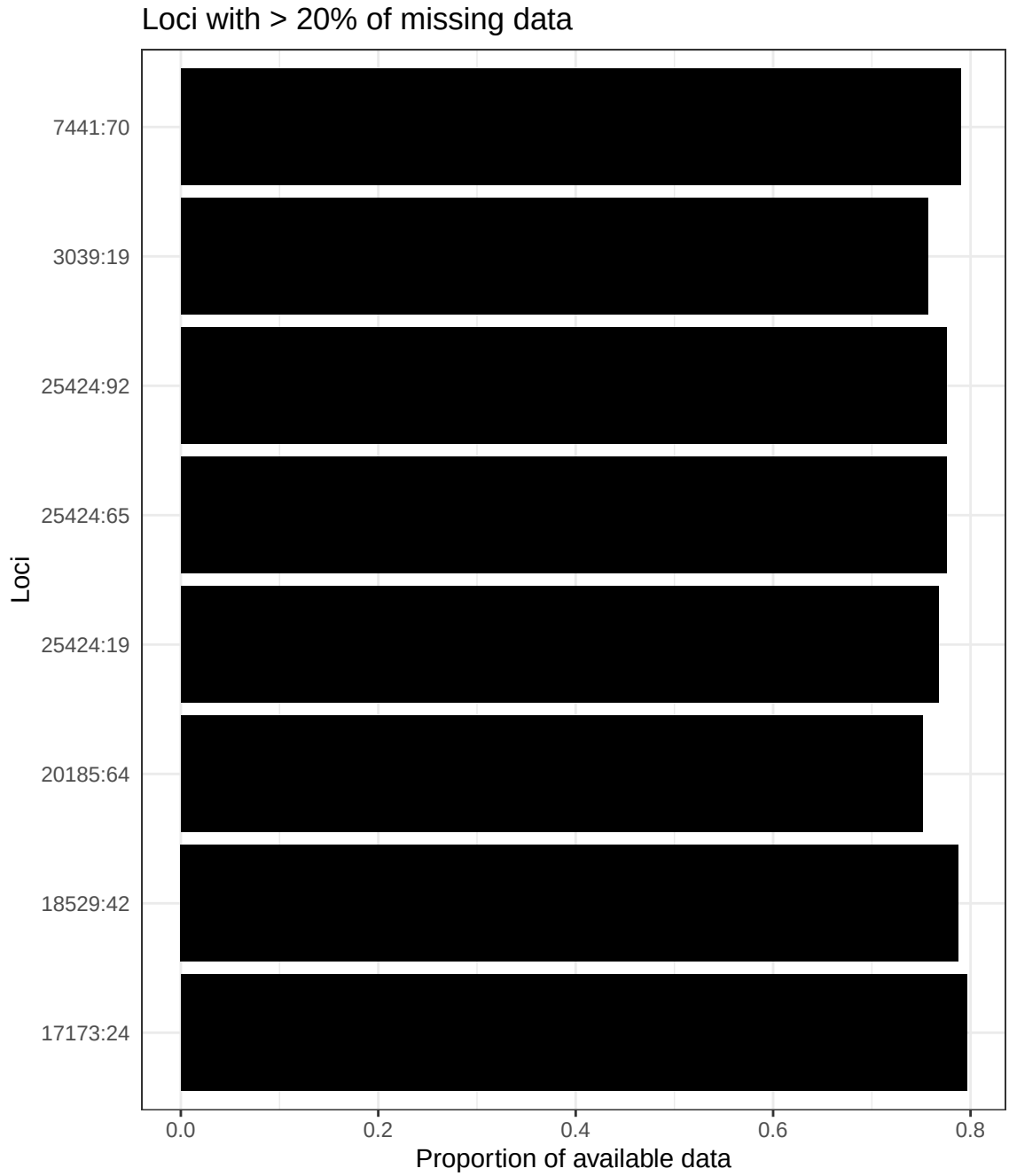
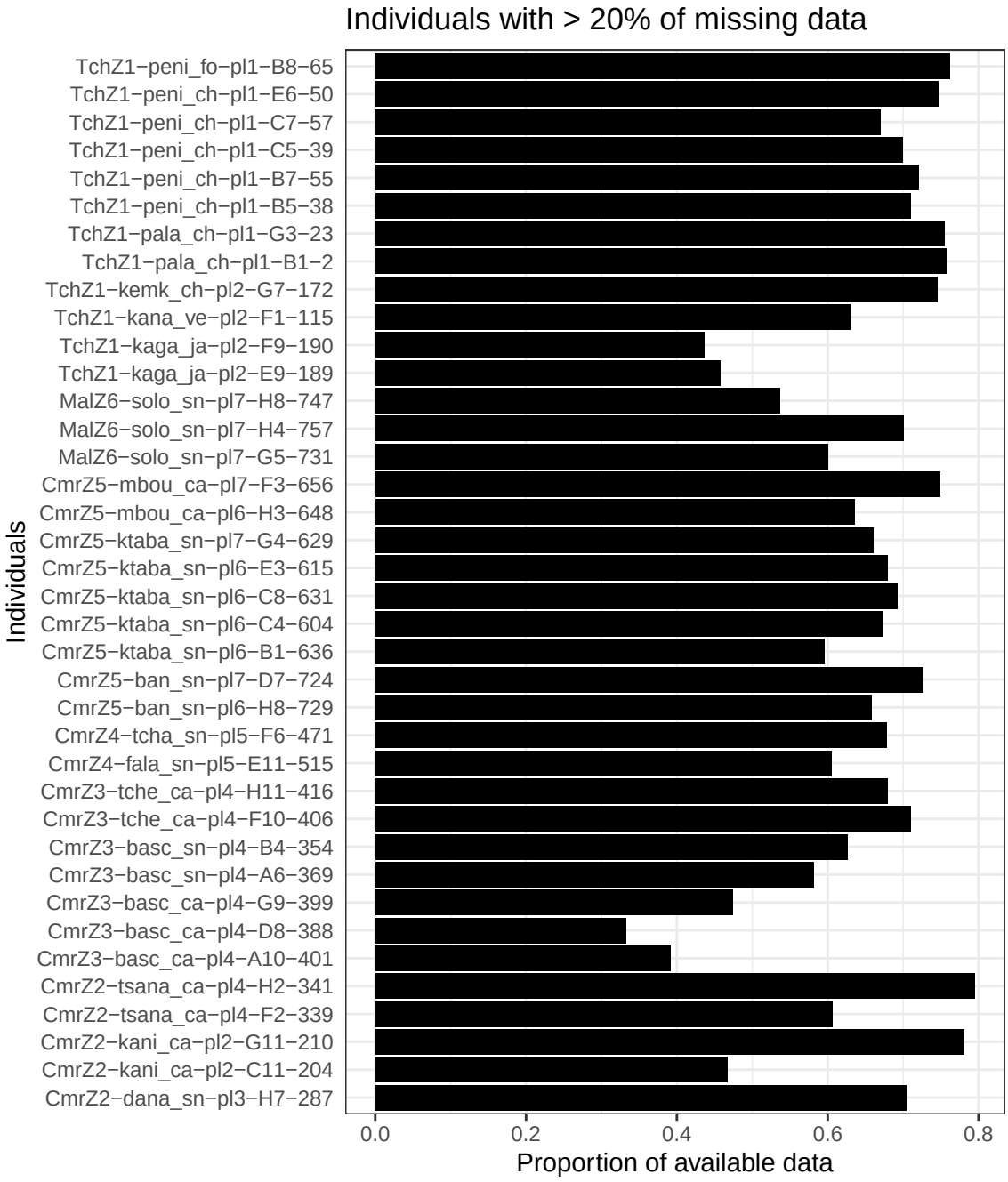
Karite genetic diversity

Karen Cristine Goncalves dos Santos

2021-10-05





Stats

Heterozygosity graphs show the expected and observed heterozygosities (Hexp and Hobs, respectively) for each locus in the populations.

I used Bartlett test of homogeneity of variances to test if the expected and observed heterozygosities were similar.

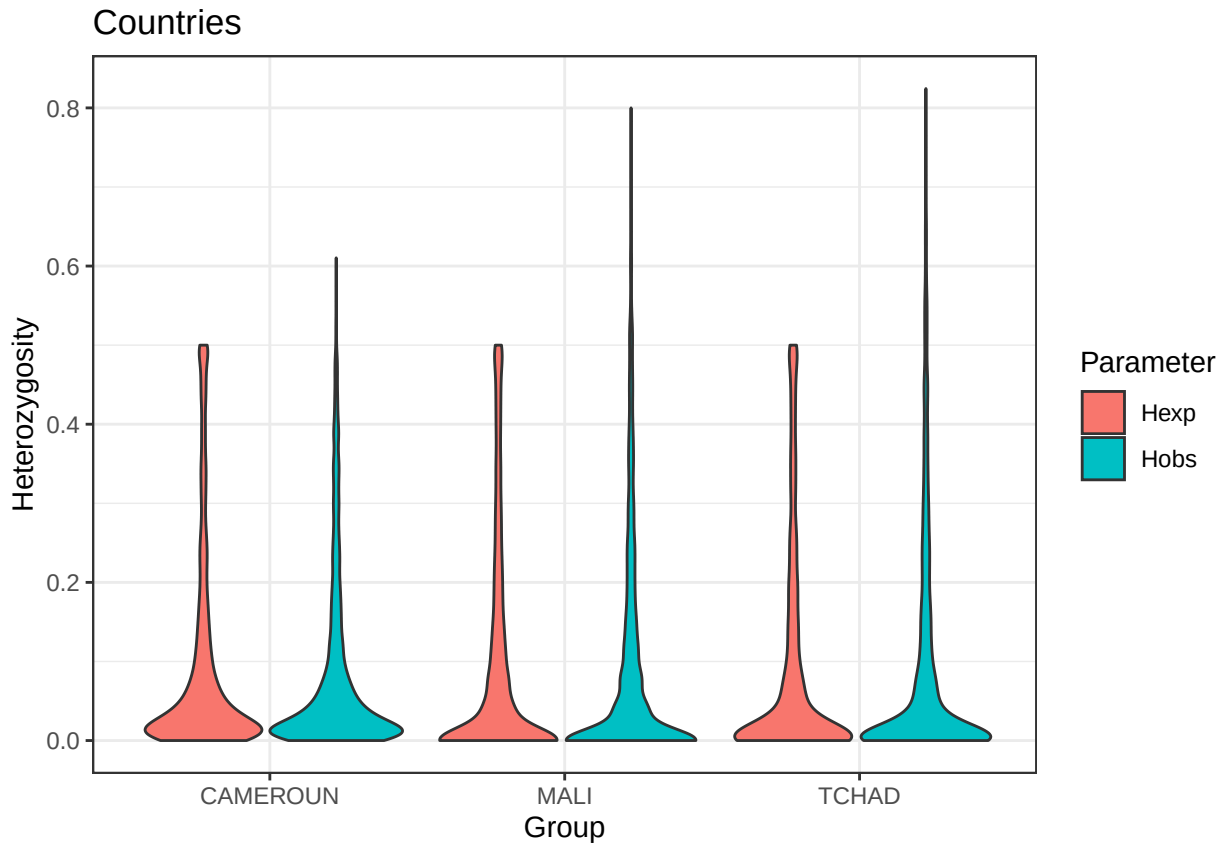
In these tables, where it shows 'NaN' it means that the parameter cannot be calculated.

In the case of Mali, there is only one region and one population location, so the values calculated for the country, the region and the population are the same.

Countries

Group	N individuals	Missing data %
TCHAD	149	10.555805
CAMEROUN	312	8.134396
MALI	29	10.756651

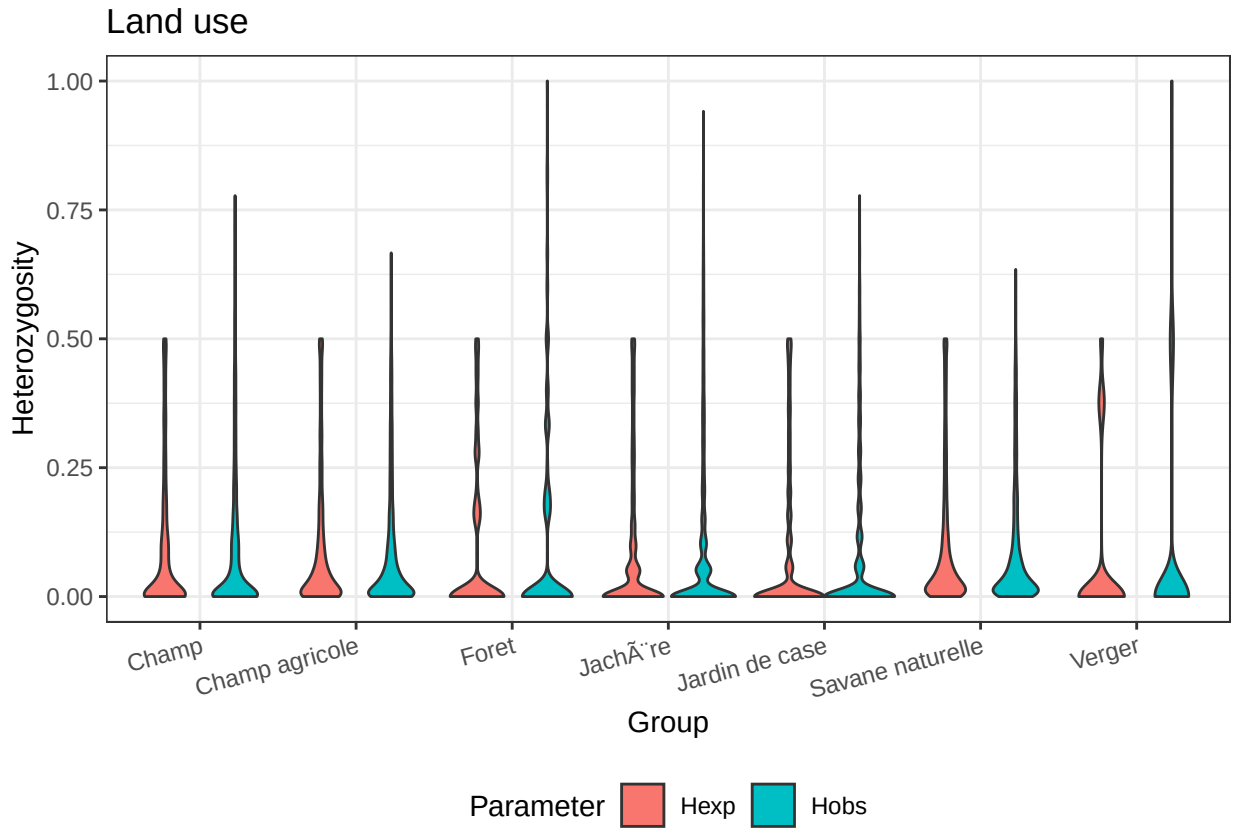
		TCHAD	CAMEROUN	MALI
Ho	Heterozygosity within population	0.0936	0.0836	0.0900
Hs	Genetic diversity within population	0.0804	0.0876	0.0964
Ht	Overall gene diversity	0.0952	0.0986	0.0964
Dst	Gene diversity among samples	0.0148	0.0109	0.0000
Htp	Corrected Ht	0.0967	0.0991	NaN
Dstp	Corrected Dst	0.0163	0.0115	NaN
Fst	Fixation index	0.1559	0.1110	0.0000
Fstp	Corrected Fst	0.1688	0.1160	NaN
Fis	Inbreeding coefficient per overall loci	-0.1640	0.0456	0.0663
Dest	Measure of population differentiation	0.0178	0.0126	NaN



Types of land use

Group	N individuals	Missing data %
Champ	93	10.980273
Foret	6	8.219178
JachÃ¨re	22	12.089348
Verger	32	8.204708
Champ agricole	98	9.916052
Savane naturelle	221	8.039946
Jardin de case	18	4.171758

	Ho	Hs	Ht	Dst	Htp	Dstp	Fst	Fstp	Fis	Dest
Champ	0.0882	0.0886	0.0965	0.0079	0.0981	0.0095	0.0821	0.0969	0.0042	0.0104
Foret	0.1021	0.0815	0.0937	0.0121	0.1059	0.0244	0.1297	0.2301	-0.2531	0.0265
JachÃ¨re	0.0902	0.0722	0.0837	0.0116	0.0953	0.0232	0.1385	0.2436	-0.2498	0.0250
Verger	0.0907	0.0590	0.0724	0.0134	0.0859	0.0269	0.1857	0.3132	-0.5375	0.0286
Champ agricole	0.0858	0.0893	0.0978	0.0085	0.0995	0.0102	0.0869	0.1025	0.0391	0.0112
Savane naturelle	0.0861	0.0901	0.0999	0.0098	0.1007	0.0106	0.0980	0.1053	0.0444	0.0117
Jardin de case	0.0732	0.0778	0.0894	0.0116	0.1010	0.0232	0.1296	0.2296	0.0586	0.0251

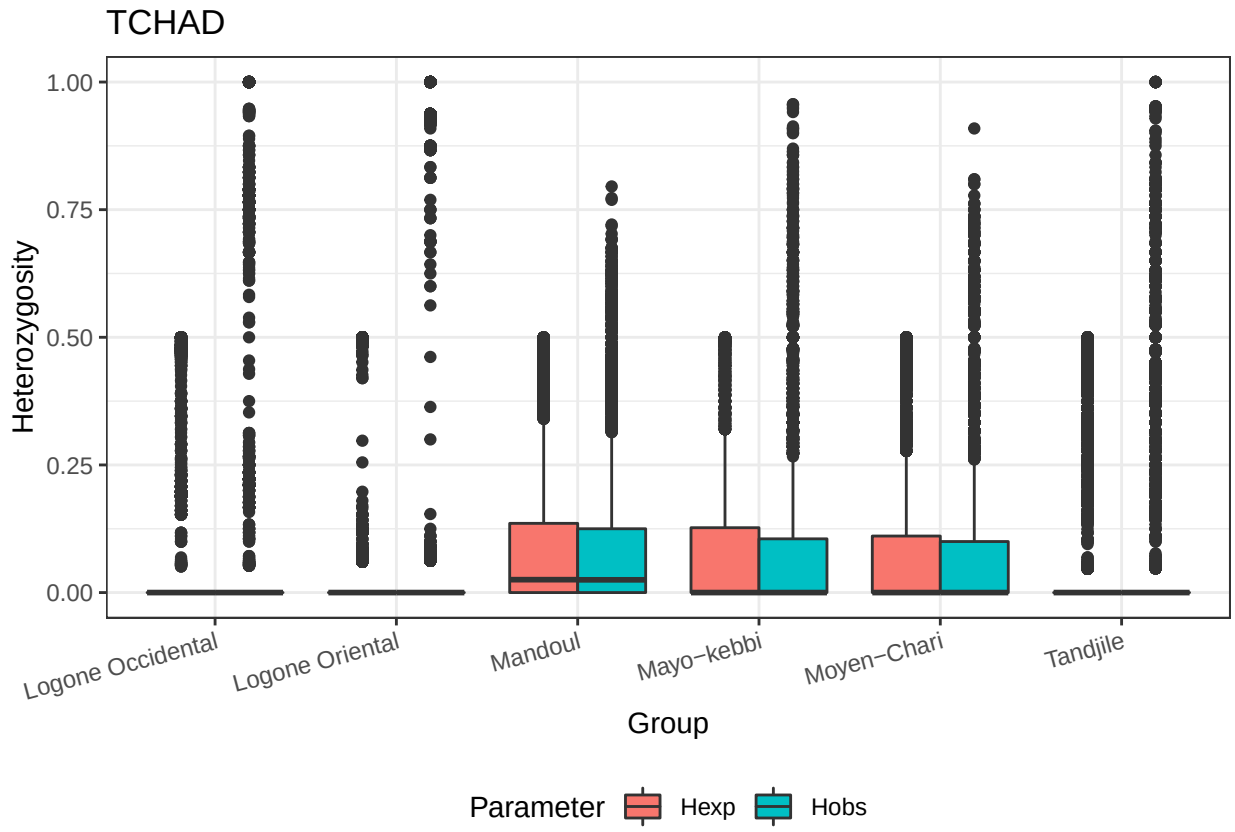


TCHAD

Regions

Group	N individuals	Missing data %
Mayo-kebbi	24	10.815808
Mandoul	46	11.965540
Logone Oriental	16	7.066371
Logone Occidental	19	8.710663
Tandjile	21	8.213666
Moyen-Chari	23	13.555185

	Ho	Hs	Ht	Dst	Htp	Dstp	Fst	Fstp	Fis	Dest
Mayo-kebbi	0.0952	0.0914	0.0914	0.0000	NaN	NaN	0.0000	NaN	-0.0416	NaN
Mandoul	0.0982	0.0973	0.0987	0.0014	0.0994	0.0021	0.0139	0.0208	-0.0093	0.0023
Logone Oriental	0.0888	0.0494	0.0491	-0.0003	0.0486	-0.0007	-0.0070	-0.0142	-0.7982	-0.0007
Logone Occidental	0.0923	0.0657	0.0657	0.0000	NaN	NaN	0.0000	NaN	-0.4049	NaN
Tandjile	0.0950	0.0728	0.0752	0.0023	0.0774	0.0047	0.0311	0.0605	-0.3045	0.0050
Moyen-Chari	0.0898	0.0922	0.0943	0.0021	0.0965	0.0043	0.0227	0.0444	0.0256	0.0047

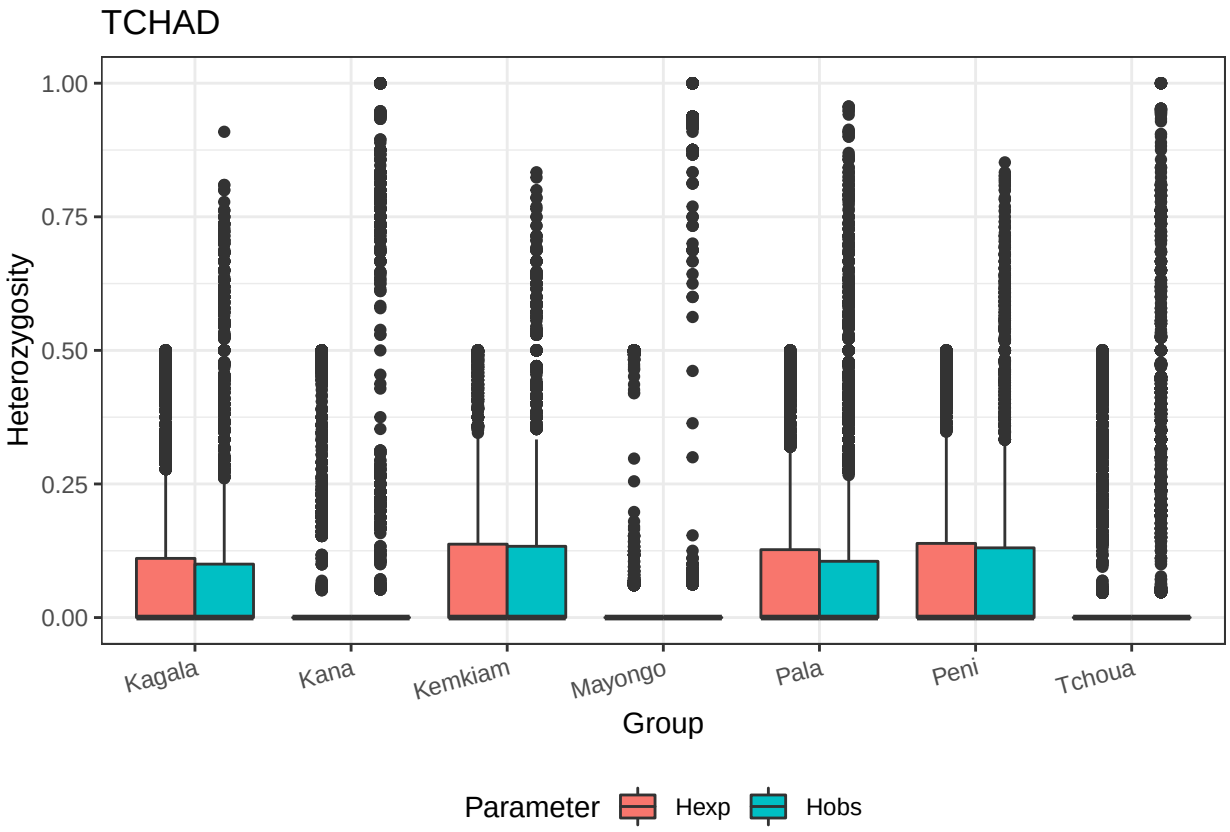


'Localities'

Group	N individuals	Missing data %
Pala	24	10.815808
Peni	29	11.819809

Group	N individuals	Missing data %
Mayongo	16	7.066371
Kana	19	8.710663
Tchoua	21	8.213666
Kemkiam	17	12.214139
Kagala	23	13.555185

	Ho	Hs	Ht	Fis
Pala	0.0952	0.0914	0.0914	-0.0416
Peni	0.1016	0.0975	0.0956	-0.0425
Mayongo	0.0888	0.0494	0.0491	-0.7982
Kana	0.0923	0.0657	0.0657	-0.4049
Tchoua	0.0950	0.0728	0.0752	-0.3045
Kemkiam	0.0915	0.0967	0.0967	0.0540
Kagala	0.0898	0.0922	0.0943	0.0256

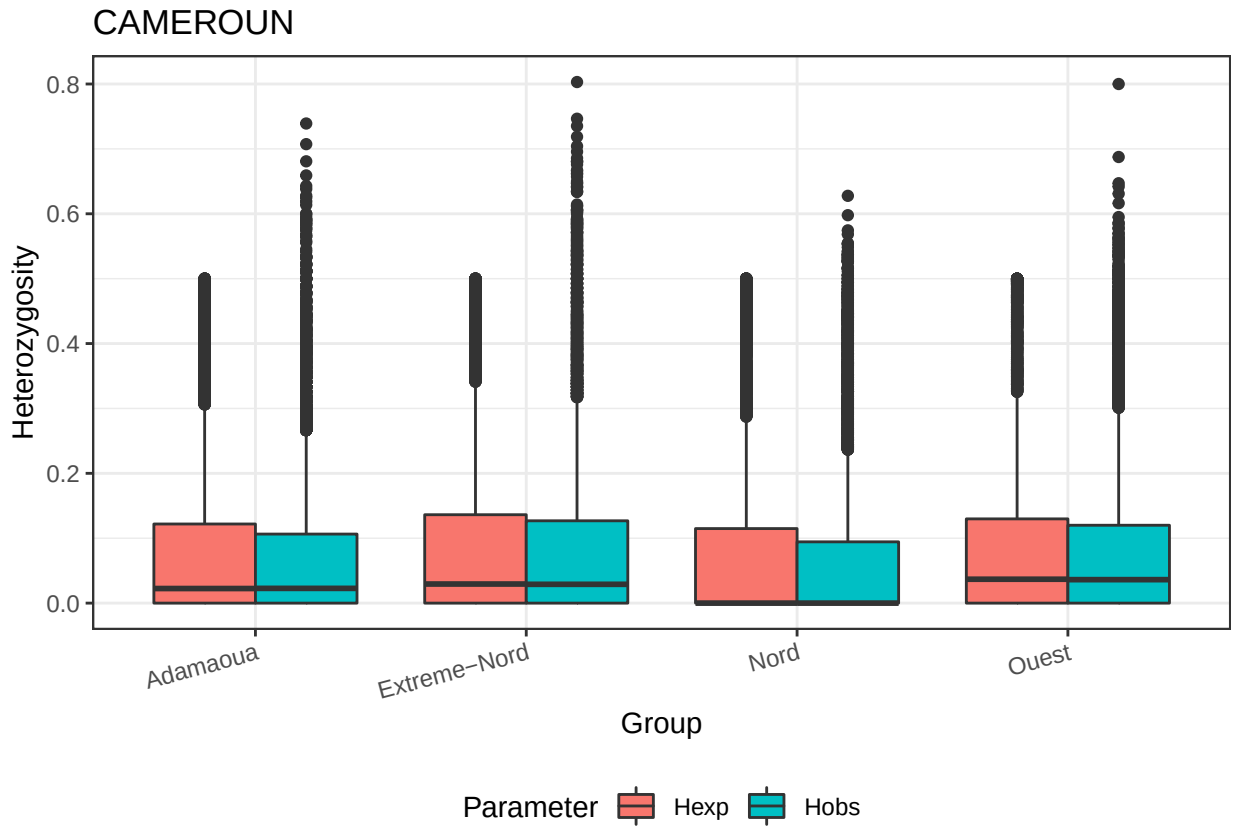


CAMEROUN

Regions

Group	N individuals	Missing data %
Extreme-Nord	74	8.077603
Nord	101	8.204278
Adamaoua	48	7.394768
Ouest	89	8.501215

	Ho	Hs	Ht	Dst	Htp	Dstp	Fst	Fstp	Fis	Dest
Extreme-Nord	0.0929	0.0913	0.0980	0.0067	0.0994	0.0081	0.0685	0.0813	-0.0177	0.0089
Nord	0.0758	0.0811	0.0873	0.0062	0.0886	0.0075	0.0713	0.0844	0.0649	0.0081
Adamaoua	0.0767	0.0835	0.0923	0.0088	0.0945	0.0110	0.0955	0.1165	0.0807	0.0120
Ouest	0.0901	0.0968	0.0982	0.0014	0.0986	0.0019	0.0142	0.0189	0.0686	0.0021

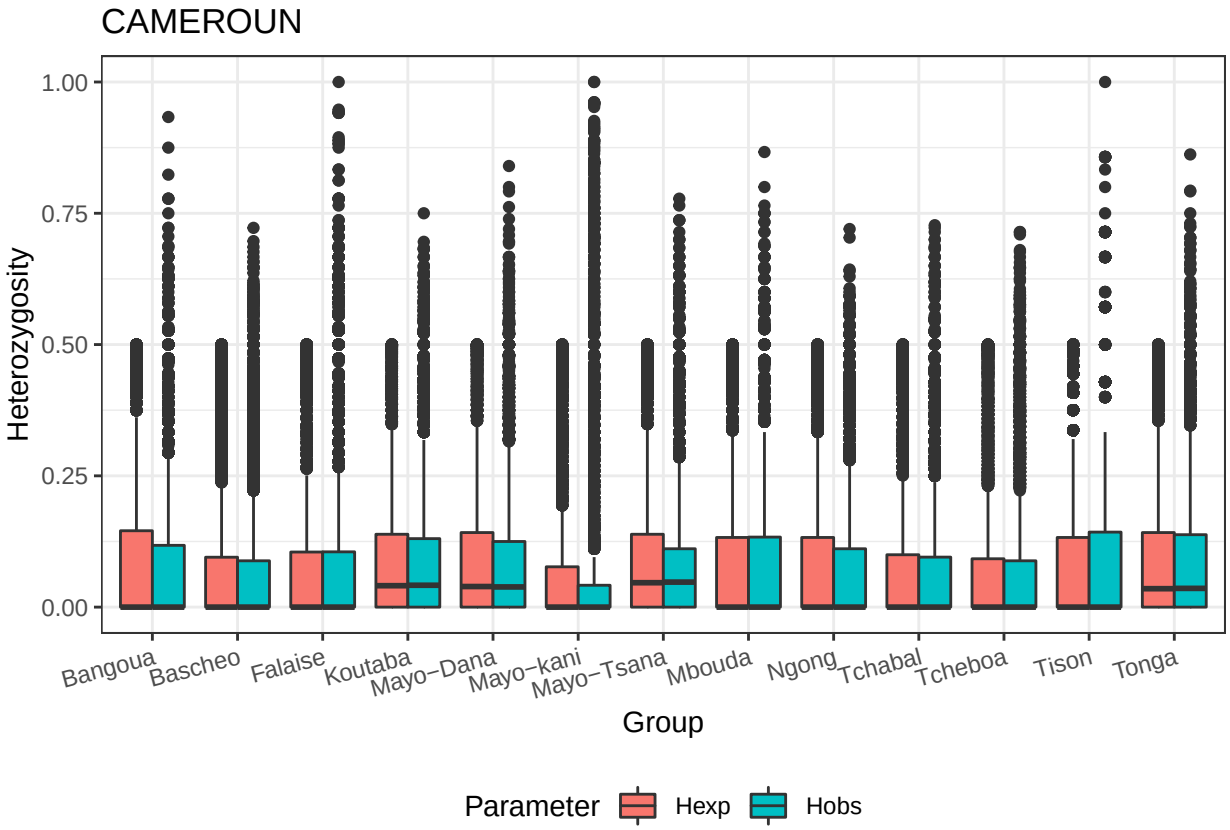


'Localities'

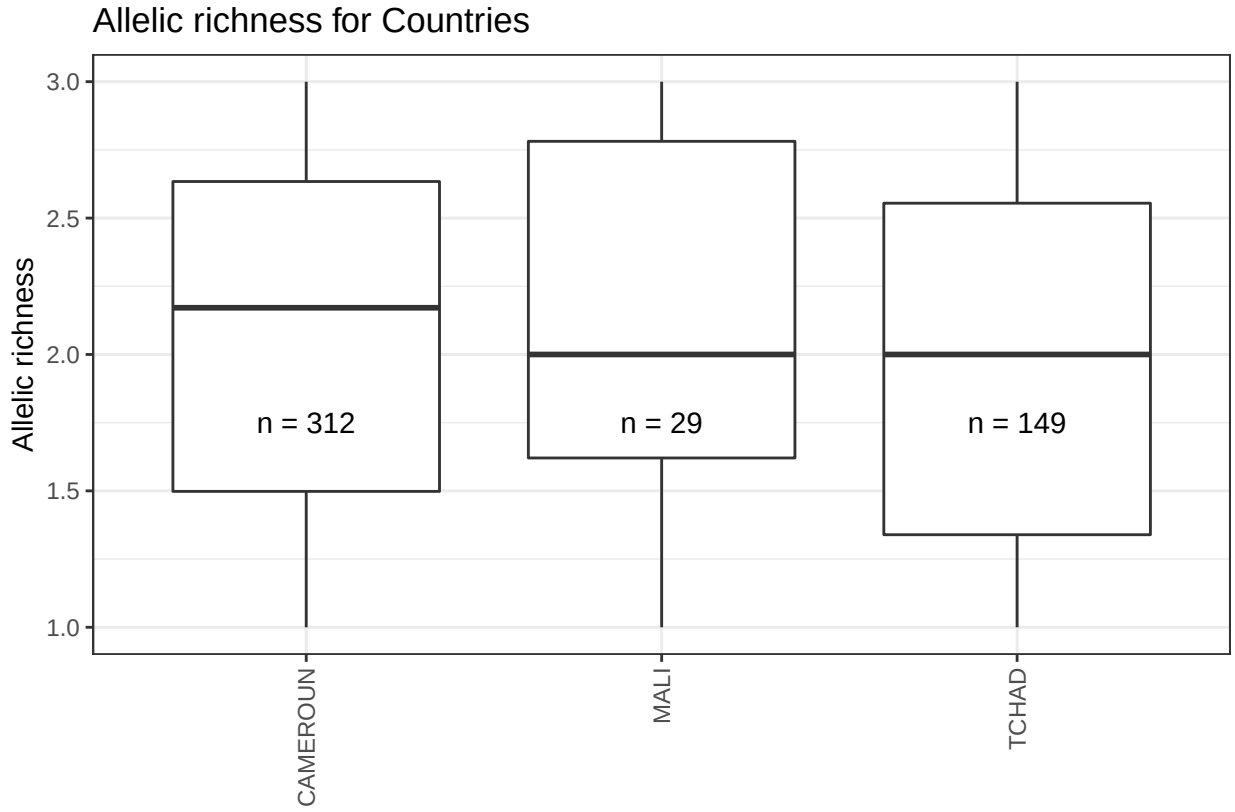
Group	N individuals	Missing data %
Mayo-kani	27	7.760413
Mayo-Dana	26	8.407664
Mayo-Tsana	21	8.076771
Bascheo	38	11.598140
Tcheboa	35	6.651415
Ngong	28	5.539401

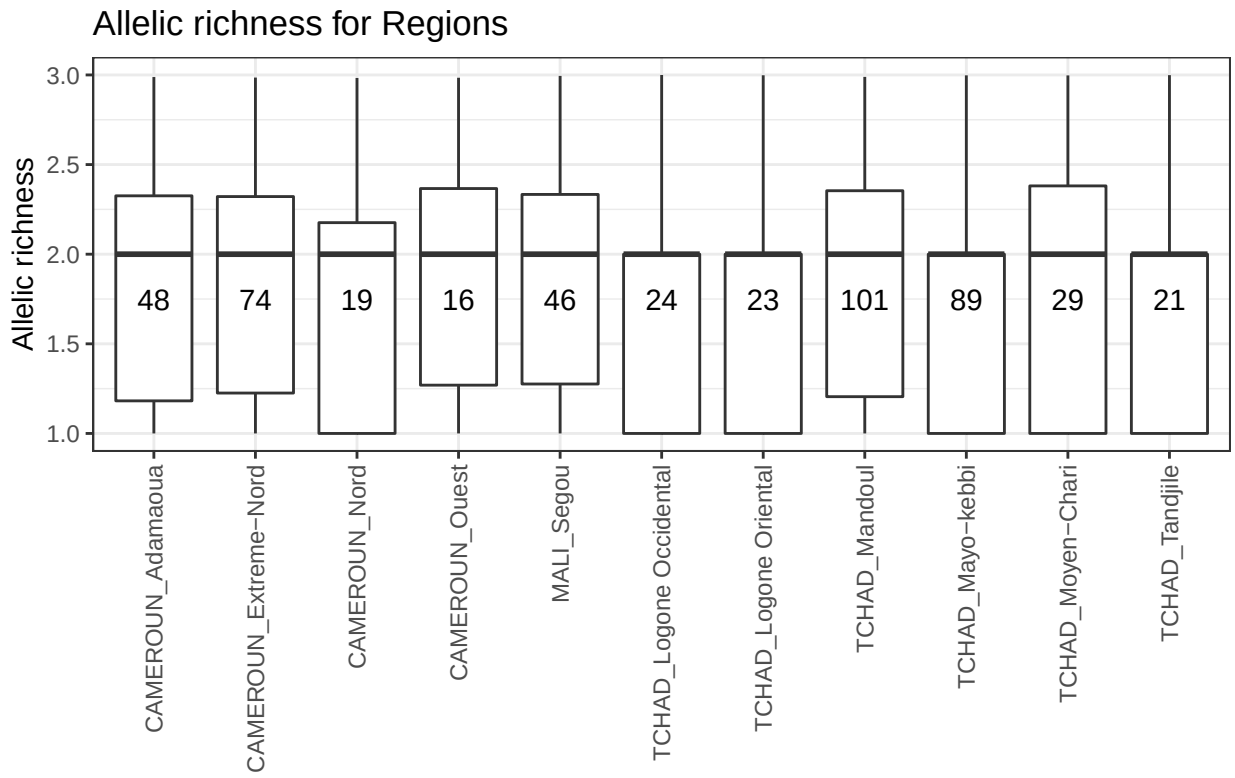
Group	N individuals	Missing data %
Tchabal	22	7.717538
Falaise	19	7.534754
Tison	7	6.000386
Koutaba	24	10.707280
Mbouda	17	9.705939
Tonga	29	6.335034
Bangoua	19	7.942971

	Ho	Hs	Ht	Fis
Mayo-kani	0.0948	0.0805	0.0824	-0.1780
Mayo-Dana	0.0927	0.0994	0.0995	0.0675
Mayo-Tsana	0.0914	0.0976	0.0983	0.0637
Bascheo	0.0759	0.0807	0.0817	0.0592
Tcheboa	0.0770	0.0819	0.0825	0.0601
Ngong	0.0746	0.0807	0.0904	0.0757
Tchabal	0.0685	0.0783	0.0792	0.1254
Falaise	0.0900	0.0852	0.0852	-0.0566
Tison	0.0883	0.0975	0.0975	0.0948
Koutaba	0.0903	0.0967	0.0967	0.0658
Mbouda	0.0894	0.0970	0.0970	0.0783
Tonga	0.0931	0.0995	0.0995	0.0648
Bangoua	0.0877	0.0938	0.0938	0.0652

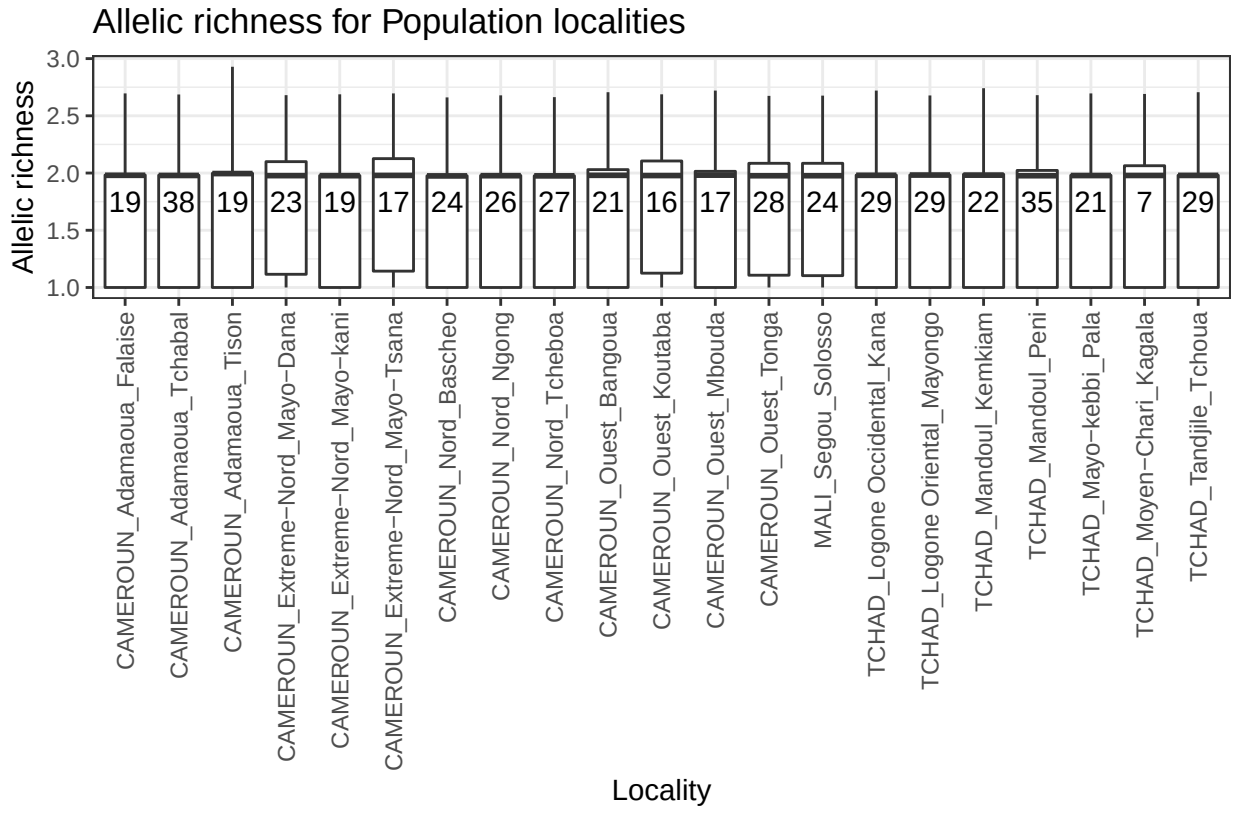


Allelic richness

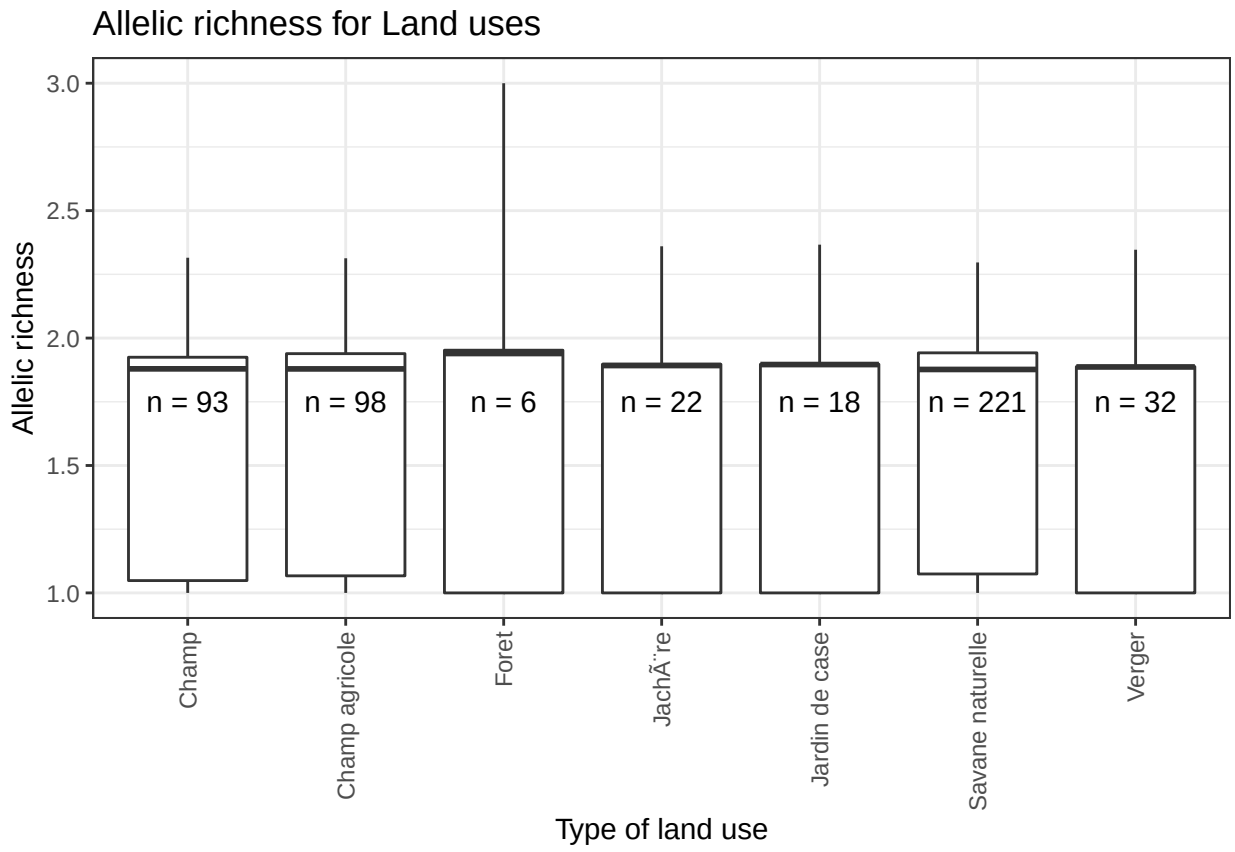




Value inside indicates sample size



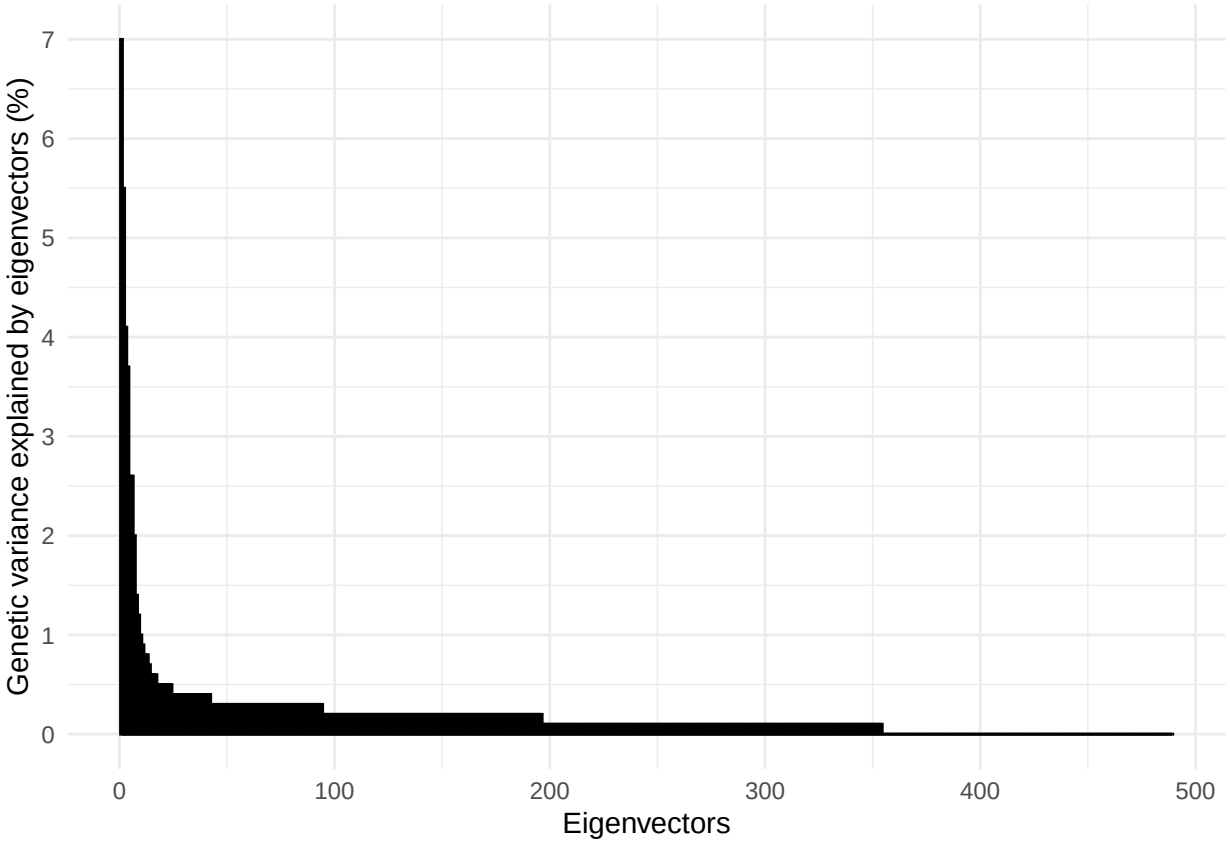
Value inside indicates sample size

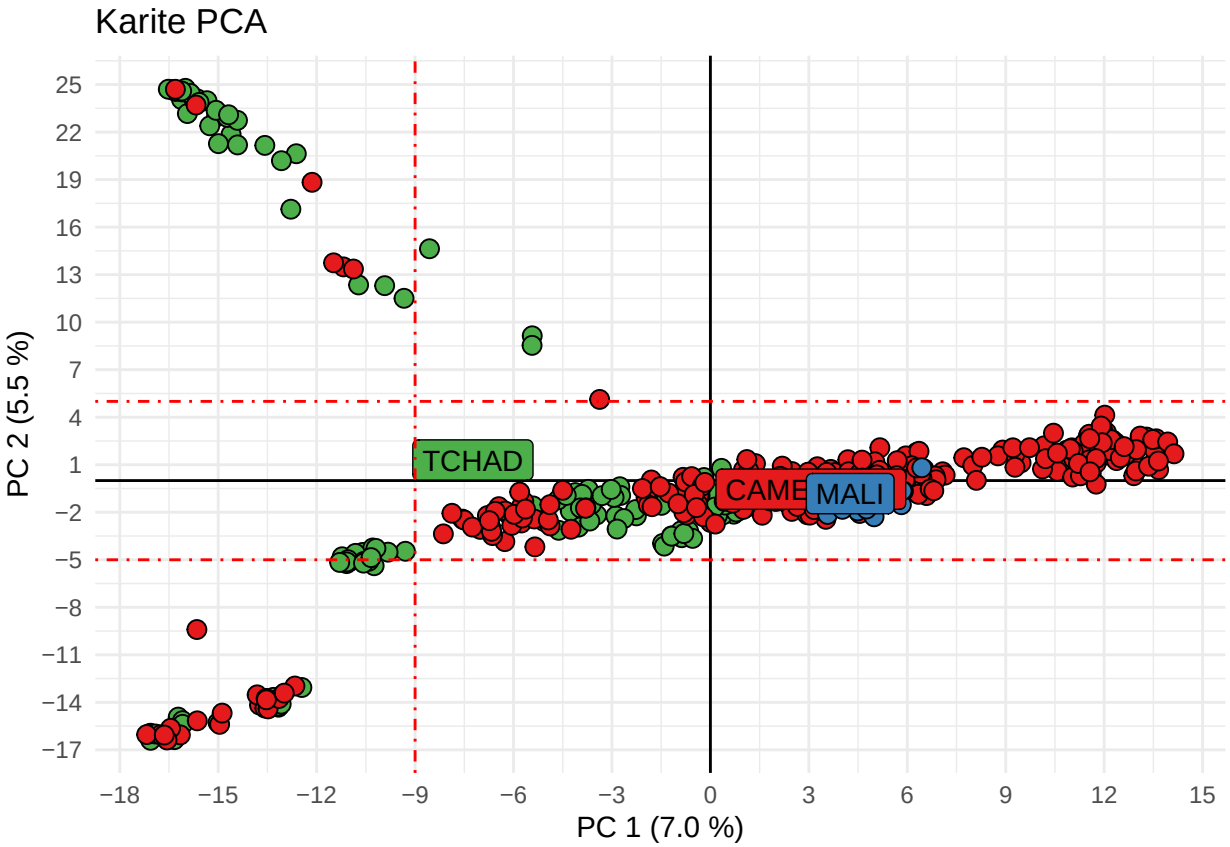


Genetic distance between samples

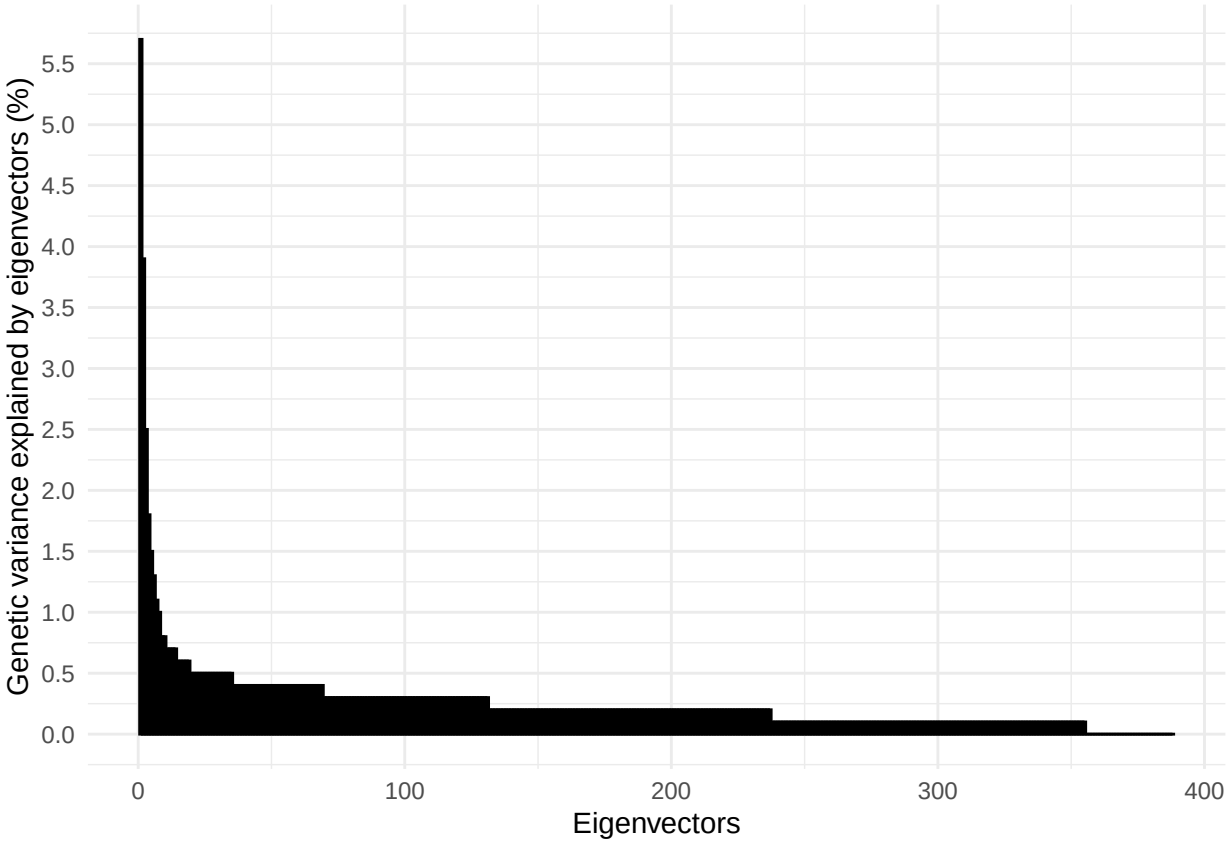
PCA

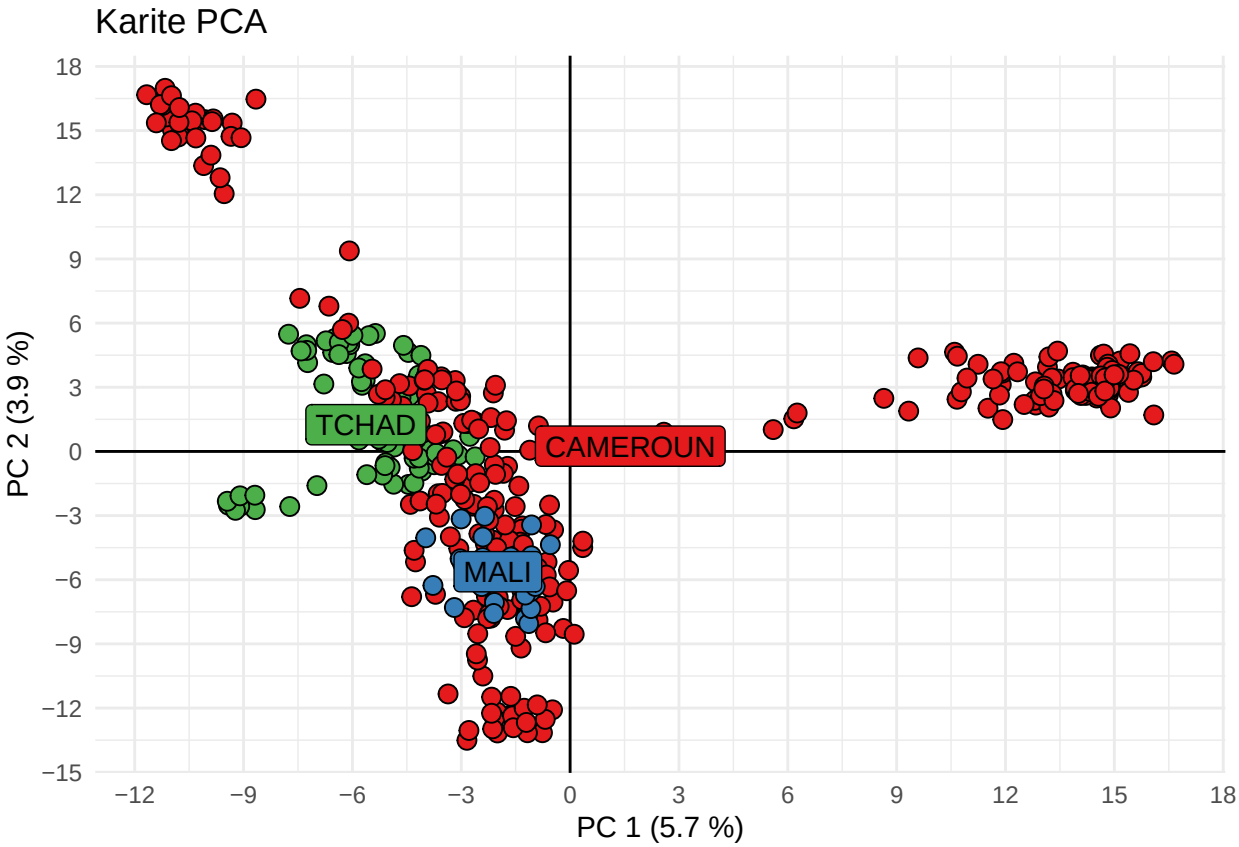
I used Tassel5 to perfom principal component analysis on the SNP data. Each component explains very little of the variance of the data, so here are two graphs with the first 4 principal components.


Genetic variance explained by eigenvectors (%)
7
6
5
4
3
2
1
0
0
100
200
300
400
500
Eigenvectors

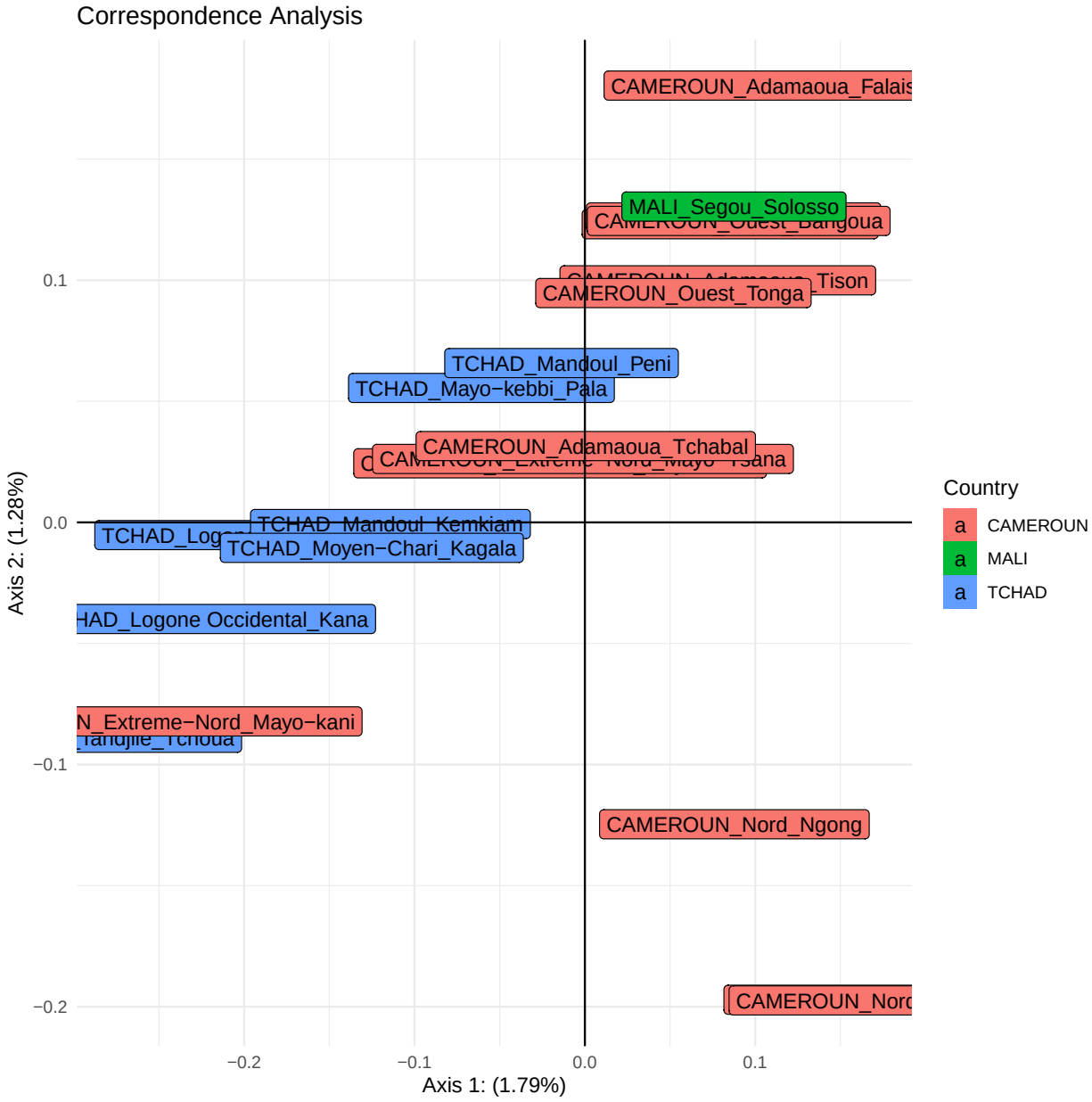


I repeated the PCA with only the individuals in the big group (those located between the red dotdashed line in PC2 & to the right of the red dotdashed line in PC1).


Genetic variance explained by eigenvectors (%)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
100
200
300
400
Eigenvectors


Karite PCA
PC 2 (3.9 %)
PC 1 (5.7 %)
TCHAD
CAMEROUN
MALI
18
15
12
9
6
3
0
−3
−6
−9
−12
−15
−12
−9
−6
−3
0
3
6
9
12
15
18

Correspondance analysis



AMOVA

Table 13: Result AMOVA. Hierarchy: *Country - Region - Location.Population.*

	Df	Sum Sq	Mean Sq
Between Country	2	5665.020	2832.5100
Between Region Within Country	8	17953.530	2244.1912
Between samples Within Region	10	8622.499	862.2499
Within samples	469	108808.122	232.0003

	Df	Sum Sq	Mean Sq
Total	489	141049.171	288.4441

Table 14: Components of covariance.

	Sigma	%
Variations Between Country	5.307309	1.789748
Variations Between Region Within Country	32.118674	10.831163
Variations Between samples Within Region	27.113225	9.143209
Variations Within samples	232.000261	78.235879
Total variations	296.539468	100.000000

Table 15: Phi - population variation statistics.

	Phi
Phi-samples-total	0.2176412
Phi-samples-Region	0.1046384
Phi-Region-Country	0.1102855
Phi-Country-total	0.0178975

Table 16: Monte-Carlo tests (999 repetitions). Sample = population localities

Test	Obs	Std.Obs	Alter	p.Value
Variations within samples	232.000261	-123.5137417	less	0.001
Variations between samples	27.113225	36.0279608	greater	0.001
Variations between Region	32.118674	4.3283223	greater	0.001
Variations between Country	5.307309	0.8212139	greater	0.232

Nei's distance

I used the function "dist.genpop" from the package *adegenet* to calculate the Nei's distance between the samples.

In the heatmap below, darkgray to black indicates low genetic distance between the populations, while light gray to white indicates high genetic distance.

We can see they are very similar but there are some groups:

- The samples from Tchad: Moyen-Chari - Kagala, Mayo-kebbi - Pala, Mandoul - Peni and Mandoul - Kemkiam (top right)
- The sample from Mali with samples from Cameroun Ouest (center)
- The samples from Cameroun Nord (down and left from the previous ones)
- Two samples from Cameroun Extreme-Nord: Mayo-Tsana and Mayo-Dana (down and left from the previous ones)

We also see that the samples from Tchad Logone Oriental are different from all the others and from each other.

Differentiation metrics

I used the package *mmod* to calculate the Jost's D, Hendrick's G'st and Nei's G'st between all populations.

Countries

Table 17: Jost's D

	TCHAD	CAMEROUN	MALI
TCHAD	0.0000000	0.0044862	0.0083408
CAMEROUN	0.0044862	0.0000000	0.0034078
MALI	0.0083408	0.0034078	0.0000000

Table 18: Nei's G'st

	TCHAD	CAMEROUN	MALI
TCHAD	0.0000000	0.0204762	0.0379374
CAMEROUN	0.0204762	0.0000000	0.0155284
MALI	0.0379374	0.0155284	0.0000000

Table 19: Hendrick's G'st

	TCHAD	CAMEROUN	MALI
TCHAD	0.0000000	0.0444369	0.0808326
CAMEROUN	0.0444369	0.0000000	0.0338855
MALI	0.0808326	0.0338855	0.0000000

Type of land use

Table 20: Jost's D

	Champ	Foret	JachÃ¨re	Verger	Champ agricole	Savane naturelle	Jardin de case
Champ	0.00000	-0.00069	0.00302	0.01147	0.00347	0.00361	0.01049
Foret	-0.00069	0.00000	0.00680	0.01676	0.00422	0.00409	0.01164
JachÃ¨re	0.00302	0.00680	0.00000	0.00301	0.00546	0.00651	0.01310
Verger	0.01147	0.01676	0.00301	0.00000	0.01442	0.01593	0.02272
Champ agricole	0.00347	0.00422	0.00546	0.01442	0.00000	0.00104	0.00363
Savane naturelle	0.00361	0.00409	0.00651	0.01593	0.00104	0.00000	0.00513
Jardin de case	0.01049	0.01164	0.01310	0.02272	0.00363	0.00513	0.00000

Table 21: Nei's G'st

	Champ	Foret	JachÃ¨re	Verger	Champ agricole	Savane naturelle	Jardin de case
Champ	0.00000	-0.00314	0.01432	0.05763	0.01584	0.01629	0.04875
Foret	-0.00314	0.00000	0.03124	0.08117	0.01888	0.01812	0.05306
JachÃ¨re	0.01432	0.03124	0.00000	0.01649	0.02561	0.02998	0.06236
Verger	0.05763	0.08117	0.01649	0.00000	0.07139	0.07729	0.11394
Champ agricole	0.01584	0.01888	0.02561	0.07139	0.00000	0.00475	0.01740
Savane naturelle	0.01629	0.01812	0.02998	0.07729	0.00475	0.00000	0.02411
Jardin de case	0.04875	0.05306	0.06236	0.11394	0.01740	0.02411	0.00000

Table 22: Hendrick's G'st

	Champ	Foret	JachÃ¨re	Verger	Champ agricole	Savane naturelle	Jardin de case
Champ	0.00000	-0.00698	0.03118	0.11920	0.03454	0.03557	0.10249
Foret	-0.00698	0.00000	0.06698	0.16439	0.04113	0.03955	0.11124
JachÃ¨re	0.03118	0.06698	0.00000	0.03535	0.05512	0.06435	0.12895
Verger	0.11920	0.16439	0.03535	0.00000	0.14577	0.15713	0.22265
Champ agricole	0.03454	0.04113	0.05512	0.14577	0.00000	0.01048	0.03771
Savane naturelle	0.03557	0.03955	0.06435	0.15713	0.01048	0.00000	0.05196

	Champ	Foret	JachÃ¨re	Verger	Champ agricole	Savane naturelle	Jardin de case
Jardin de case	0.10249	0.11124	0.12895	0.22265	0.03771	0.05196	0.00000

UPGMA

Values indicate node support using 500 samples.
Calculated using Euclidean distance matrix