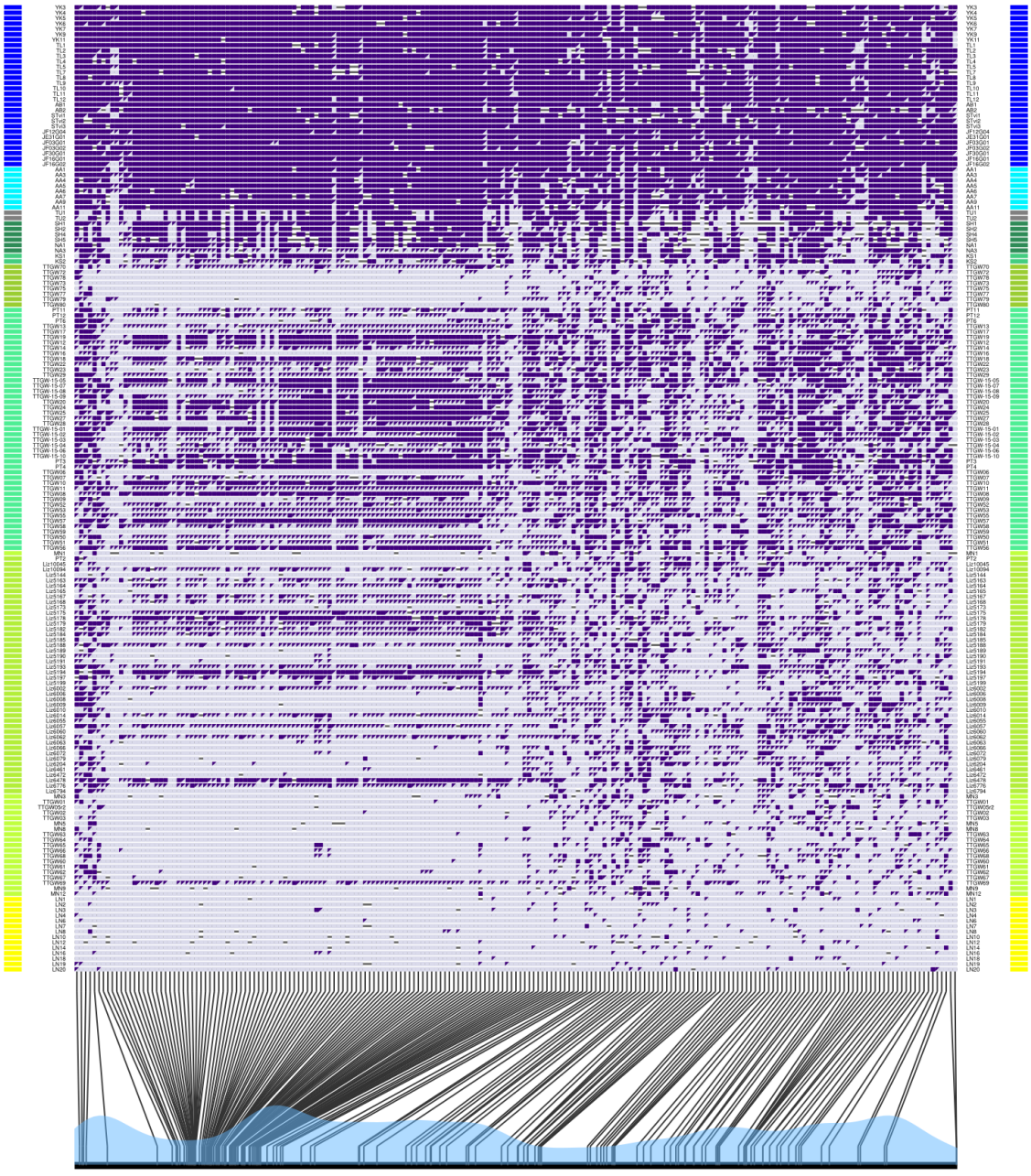
```
groups = ["vir","lud_PK","troch_LN"] # for purpose of calculating pairwise Fst and Fst_group (to determ
plotGroups = ["vir", "vir_misID", "vir_S", "nit", "lud_PK", "lud_KS", "lud_central", "lud_Sath", "lud_M
plotGroupColors = ["blue", "blue", "turquoise1", "grey", "seagreen4", "seagreen3", "seagreen2", "olived
group1 = "vir"   # these groups will determine the color used in the graph
group2 = "troch_LN"
groupsToCompare = "Fst_among" #"vir_plumb"
Fst_cutoff = 0.8
missingFractionAllowed = 0.2  # only show SNPs with less than this fraction of missing data among indiv

# Calculate allele freqs and sample sizes (use column Fst_group)
freqs, sampleSizes = getFreqsAndSampleSizes(genosOnly_with_missing, ind_with_metadata_indFiltered.Fst_g
println("Calculated population allele frequencies and sample sizes")

Fst, FstNumerator, FstDenominator, pairwiseNamesFst = getFst(freqs, sampleSizes, groups; among=true)  #
println("Calculated Fst values")

scaffolds_to_plot = ["gw1"]

for i in 1:length(scaffolds_to_plot)
    chr = scaffolds_to_plot[i]
    regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr
    plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff,
        missingFractionAllowed, regionInfo,
        pos_SNP_filtered, Fst, pairwiseNamesFst,
        genosOnly_with_missing, ind_with_metadata_indFiltered, freqs,
        plotGroups, plotGroupColors;
```

```
        indFontSize=4, figureSize=(1200,1600))
    println("Completed the figure for ", chr, ".")
end

#for chr in scaffolds_to_plot
#    println(chr)
#end
# plotInfo contains a tuple with: (f, plottedGenotype, locations, plottedMetadata)
```

```
Calculated population allele frequencies and sample sizes
Calculated Fst values
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```

# 1 1 to 121941280: genotypes Fst>0.8 loci between Fst_



```
Completed the figure for gw1.
```

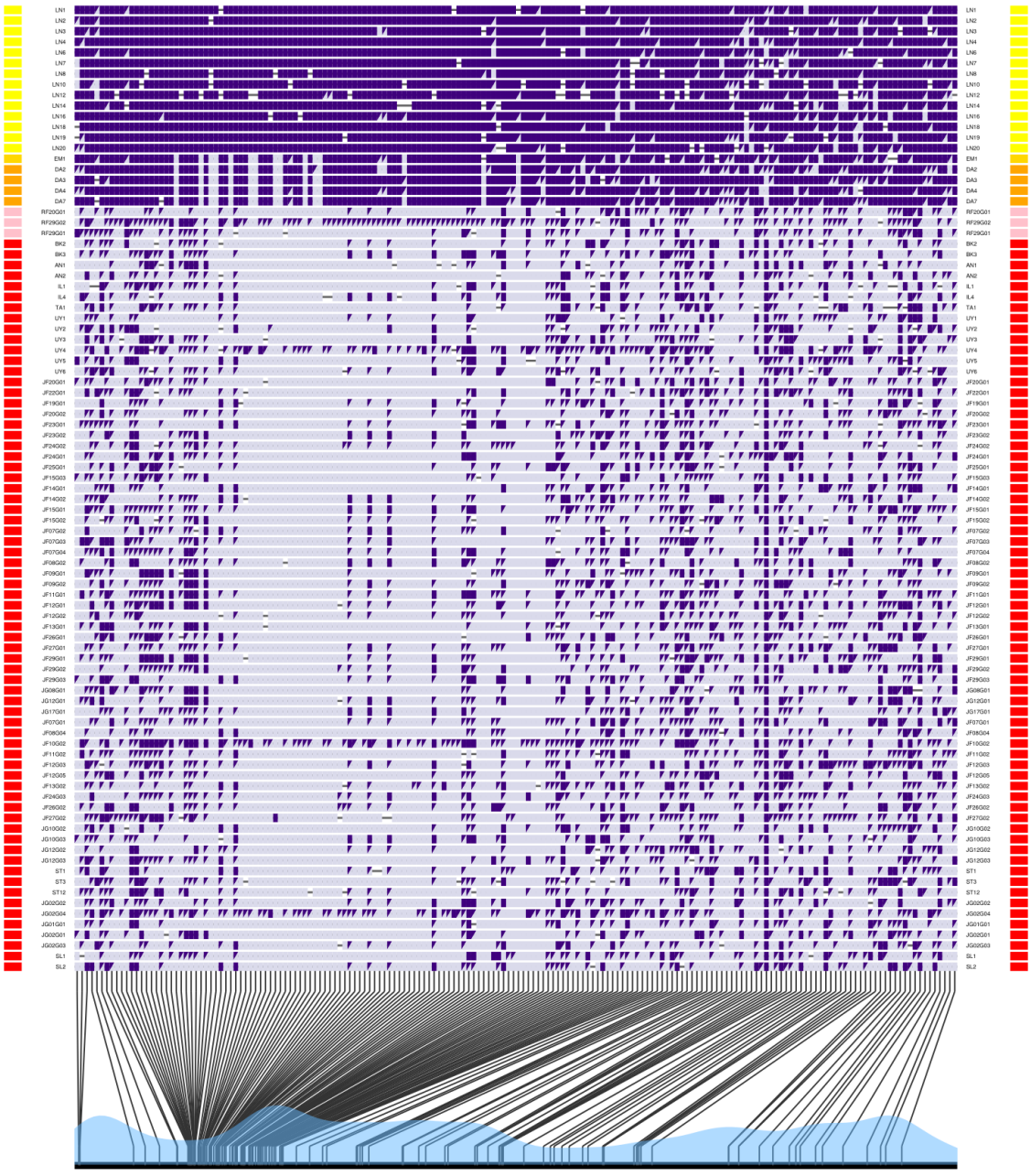
## Make GBI plots showing just east individuals:

```
groups = ["troch_LN","obs","plumb"] # for purpose of calculating pairwise Fst and Fst_group (to determi
plotGroups = ["troch_LN", "troch_EM", "obs", "plumb_BJ", "plumb"] # east GW individuals
plotGroupColors = ["yellow", "gold", "orange", "pink", "red"]
group1 = "troch_LN"   # these groups will determine the color used in the graph
group2 = "plumb"
groupsToCompare = "Fst_among" #"vir_plumb"
Fst_cutoff = 0.8
missingFractionAllowed = 0.2  # only show SNPs with less than this fraction of missing data among indiv

# Calculate allele freqs and sample sizes (use column Fst_group)
freqs, sampleSizes = getFreqsAndSampleSizes(genosOnly_with_missing, ind_with_metadata_indFiltered.Fst_g
println("Calculated population allele frequencies and sample sizes")

Fst, FstNumerator, FstDenominator, pairwiseNamesFst = getFst(freqs, sampleSizes, groups; among=true)  #
println("Calculated Fst values")

for i in 1:length(scaffolds_to_plot)
    chr = scaffolds_to_plot[i]
    regionInfo = chooseChrRegion(pos_SNP_filtered, chr; positionMin=1, positionMax=scaffold_lengths[chr
    plotInfo = plotGenotypeByIndividualWithFst(groupsToCompare, Fst_cutoff,
        missingFractionAllowed, regionInfo,
        pos_SNP_filtered, Fst, pairwiseNamesFst,
        genosOnly_with_missing, ind_with_metadata_indFiltered, freqs,
        plotGroups, plotGroupColors;
        indFontSize=4, figureSize=(1200,1600))
    println("Completed the figure for ", chr, ".")
end
# plotInfo contains a tuple with: (f, plottedGenotype, locations, plottedMetadata)
```

```
Calculated population allele frequencies and sample sizes
Calculated Fst values
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `Scene`s and
└ @ Makie ~/.julia/packages/Makie/pFPBw/src/scenes.jl:238
```

# 1 1 to 121941280: genotypes Fst>0.8 loci between Fst_



Completed the figure for gw1.