Methods

Building the Reference Database: Our Approach and Process
Publicly available molecular markers were deployed for identifying Arabica coffee varieties, from a set of 96 Single nucleotide polymorphism (SNP) markers recently published by Zhang et al. (2021). From this set of markers, we refined our selection to a subset of 41 markers that most differentiated between multiple coffee varieties and exhibited efficient high-throughput genotyping performance. Various sources of the most commonly cultivated Arabica coffee varieties were sampled in Latin America and Africa. These samples were then sent to Intertek/Agritech (https://www.intertek.com/agriculture/agritech/), to genotype the samples. Once the genotyping process was completed, we established the SNP fingerprints of the samples and developed the reference database.

Due to the incomplete inbreeding of most Arabica coffee varieties at the time of their release, in some cases, there is considerable intra-variety variation. Additionally, certain varieties with the same genetic backgrounds and parentage but were given different names in different countries or regions, making the discrimination of these varieties impossible. This is further complicated by the fact that closely related varieties often exhibit very similar fingerprint marker patterns. As a result, we opted to collect samples from reliable sources and genotype multiple samples of the same variety to establish the most precise SNP fingerprint profile. For that reason, a reference database was developed from multiple samples per variety, to capture and use that variation to better aid the identification and verification process.

After the reference database was completed, we used the R-package assignPOP developed by (Chen et al., 2018) to identify varieties or variety groups from the SNP markers database. This package facilitates the assignment of populations from genetic datasets using a machine-learning approach. Key features of this package include principal component analysis (PCA) for dimensionality reduction, Monte-Carlo cross-validation for assignment accuracy estimation, and K-folds cross-validation for membership probability estimation.

It is important to note that assignPOP is not the only package that can be used for variety assignment. There are other packages such as STRUCTURE, Discriminant Analysis of Principal Components (DAPC), Structure, and LEA (Landscape and Ecological Association Studies), among others that also enable the assignment of populations based on markers data. The reason we use assignPOP is due to its internal machine-learning validation approaches, effective discrimination power in predicting sources of populations, ability to handle low-SNP data, and ability to process relatively large datasets.

Conducting the Variety ID Assessment with Genotypic Data

Once the samples are genotyped these data will be used to conduct the variety ID assessment with the assignPOP R-package. For more details regarding the assignPOP R-package please visit: https://alexkychen.github.io/assignPOP/

In the following section, we provide a detailed guide on how to carry out the variety assessment using the R studio software (R Studio Team, 2020).

R-Script considerations

First work on the reference database:

To access the Excel reference database (low-density SNP marker panel), please download it from the WCR resources page. Save the file in a designated working folder on your computer. The table below provides an overview of the reference database's structure. Please note that in the R-script you will need an identification column giving the individual names of the samples and a population column giving the population of each individual.

Example reference database

id	TargetCountries	testAna14	Variety	Pedigree	Pop	Subpop	Ca0468	Ca0516	Ca0711	Ca2024	Ca2044	Ca2285	Ca3052
S747	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S748	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S749	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S750	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S751	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S752	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S753	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S754	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S755	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S756	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S757	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S758	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S759	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S760	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S761	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S762	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S763	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S764	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S765	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S766	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S768	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S769	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S770	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S771	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S772	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S773	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S774	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C
S775	LatAm	Y	Anacafe14	(Timor Hybrid 832/1 x Caturra) x Pacamara	Anacafe14	Anacafe14	C:C	T:T	G:G	A:A	G:G	T:T	C:C

Reference database to compare variety samples

Depending on the variety assigning R-package better results are achieved by using a subset of the reference varieties, as having many fills the genetic space and decreases the power of the analyses. Especially in the case of varieties that are genetically similar, but were given different names in different countries. In general, it is important to include varieties representative of the broad diversity while also including varieties that are known candidate varieties for possible confounders (neighbor varieties at the seed production level, or at the nursery). To address this concern, we developed a table that lists each variety in the reference panel for comparison. This enables the creation of appropriate subsets of references, mitigating possible confounding results when generating the assigning analysis. Please note that this table was created based on previous knowledge of the genetic background of the varieties and this table and its

discretion rely on WCR internal expertise, which involved the analysis of multiple samples from the listed varieties in this reference dataset.

Make sure you use the subset of the reference varieties accordingly.

R- Script
[Link to the R script]

Install required packages if not already installed

```
required_packages <- c("readxl", "adegenet", "graph4lg", "assignPOP", "openxlsx")
new_packages <- required_packages[!(required_packages %in%
installed.packages()[,"Package"])]
if (length(new_packages) > 0) {
  install.packages(new_packages)
}
```

Load required libraries and sources

```
library(readxl)
library("adegenet")

library(graph4lg)

library(assignPOP)
library(openxlsx)
```

Set the working directory

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

Read the reference file

Make sure you use the subset of the reference varieties accordingly. You have the flexibility to subset the reference database according to your requirements. For example, you may want to include all reference varieties in a single database and then create subsets based on preferred varieties for assessment. To subset the 'snp' data frame, we will use the subset() function with specific criteria.

```
snp <- read.csv("FinalWCRFullReferencesLDPA-07-03-23.csv")
snp<- subset(snp, testAna14=="Y")
#only use the first 45 SNPs that we routinely use
snp2 <- snp[,c(10:54)] # Make sure you use only the columns where the markers are in the
database. In the example code line, the markers are located between columns 10 and 54
```

Read the file in adegenet and convert the reference file to genind object

?df2genind *#The function df2genind converts a data.frame (or a matrix) into a genind object.* obj <- **df2genind**(snp2, ploidy **=** 2, NA.char **=** "", sep **=** ":", ind.names **=** snp**$**id, pop **=** snp**$**Variety) *# Remember that this code is used for the references and the ind.names= and pop = have to match according to the database. In the example database the ind.names=snp$id, the id is the vector giving the individuals names in the database named snp. The pop = snp$Variety, is the vector giving the population of each individual.*

Convert genind object to genepop format

genind_to_genepop(obj, output **=** "Referencetogenepop_test.txt") *# In this case we named the output as "Referencetogenepop.txt" You can name it as you wish.*

Replace "000000" with "0000" in Referencetogenepop1.txt

IMPORTANT - When the file is converted it makes NAs as 000000 (six zeroes), while it needs 0000 (four zeroes). Go to the file produced in the folder 'Referencetogenepop_test.txt' and replace 000000 to 0000

Using AssignPop

https://alexkychen.github.io/assignPOP/. It is important to put the `pop.name` in the alphabetic order

Read reference genepop file

YourGenepoprefs <- **read.Genepop**("Referencetogenepop_test.txt", haploid **=** FALSE, pop.names**=**c("Anacafe14", "Catigua Mg2", "Centroamericano", "Marsellesa", "Pacamara", "PacasTekisic", "SanPacho")) *# In the script, these references have to be listed in alphabetic order: If not sure of the order look at the Referencetogenepop.txt created in the step before*

Next step is predicting the samples

Example of the samples table

PLATE_ID	Variety	Code	id	Ca0516	Ca0711	Ca2024	Ca2044	Ca2285	Ca3165	Ca3177	Ca3457	Ca3594	Ca3700	Ca3728	Ca3794	Ca397(
SE-23-0920-06_SLV-MH-12	A01	SLV-MH-Anacafe14-01	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G		T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A02	SLV-MH-Anacafe14-02	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A03	SLV-MH-Anacafe14-03	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A04	SLV-MH-Anacafe14-04	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A05	SLV-MH-Anacafe14-05	Anacafe14-0	T:T	C:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A06	SLV-MH-Anacafe14-06	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A07	SLV-MH-Anacafe14-07	Anacafe14-0	T:T	C:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A08	SLV-MH-Anacafe14-08	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A09	SLV-MH-Anacafe14-09	Anacafe14-0	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A10	SLV-MH-Anacafe14-10	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A11	SLV-MH-Anacafe14-11	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	A12	SLV-MH-Anacafe14-12	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B01	SLV-MH-Anacafe14-13	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B02	SLV-MH-Anacafe14-14	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B03	SLV-MH-Anacafe14-15	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B04	SLV-MH-Anacafe14-16	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B05	SLV-MH-Anacafe14-17	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B06	SLV-MH-Anacafe14-18	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B07	SLV-MH-Anacafe14-19	Anacafe14-1	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B08	SLV-MH-Anacafe14-20	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B09	SLV-MH-Anacafe14-21	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B10	SLV-MH-Anacafe14-22	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B11	SLV-MH-Anacafe14-23	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	B12	SLV-MH-Anacafe14-24	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	C01	SLV-MH-Anacafe14-25	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	C02	SLV-MH-Anacafe14-26	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	C03	SLV-MH-Anacafe14-27	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C
SE-23-0920-06_SLV-MH-12	C04	SLV-MH-Anacafe14-28	Anacafe14-2	T:T	G:G	A:A	G:G	T:T	G:G	G:G	T:T	C:C	G:G	A:A	T:T	C:C

IMPORTANT: Make sure the heterozygous markers in the samples database are in the same allele order as the reference database.

Het marker in ref. DB	Het marker in Sample DB	Het marker in the sample DB after modification
A:C	C:A	A:C
A:G	G:A	A:G
A:T	T:A	A:T
C:G	G:C	C:G
C:T	T:C	C:T
G:T	T:G	G:T

If the heterozygous markers are not in the same allele order as the reference database, then order the markers accordingly. (I.e, if the marker in the reference database is A:C and the marker in the samples database is C:A, replace C:A for A:C in the samples database)

Read sample file

```
samples <- read.csv("Rawdata 1-16 149.016_report.csv")

samples2 <- samples[, 5:45] # Make sure you use only the columns where the markers are in
the database. In the example code line, the markers are located between columns 4 and 45
```

```
samplesg <- df2genind(snp2, ploidy = 2, NA.char = "", sep = ":", ind.names = snp$id, pop =
snp$Variety) # Remember that this code is used for the references and the ind.names= and
pop = have to match according to the database. In the example database the
ind.names=snp$id, the id is the vector giving the individuals names in the database named snp.
The pop = snp$Variety, is the vector giving the population of each individual.

Samplesg
```

Assess loci and genotype missing data

```
#?? missingno

samplesg <- missingno(samplesg, type = "loci", cutoff = 0.3, quiet = FALSE, freq = FALSE) #
Confirm whether the markers are missing in all genotyped plates and varieties. If any markers
have missing data across all plates and varieties, consider eliminating the data for those
markers accordingly. In this code, the threshold for removal is set at 30% of missing data.

samplesg <- missingno(samplesg, type = "geno", cutoff = 0.3, quiet = FALSE, freq = FALSE)
#Confirm whether the genotypes are missing in all genotyped plates and varieties. If any
genotype has missing data across all plates and varieties, consider eliminating the data for
those genotypes accordingly. In this code, the threshold for removal is set at 30% of missing
data.

samplesg
```

Convert genind object to genepop format

IMPORTANT - When the file is converted it makes NAs as 000000 (six zeroes), while it it needs
0000 (four zeroes).
```
genind_to_genepop(samplesg, output = "WCR_AssigPop_test.txt")

# Replace "000000" with "0000" in WCR_AssigPop_test.txt

#replace_zeros("WCR_AssigPop_test.txt")
```

Read unknown individuals in the genepop file

```
YourGenepopunknown <- read.Genepop("WCR_AssigPop_test.txt",  haploid = FALSE)
```

Perform assignment test

The `assign.X` function utilizes markers data, along with known individuals (in this case the
references), to assign the unknown individuals to possible source populations. The results are
then saved as an 'AssignmentResult.txt' file.

?assign.X

```r
assign.X( x1=YourGenepoprefs, x2=YourGenepopunknown, dir="ResultTest/",
model="naiveBayes") #The files are saved in the dir="ResultTest/" folder. Please make sure you
create the folder before running this codel line. If you have created the folder make sure this is
empty before you run this script line.
```

Read assignment result from AssignmentResult.txt

```r
assignment_data <- read.table("ResultTest/AssignmentResult.txt",  header = FALSE)
```

After executing the script, the resulting 'AssignmentResult.txt' file needs to be opened and saved as either a .csv or .xls file. The output table will include the predicted variety and the corresponding membership probabilities, similar to the sample table provided below. It is essential to understand that when the expected variety does not match the predicted variety, it should not be assumed that the predicted variety accurately represents the sample. For instance, in the table, the expected variety of the samples was "Anacafe14," and all the trees except tree 2 had their predicted varieties matching the expected variety. However, for tree 2, the predicted variety was "Marsellesa." Due to the closely related and similar SNP profiles of Arabica coffee varieties, the assignment of "Marsellesa" is not entirely certain. So, this assignment test is only for "yes" or "no" results.

Code	Ind.ID_well	Variety	Plant	pred.pop	Anacafe14	CatiguaMg2	Centroamericano	Icatu	Marsellesa	Pacamara	PacasTekisic	SanPacho
A01Anacafe141	A01	Anacafe14	1	Anacafe14	1	4.87E-142	8.61E-36	2.28E-12	8.16E-83	2.68E-26	2.16E-20	7.58E-56
A02Anacafe142	A02	Anacafe14	2	Marsellesa	4.46E-29	2.08E-253	1.18E-41	0.014033485	0.985965356	9.83E-19	1.16E-06	2.31E-19
A03Anacafe143	A03	Anacafe14	3	Anacafe14	1	7.60E-160	3.99E-43	4.92E-15	8.33E-82	2.66E-30	2.94E-26	1.17E-62
A04Anacafe144	A04	Anacafe14	4	Anacafe14	1	4.59E-175	1.06E-31	4.06E-17	4.26E-97	7.57E-29	5.42E-24	6.92E-71
A05Anacafe145	A05	Anacafe14	5	Anacafe14	1	5.75E-152	1.21E-32	6.10E-15	2.29E-90	2.08E-27	3.14E-23	2.05E-62
A06Anacafe146	A06	Anacafe14	6	Anacafe14	1	7.36E-204	1.63E-15	1.04E-17	2.20E-119	1.72E-24	1.56E-16	9.41E-85
A07Anacafe147	A07	Anacafe14	7	Anacafe14	1	3.68E-184	1.14E-25	2.57E-17	5.38E-104	3.10E-27	3.77E-20	4.60E-74
A08Anacafe148	A08	Anacafe14	8	Anacafe14	1	4.00E-171	3.06E-34	1.39E-16	5.75E-93	4.36E-29	1.84E-23	5.85E-68
A09Anacafe149	A09	Anacafe14	9	Anacafe14	1	4.59E-175	1.06E-31	4.06E-17	4.26E-97	7.57E-29	5.42E-24	6.92E-71
A10Anacafe1410	A10	Anacafe14	10	Anacafe14	1	7.60E-160	3.99E-43	4.92E-15	8.33E-82	2.66E-30	2.94E-26	1.17E-62
A11Anacafe1411	A11	Anacafe14	11	Anacafe14	1	1.39E-188	1.89E-23	8.02E-18	2.63E-108	5.47E-27	1.60E-20	2.03E-77
A12Anacafe1412	A12	Anacafe14	12	Anacafe14	1	4.00E-171	3.06E-34	1.39E-16	5.75E-93	4.36E-29	1.84E-23	5.85E-68
B01Anacafe1413	B01	Anacafe14	13	Anacafe14	1	3.68E-184	1.14E-25	2.57E-17	5.38E-104	3.10E-27	3.77E-20	4.60E-74
B02Anacafe1414	B02	Anacafe14	14	Anacafe14	1	7.84E-191	3.50E-22	1.37E-17	1.92E-109	1.92E-26	4.98E-19	1.15E-77
B03Anacafe1415	B03	Anacafe14	15	Anacafe14	1	7.36E-204	1.63E-15	1.04E-17	2.20E-119	1.72E-24	1.56E-16	9.41E-85
B04Anacafe1416	B04	Anacafe14	16	Anacafe14	1	3.68E-184	1.14E-25	2.57E-17	5.38E-104	3.10E-27	3.77E-20	4.60E-74
B05Anacafe1417	B05	Anacafe14	17	Anacafe14	1	3.68E-184	1.14E-25	2.57E-17	5.38E-104	3.10E-27	3.77E-20	4.60E-74
B06Anacafe1418	B06	Anacafe14	18	Anacafe14	1	5.31E-164	5.96E-39	9.71E-16	4.97E-87	1.10E-29	7.42E-24	2.67E-64
B07Anacafe1419	B07	Anacafe14	19	Anacafe14	1	4.59E-175	1.06E-31	4.06E-17	4.26E-97	7.57E-29	5.42E-24	6.92E-71
B08Anacafe1420	B08	Anacafe14	20	Anacafe14	1	4.30E-176	5.28E-30	1.04E-16	5.92E-98	5.29E-28	8.09E-21	7.73E-70
B09Anacafe1421	B09	Anacafe14	21	Anacafe14	1	6.40E-184	3.62E-26	1.12E-17	1.48E-104	1.12E-27	9.91E-22	3.83E-75
B10Anacafe1422	B10	Anacafe14	22	Anacafe14	1	4.00E-171	3.06E-34	1.39E-16	5.75E-93	4.36E-29	1.84E-23	5.85E-68
B11Anacafe1423	B11	Anacafe14	23	Anacafe14	1	2.90E-131	1.45E-34	1.31E-10	5.57E-81	9.30E-25	1.95E-20	2.17E-52
B12Anacafe1424	B12	Anacafe14	24	Anacafe14	1	1.39E-188	1.89E-23	8.02E-18	2.63E-108	5.47E-27	1.60E-20	2.03E-77
C01Anacafe1425	C01	Anacafe14	25	Anacafe14	1	1.39E-188	1.89E-23	8.02E-18	2.63E-108	5.47E-27	1.60E-20	2.03E-77

References

Raymond M. & Rousset F, (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, 86:248-249

Chen, K. Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). assign POP: An R package for population assignment using genetic, non-genetic, or integrated

data in a machine-learning framework. Methods in Ecology and Evolution. 9(2)439-446. https://doi.org/10.1111/2041-210X.12897

Zhang, D., Vega, F.E., Solano, W. et al. Selecting a core set of nuclear SNP markers for molecular characterization of Arabica coffee (Coffea arabica L.) genetic resources. Conservation Genet Resour 13, 329–335 (2021)

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.