Research Article | **Open Access**

Haplotype Structure and Phylogeographic Evolution of West African Populations of *Sitophilus zeamais* (Coleoptera, Curculionidae)

NDIAYE Mama Racky[1,2*], THIAW Cheikh[1,3] and Mbacké Sembène[1,2]

[1]Department of Animal Biology, Faculty of Sciences & Techniques, University Cheikh Anta DIOP of Dakar, GENGESPOP team; BO: 5005 Dakar, Senegal

[2]BIOPASS Laboratory, UMR 022, Campus of Institute of Research for Development (IRD)/CBGP, Bel-Air BP 1386 Dakar, Senegal

[3]National Center of Agronomic Research/Senegalese Institute of Agricultural Research (CNRA-ISRA) of Bambey, Senegal

*Corresponding author: Racky NM, Department of Animal Biology, Faculty of Sciences & Techniques, University Cheikh Anta DIOP of Dakar, GENGESPOP team; BO: 5005 Dakar, Senegal, Tel: +221777301010; E-mail: mamaracky.ndiaye@ucad.edu.sn

Abstract

Knowledge of spatial and temporal distribution of genetic variability within and between populations is a crucial step for establishing strategies management of stored crops and commodities to assure food safety. The maize weevil, *Sitophilus zeamais*, is known because of his extraordinary potential for destruction of stored cereals; but little is known about its phylogeography and population structure in Africa in general. This study aimed to determine phylogenic relationships and Phylogeographic structuring of maize weevil populations in West and Central Africa. After alignments and corrections, a total of 112 and 109 samples were used for analysis of Cytochrome B and Cytochrome oxidase I, respectively. The standard indices of genetic diversity, analyses of hierarchical molecular variance (AMOVA), haplotypic network, phylogenetic relationships, demographic tests and Nested Clade Phylogeographic Analysis (NPCA) were computed. Analysis of sequences revealed, with the concatenated matrix, for example, the presence of 30 haplotypes and a high level of haplotype diversity and low nucleotide diversity were observed in West Africa (h = 0.922 ± 0.023; π = 0.0067 ± 0.00116); but these both parameters were twice lower in Central Africa's populations (h=0.512 ± 0.085; π =0.0034 ± 0.00083). The majority haplotype represented 43% of the overall population and connected West Africa to Central Africa. Maize weevil populations showed a genetic structure but indicated a lack of phylogeographic signal. They also showed that most of molecular variance was due to individual characteristics in a population (~60%); while a proportion of molecular divergence (8-14%) was related to a differentiation between two biogeographical regions in Sub-Saharan Africa. The latters were characterized by the presence of endemic haplotypes as well as common haplotypes. Analysis of mismatch distribution, neutrality tests and Bayesian inferences suggested the existence of 4 to 5 main clusters that had separated themselves from a bottleneck followed by a moderate demographic expansion since Pleistocene period.

Keywords: *Sitophilus zeamais*; Sub-Saharan Africa; Phylogeography; Genetic structure; History evolution; Cytochrome B; Cytochrome oxidase I

Introduction

Originated in Central America, more particularly in Mexico and imported to Africa in the sixteenth century by Portuguese explorers, maize (*Zea mays L*) is a staple food for many African countries. However, like all cereals, maize is not spared from damage caused by weevils which can depreciate 25% or even 40% of stored harvest in 6 months [1]. Considered as primary insects, these weevils facilitate ways for secondary insects such as *Tribolium confusum* (Du Val), but also for fungi which, in addition, secrete harmful and carcinogenic substances. Among these latters, we distinguish aflatoxin that compromises food safety and status health of most vulnerable groups in Africa. It is estimated that more of 200 million people in Sub-Saharan Africa (ASS) depend on corn like food source of safety and economic welfare [2]. In addition, among the first 20 countries that generate 96% of the total production of corn in ASS, eight ones were in West Africa [3]. Beyond, previous studies have shown extent of the damages that causes the genus of *Sitophilus* in stocks. Indeed, in rural areas where agricultural commodity conservation techniques are not enough elaborated, *Sitophilus zeamais* (Motschulsky, 1855) can cause post-harvest losses up to 90% in 5 months of storage [1]. In addition, economic importance of culture experienced by grains/seeds like corn would have led to import of grains (seeds) by rural populations. This conveyance of grains from one country to another could be accompanied by introduction of infested products [4,5]. And therefore an installation in areas that were not previously affected by losses inflicted by maize weevil. Consequently, weevils constitute an agro-economic and health problem. Moreover, characterization of populations by molecular analysis is effective only for a minority of sub-Saharan populations. This study aimed to determine phylogenetic relationships and phylogeographic structuring of maize weevil in 9 populations (countries) belonging in two biogeographical regions of West and Central Africa which are Sahelo-Sudanian and Guineo-Congolian zones.

Materials and Methods

Populations sampling

Populations sampling was carried out in seven countries of West Africa (Burkina Faso, Ivory Coast, Ghana, Mali, Niger, Republic of Guinea and Senegal) and two countries of Central Africa (Cameroon and Central African Republic, CAR). These latest populations were added to highlight effects of ecological differences between two biomes

[6]: tropical and subtropical moist broadleaf forests (Guineo-Congolian region) and tropical and subtropical grasslands, savannas and shrub lands (transitional zone of Sahel and centre of endemism of Sudanian region) on the one hand; and on the another hand, to underline anthropogenic effects with samples came out of the integrated economic areas of West Africa. Moreover, many studies have determined biogeographical regions and/or biomes based on climate and bio-ecological parameters in West and Central Africa subregion [6-10]. However, the fundamental differences between these two biogeographical regions are strongly linked to climate, particularly to rainfall which is much more substantial in the South than in the North. Thus, we globally considered 2 groupes of populations, in this study, Guineo-Congolian and Sahelo-sudanian biogeographical regions. Guineo-Congolian region was so defined by Anhuf (2000) (8000 B.P) and corresponds to the putative rainforest refugia proposed by Maley [11] during climatic changes of Pleistocene (18000 B.P.). We so highlight that the term "Sahelo-sudanian zone" includes the transition zone of Sahel and the center of endemism of Sudanian region in our analysis. Then, the Sahel is a semiarid region situated in the south of the Sahara, occurs mostly between 10 and 20" N [12]. The Sudanian region forms a belt from the west coast in Guinea to the Red Sea coast in Sudan and Eritrea. Its pluviometry gives around 1000 mm per annum in the south, where there is woodland, to about 600 mm per annum in the grasslands of the north. Finally, samples came from of each country are considered like a population and were stored in absolute ethanol at 90°C until their use by PCR-sequencing techniques.

DNA extraction and PCR-sequencing

Samples were extracted using a standard QIAGEN® protocol. Fragments of two mitochondrial genes were amplified using specific primers developed by Simon et al. [13]: CB1 (5-GGATCACCTGATATAGCATTCCC-3) and CB2 (5-CCCGGTAAAATTAAAATATAAACTTC-3) for Cytochrome B gene (Cyt-B); Ron II (5-TATGTACTACCATGAGGACAAATATC-3) and Nancy (5-ATTACACCTCCTAATTTATT AGGAAT-3) for Cytochrome oxydase I gene (COI). PCR protocol for Cyt-B consisted of an initial denaturation at 93°C for 1 min 30s, followed by 36 cycles of 35 s at 93°C, 1 min at annealing temperature of 47°C and 2 min extension at 72°C, then final extension at 72°C for 8 min For COI gene, PCR reactions were done using the following conditions: initial denaturation at 94°C for 3 min 30 s, followed by 35 cycles of 1 min at 94°C, 1 min of annealing temperature at 47°C and 1 min extension at 72°C, then the final extension at 72°C for 10 min. For each gene, the PCR reaction mixture for each individual contained 25 ml with 19.3 or 18.3 µl MilliQ and 1 or 2 µl DNA matrix for respectively COI and Cyt-B which we added 2.5 ml enzyme buffer supplied by the manufacturer, 1 µl MgCl$_2$, 0.5 µl dNTP, 0.25 µl of each primer and finally 0.2 µl Taq polymerase. The consensus sequence of each population has been deposited in GenBank under accession numbers: MF581389-MF581406.

Sequence alignment and genetic diversity inferences of the data sets

Sequences were aligned and edited by BioEdit ver. 7.2.5 [14] which uses Clustal W method [15]. Standard indices of diversity such number of total haplotypes (HT), number of endemic/private haplotypes (HP), polymorphic sites (S), parsimony informative sites (SIP), singleton variable sites (SS) and saturation rate of A+T were assessed by MEGA7 [16], ver. 7.0.18. Haplotype and nucleotide diversities (Hd and π) were estimated with software DnaSP [17], ver. 5.10.01. And finally, we have realized T-Student tests of the average values of π, Hd and K between the biogeographical regions (i.e. between Sahel-Sudanian and Guineo-Congolian regions) with the Rcmdr package [18] of the statistical software R [19], ver. 3.2.3. We have used the MG94xREV codon model [20] for estimating the global non-synonymous to synonymous rate ratios in HyPhy [21] for each marker. Afterwards, method FUBAR (Fast, Unconstrained Bayesian AppRoximation) [22], which uses a Bayesian approach to infer nonsynoymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny as implemented in HyPhy, was employed to perform tests to find sites which have experienced pervasive diversification. This method assumes that selection pressure for each site is constant along the entire phylogeny.

Genetic and phylogeographic structuring

The presence of an isolation by distance pattern [23] was tested by performing Mantel tests [24] which were conducted in XLSTAT [25] using the genetic differentiation matrix from MEGA7 and a matrix of geographic distances between breeding sites. Genetic differentiation indexes (FST, NST and GST), diversity indexes of Nei [26], (HT, HS and DST), pairwise genetic divergences between populations (FST) [27] with using haplotype frequencies, hierarchical analysis of molecular variance (AMOVA, [28]) and genetic distances were calculated. Two AMOVA tests were conducted. The first assessed the distribution of genetic variance among populations and the second test used rainfall parameters as structuring factors. A hierarchical ascending classification (HAC) has been performed also with differentiation matrix of concatenated data in statistic software R with the function hclust of stats package and pvclust package [29]. Others analysis were processed using the following softwares: Arlequin [30], ver. 3.5.1.3, PERMUT ver. 2.0 [31] and Mega 7.

Detection of possible barriers to gene flow was inferred by using the method of spatial analysis of molecular variance (SAMOVA) implemented in SAMOVA ver. 2.0 [32]. The test is based on a simulated annealing procedure to explore potential number of groups (K). K varied between 2 and 5 and 100 simulated annealing processes were used for each value of K. Hierarchical clustering analyses were carried out by the Bayesian method in BAPS ver. 6.0 [33]. And finally, a subsequent admixture was executed using the number of clusters chosen in the mixture analysis.

Network structure and population history inferences from NCPA

The haplotype network was delineated by using Network, ver. 5.0.0.0 according to the median-joining network method [34]. The resulting network was used to construct the nested clade design with the procedure described in Templeton et al. [35]. We performed the NCPA (nested-clade phylogeographic analysis) using GeoDis [36], ver. 2.6, based on 1000 bootstrap resamples to test the significance of departures of the clade distance (Dc) and the nested clade distance (Dn). So as to infer causal processes (e.g. restricted gene flow, population fragmentation, range expansion, etc.), we used the Templeton's inference key [37] to determine the likely mechanism of the observed genetic structure. This inference key was constructed based on the expected patterns of geographical association that can occur under three patterns of historical events: restricted gene flow, range expansion and allopatric fragmentation [38]. Thus, the NCPA allows to distinguish historical processes (fragmentation, range

expansion, etc.) from current processes (gene flow, genetic drift, etc…) that are responsible of the observed pattern of genetic variation [39].

Phylogenetic reconstructions

Phylogenetic relationships were established by neighbor-joining (NJ; [40]), maximum likelihood (ML; [41]) and Bayesian inference (BI; [42]) methods. The best-fit evolutionary models of our different matrix were appreciated with Jmodeltest ver. 2.1.10 [43] by using Akaike information criterion corrected (AICc) [44] and likelihood ratio test (LRT), including the out-group. The determination of a phylogenetic signal was tested by the method of Steel et al. [45] with modifications as implemented in DAMBE program ver 5.6.8 [46] and according to the criterions of Lyons-Weiler et al. [46]. Steel's test was performed by sampling all possible quartets in the dataset and evaluated using the mean -value for each sequence [47]. Thus, sequences with a mean -value smaller than 0.04 may be interpreted as lacking phylogenetic signal. Moreover, test of substitution saturation [48,49] was done through the entropy-based information method [48,49] as implemented in DAMBE program.

Trees were rooted using an out-group as *Sitophilus oryzae* (Linnaeus, 1763). The NJ method based on the substitution model of Kimura 2-parameter [50] with gamma and 1000 bootstrap replications was performed by using Neighbor program of Phylip package [51]. Then ML method was executed by PhyML [52], ver. 3.0. Then, the BI was done using MrBayes program [53], ver. 3.2.5, based on 3,000,000 generations of Markov Chain Monte Carlo (MCMC). In this latter, the 25% first generations are ruled out as burn-in and the posterior probabilities were estimated from the remaining trees. Likelihood-maps [54] implemented in Tree-Puzzle [55], ver. 5.2, was used to evaluate conflicting phylogenetic hypothesis (topology of trees) among methods after phylogenetic constructions. The likelihood mapping analysis allowed checking the content of phylogenetic information in an alignment and estimating of quartet support of relationships among groups of sequences. Moreover, internal branches supported by ≥ 70% of replicates are considered as statistically significant.
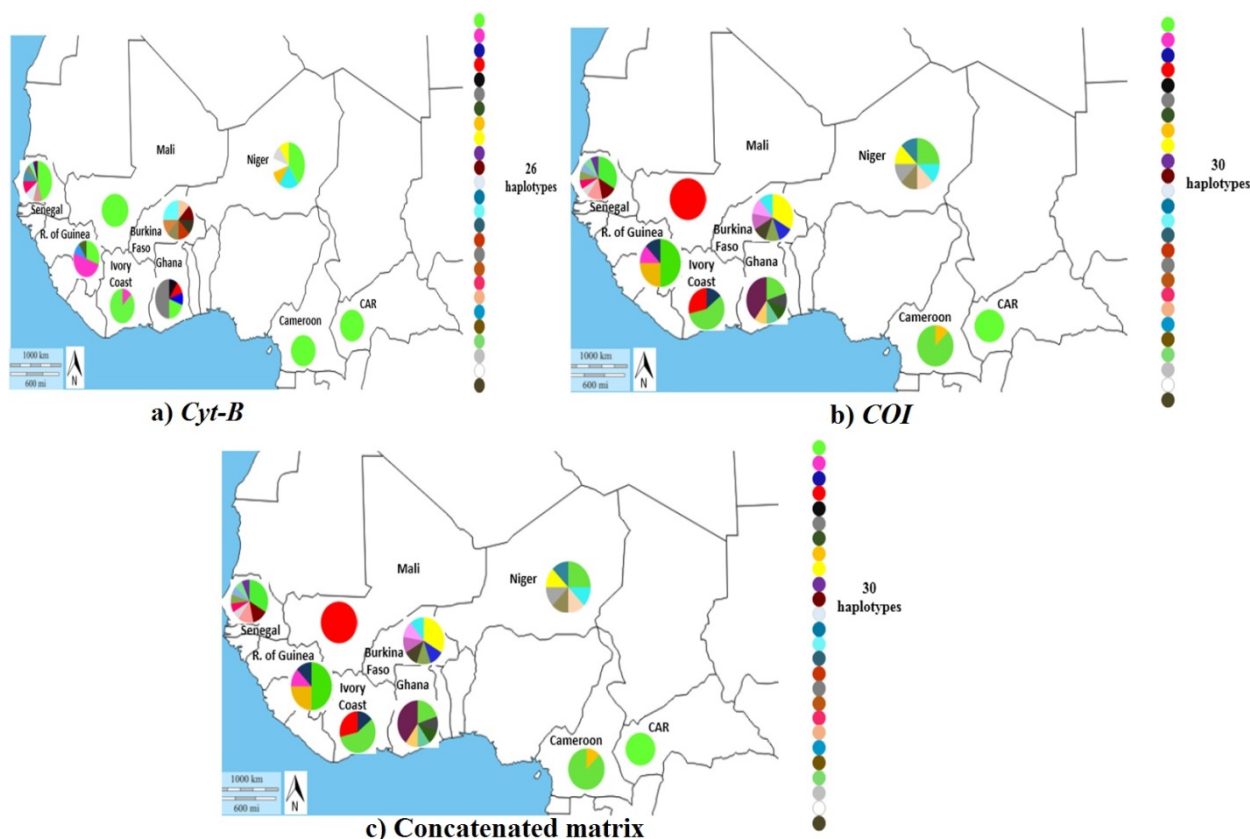

Figure 1: Map of spatial distribution of haplotypes of *Sitophilus zeamais*.

Demographic history, coalescence analysis and time of divergence

Demographic history was explored by inferring mismatch distribution analysis, R2 [56], haplotype and nucleotide diversities [57,58] and S/K ratio or expansion coefficient [59] with the DnaSP program. Besides, the goodness of fit of the observed distribution with the expected values was estimated by the demographic indexes (SSD, Sum of square deviations and Rg, Harpending's raggedness index; Harpending, [60]) with Arlequin program. For neutrality tests, Tajima's D [61] and Fu's Fs (Fu, [62]), which are sensitive to demographic changes, were ascertained by handling DnaSP. These significance tests of neutrality were estimated based on 1000 coalescent simulations.

Coalescence among clades and Bayesian molecular dating of concatenated matrix were computed using BEAST package [63], ver. 1.8.2. The length of the Markov Chain Monte Carlo (MCMC) was set to 10,000,000 generations and log parameters were sampled every 1000th generation. It ran with a GTR model of evolution and a substitution rate of 1.5% and 2.3% per lineage per million years respectively for Cyt-B and COI [64,65]. It so ran with an uncorrelated relaxed molecular clock approach. TreeAnnotator ver. 1.4.8 was used to obtain a consensus tree with posterior probabilities from the tree output file with the burn-in parameter set to 10%, corresponding to the initial 1,000,000 generations of the MCMC chain. FigTree [66], ver. 1.4.2, was handled to visualize the consensus tree from TreeAnnotator output. Thus, this allowed know the different tMRCAS estimated, with 95% highest posterior density (HPD) interval.

Results

Molecular diversity of mitochondrial DNA

For the entire population, 112, 109 and 99 sequences were screened respectively for Cyt-B (442 pb), COI (415 pb) and concatenated matrix (857 pb) after alignment, corrections and then concatenation of 120 and 110 sequenced individuals of two genes cited previously and respectively. There were 58, 9 and 56 polymorphic sites respectively for Cyt-B, COI and concatenated matrix (Table 1). Twenty six (26), five (5) and Thirty (30) haplotypes were identified with Cyt-B, COI and concatenated matrix respectively (Figure 1).

Geographical groups	Climate	N	Nμ	S	SIP	SS	K	HT	HP	%HP	Hd (± SD)	Π (± SD)
GSA		60	58	51	28	23	4.9898	22	20	90.91	0.8203	0.01129
											-0.045	-0.00205
	semi-arid to semi-humid	55	4	4	1	3	0.5899	4	2	50	0.5165	0.00142
											-0.041	-0.00021
		50	48	43	29	14	5.763	24	21	87.5	0.922	0.00673
											-0.023	-0.00116
		52	12	12	1	11	2.3718	6	4	66.66	0.3424	0.00537
											-0.082	-0.00151
GGC	Humid to Subhumid	54	6	6	1	5	0.3564	3	1	33.33	0.2047	0.00086
											-0.069	-0.00045
		49	18	18	12	6	2.889	9	6	66.66	0.512	0.00337
											-0.085	-0.00083
Global population		112	68	58	36	22	3.9574	26			0.6448	0.00895
											-0.052	-0.00145
		109	9	9	1	8	0.5134	5			0.3986	0.00124
											-0.046	-0.00026
		99	63	56	37	19	4.602	30			0.789	0.00538
											-0.039	-0.00082
Chi2 test (GSA *vs.* GGC)							NS				NS	*/**

Table 1: Standard indices of genetic diversity with Cyt-B, COI and concatenated matrix. Cyt-B (bold); COI (italics); concatenated matrix (bold and italics); N: number of individuals; N_μ: number of mutations; S: number of segregating sites; S_{IP}: Parsimony informative sites; S_S: singleton sites ; K: mean pair-wise nucleotide; H_T: number of total haplotypes; H_P: number of private haplotypes; %H_P: percentage of private haplotypes ; Hd: gene diversity; π: nucleotide diversity; GSA vs GGC: Group of Sahelo-soudanian populations versus Group of Guineo-congolian populations; SD: Standard Deviation; P-values significance tests: [*] :P<0.05 ; [**]: 0.05<P<0.01; [***] :0.01<P<0.001; NS (P>0.005).
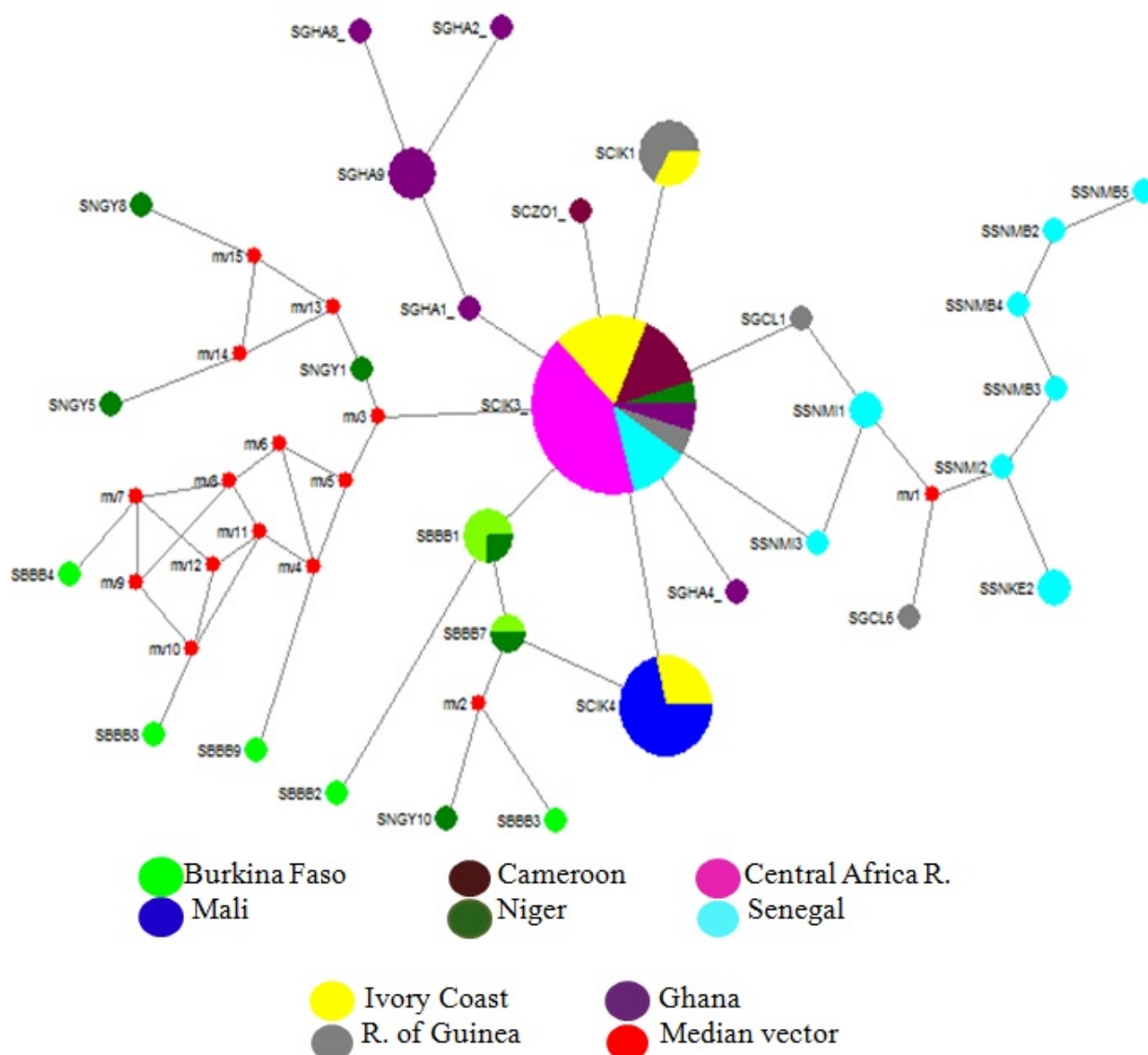
Figure 2: Haplotypes Network of concatenated matrix.

The majority haplotype represented 43% of the sampled individuals and was shared by all populations except in Burkinabe's population and connected essentially West Africa to Central Africa (Figure 2). The global dN/dS ratio over the entire population of *Sitophilus zeamais* was estimated at 0.39, 0.17 and 0.93 respectively for Cyt-B, COI and concatenated matrix. In the latter, there were 11 sites with significant evidence of pervasive diversifying selection, of which 46% are expected to be false positives according to Bayesian analyses of FUBAR method. On average, the richness of A+T is important in both genes (70% for Cyt-B versus 66% for COI). Whereas this richness is higher great at the third codon position. This richness is higher important at the third codon position. Indeed, fragments of genes Cyt-B and COI had a saturation rate of A+T=79.5% and A+T=84.1 respectively. In addition,

mutation rate is in favor of the transitions which are 52, 51 and 50.03% against 48, 49 and 50.07% for the transitions respectively for Cyt-B, COI and concatenated matrix. The evolutionary model estimation indicated that GTR model best fitted both Cyt-B and concatenated data but it underlined the F81 model [40] for COI gene. The Chi 2 test revealed a significant difference of the nucleotide diversity (π) between the two biogeographical areas (Cyt-B and concatenated matrix for P<0.05; and COI matrix with P<0.01). For the haplotype diversity (Hd) and the mean number of nucleotide differences between pairs of haplotypes (K), no differences have been detected (Table 1).

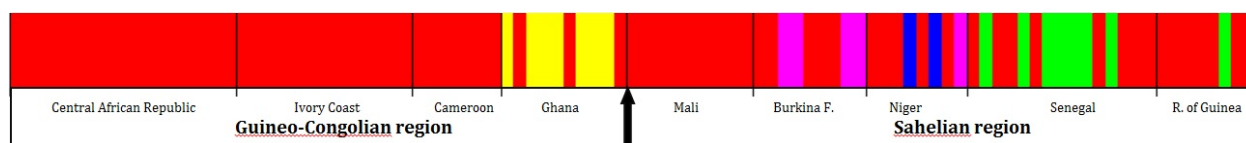**Genetic and phylogeographic structuring**

Figure 3: Clustering of populations by mixture model with concatenated matrix, clusters presented by color.

Source of Variation	degrees of freedom			sum of squares			Variance components			Percentage variation			Fixation Index		
Repartition of molecular variance between biogeographic zones															
between Zones	1	1	1	11.953	2.353	14.958	0.181	0.0388	0.25836	8.73	14.07	10.64	FST : 0.0873	FST : 0.14071	FST : 0.1086
							Va (FST)	Va (FST)	Va (FST)						
Within zones	110	107	97	207.681	25.372	210.547	1.888	0.2371	2.17059	91.27	85.93	89.36	Significance tests		
							Vb	Vb	Vb						
Total	111	108	98	219.634	27.725	225.505	2.069	0.276	2.42895	100	100	100	P-value : 0.0000	P-value : 0	P-value : 0
Repartition of molecular variance between populations															
between populations	8	8	8	81.851	9.395	93.451	0.72641	0.08343	0.941	35.19	31.28	39.07	FST : 0.3519	FST : 0.3128	FST : 0.3907
							Va (FST)	Va (FST)	Va(FST)						
Within populations	103	100	90	137.783	18.33	132.054	1.3377	0.1833	1.467	64.81	68.72	60.93	Significance tests		
							Vb	Vb	Vb						
Total	111	108	98	219.634	27.725	225.505	2.06411	0.26673	2.408	100	100	100	P-value : 0.0000	P-value : 0.0000	P-value : 0.000

Table 2: Repartition of variance molecular by Analysis of MOlecular VAriance (AMOVA). Results: Cyt-B (bold), COI (italics) & concatenated matrix (bold and italics); Significance tests (1000 permutations)

There was no significant effect of isolation by distance (IBD) according to the results of Mantel test; the correlation between genetic and geographic distances is not significant, as well for the Cyt-B (r = -0.178; P = 0.304), the COI (r = -0.010; P = 0.975) and the concatenated matrix (r = 0.109; P = 0.514). The genetic divergence analysis and matrices of differentiation based on a clustering analysis of concatenated data has revealed a differentiation between Sahelo-sudanian populations and those of Guinea-Congolese area by the AU p-values (approximately unbiased) in red ; and the BP p-values (Bootstrap probability) in green (Figure 3).

However, the observed differentiation was attenuated by processes of introgression between biogeographical areas. The AMOVA analysis indicated that 31 to 39% of the genetic variation was partitioned among populations; and 60 to 68% within populations according to molecular markers (Table 2). Considering two groups of populations defined according to isohyetes (rainfall), namely the populations group

of the Sahel-sudanian zone (GSA) and the populations group of the Guineo-Congolian zone (GGC), our AMOVA analysis confirms that the most part of molecular variance is mainly due to genetic variability within a population with more than 85 to 91% of the total variation. However, the differentiation between these 2 groups is significant with 8 to 14% of the total variation. Thus, the analyses of molecular variance showed that variation among individuals within populations explained the essential of the observed variance. These results revealed that *S. zeamais* presented levels of genetic variation between populations no negligible and indicated a population structure which has tendency to attenuate. The results of SAMOVA showed that there are no genetic barriers (Table 3), while BAPS's ones showed that there is a differentiation between the Sahelo-sudanian zone and the Guineo-Congolian one by identifying 5 (mixture model) or 4 (admixture model) clusters (Figure 4).

Source of Variation	degrees of freedom			sum of squares			Variance components			Percentage variation			Fixation Index		
Within groups of populations	1	1	1	40,279	6,39	40,21	1,928	0,330	1,835	53,06	61,85	47,51	1 FSC = 0,216	1 FSC = 0,100	1 FSC = 0,277
							Va	Va	Va				2 FST = 0,632	2 FST = 0,657	2 FST = 0,620
													3 FCT = 0,531	3 FCT = 0,618	3 FCT = 0,475
between populations within groups	7	7	7	41,572	3,00	53,24	0,368	0,020 Vb	0,561	10,12	3,81	14,52			
							Vb		Vb						
Within populations	103	100	90	137,78	18,33	132,05	1,338	0,183	1,467	36,82	34,34	37,98	Significance tests		
							Vc	Vc	Vc						
Total													1 P-value = 0,000	1 P-value = 0,000	1 P-value = 0,000
	111	108	98	219,63	27,73	225,51	3,633	0,534	3,864	100	100	100	2 P-value = 0,000	2 P-value = 0,015	2 P-value = 0,000
													3 P-value = 0,112	3 P-value = 0,119	3 P-value = 0,121

Table 3: Spatial Analysis of MOlecular VAriance (SAMOVA). Results: Cyt-B (bold), COI (italics) & concatenated matrix (bold and italics); Significance tests (1000 permutations)

In addition, the Guineo-Congolian zone has less genetic variability than the Sahelo-sudanian's one. Moreover, the differentiation indices GST and NST are statistically different from zero; and NST values are not significantly greater than those of the GST both between biogeographical regions and populations (Table 4). The analysis of haplotype relationships by NCPA (Figure 5) revealed that the historical events that led to the current population structure of *S. zeamais* consisted of restricted gene flow along with isolation by distance and long-distance dispersal events (Table 5).

Phylogenetic signal and relationships, and dating analysis

For the phylogenetic reconstruction methods of the ML and the BI, the trees were inferred using GTR [67,68] evolutionary model with the concatenated matrix. At this model it is added corrections for heterogeneity in site substitution rates (shape parameter α=0.958) and invariants (I=0.348) for Cyt-B but for COI, the model best fitted is the transition model (TIM2) [69]. In all cases, the trees of probabilistic methods showed essentially the same topology. The clade support for some internal nodes was different depending on the gene being analysed and the method being used, which would due to the relative low number of informative sites for each gene. Thus, we used the concatenated matrix and all subsequent analyses were performed using this matrix. There is an important phylogenetic signal within the sequences analyzed according to the inferences of Lyons-Weilers et al. [47] and the mean of Steel's test which is smaller than 0.04. The entropy-based information method of Xia et al. [49] showed so that the sequences used are unsaturated because the index of substitution saturation (Iss=0.0169) is smaller than its critical value (Iss.c=0.7432).
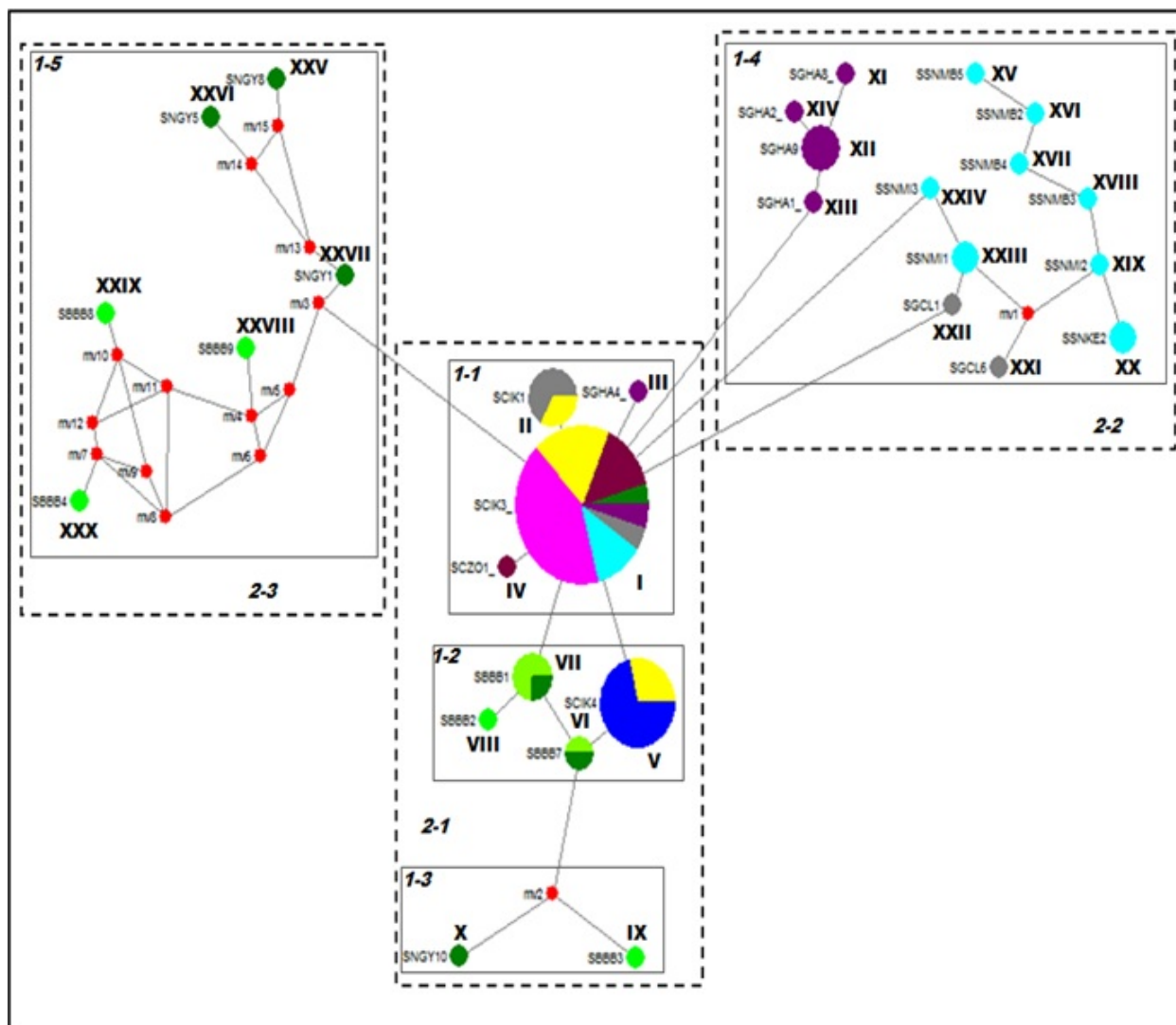
Figure 4: Haplotype network of NCPA (Nested Clade Phylogeographic Analysis) Circles with Roman numbers (I with XXI) refer to haplotypes (0-step clades); boxes in single circuit line are marked 1-1 to 1-4 and represent 1-step clades; tear lines are marked 2-1 to 2-2 indicate 2-step clades; a fine and gray line indicates a connection between haplotypes or n-steps clades.

Concatenated matrix showed a good resolution for solving relationships in the three hierarchical levels analyzed by Likelihood-map analyses. This latter showed a phylogenetic signal where 9.8% of sequences are no resolved against 5.6% ones which are partially resolved and 84.5% entirely resolved. The results of the phylogenetic reconstruction of the probabilistic methods of the cladistic approach (ML and BI) of concatenated matrix presented globally 5 haplogroups (i.e. phylogenetic groups which are monophyletic clusters of haplotypes) (Figures 6-8) corresponding to clusters already identified by the Bayesian analysis of the BAPS.

The trees are represented with leaves colored according to their assignment in the different clusters detected by the BAPS mixture model. However, there are non-resolving haplotypes. The thirty haplotypes identified in the concatenated genes were clustered into four or five clades and showed that some haplotypes of Republic of Guinea are more closely related to some Senegal's haplotypes; and haplotypes to come from Niger are more closely related to ones sampled in Burkina Faso. Bayesian dating showed that the TMRCA (the most recent common ancestor) of the identified haplotypes of *S. zeamais* is aged about 1 MA according to the analysis of concatenated genes (Figures 9-11). So they were diverged during the Pleistocene (2.58 Ma to 11,700 BP). According to the COI and Cyt-B genes, the estimated times of divergence ranged from: 0.13 Ma to 0.55 Ma and 0.10 Ma to 2.35 Ma respectiveley (Figures 10 and 11). The results of Bayesian dating suggested that the two weevil twin species were diverged about 6.52, 8.16 and 8.75 Ma ago respectively for COI, Cyt-B and concatenated matrix from the TMRCA (Figures 9-11). These results were in accordance with those of Corrêa et al. [70].
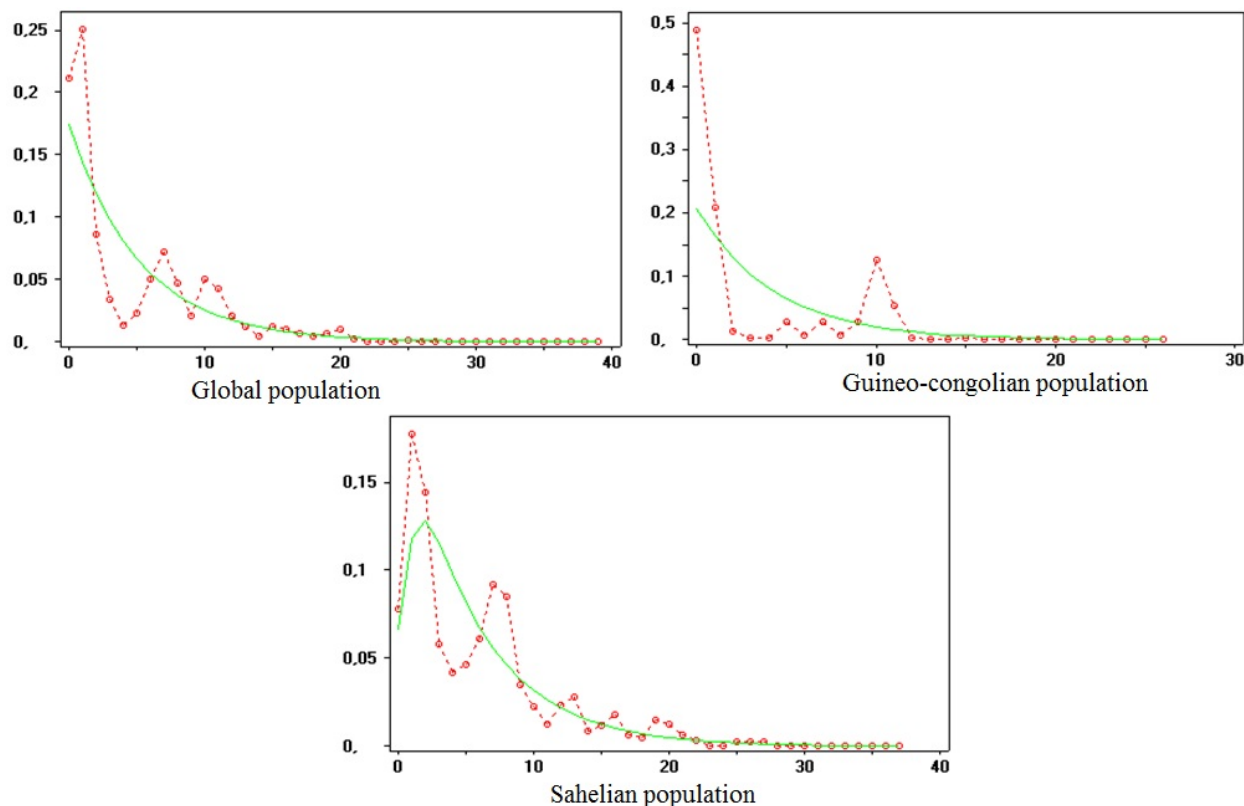
Figure 5: Mismatch distributions analysis by concatenated matrix under expansion model.

Molecular markers	Zones	GST ± (SE)	NST±(SE)	GST *vs* NST	HT ± (SE)	HS ± (SE)	DST	Nm
	GSA vs GGC	0.137	0.090 (NC)	NS	0.674	0.581	0.093	5.225
Cyt-B		(NC)			0	-0.239		
	Between populations	0.365	0.366	NS	0.727	0.462	0.265	0.4359
		-0.0542	(NC)		-0.1411	-0.1297		
	GSA vs GGC	0.167	0.141	NS	0.433	0.361µ	0.153	3.053
COI		(NC)	(NC)		0	-0.1559		
	Between populations	0.381	0.280 (0.1910)	NS	0.452	0.28	0.172	0.407
		-0.2325			-0.0983	-0.0854		
	GSA vs GGC	0.221	0.066	NS	0.854	0.665	0.189	4.201
concatenated matrix		(NC)	(NC)		0	-0.2659		
	Between populations	0.350 (0.1353)	0.369 (0.0526)	NS	0.846	0.55	0.296	0.4645
					-0.0902	-0.1312		

Table 4: Indices of signal phylogeographic, Nei (1973;1987) & flow gene between populations and biogeographic zones. Estimates of average gene diversity within populations (H_S), total gene diversity (H_T), inter-population differentiation (G_{ST}), and number of substitution types (N_{ST}) (mean ± SE (Standard Error) in parentheses); Number of migrants (N_m); GSA vs GGC: Group of Sahelo-soudanian populations versus Group of Guineo-congolian populations; Significance tests: *: $0.05 < P < 0.01$; ***: $0.01 < P < 0.001$; NS: non-significant ($P > 0.05$).

Clades	Inferences chain	Infered Pattern
Clade 1-1	1-2-3-4: NO	Restricted Gene Flow with Isolation by Distance
Clade 1-2	1-2-11-17: NO	Inconclusive outcome
Clade 1-3	1-2-11-17: NO	Inconclusive outcome
Clade 1-4	1-2-11-17: NO	Inconclusive outcome
Clade 1-5	1-2-11-17: NO	Inconclusive outcome
Clade 2-1	1-2-3-5-6-7: YES	Restricted Gene Flow/Dispersal but with some Long Distance Dispersal
Total cladogram	1-2-3-4: NO	Restricted Gene Flow with Isolation by Distance

Table 5: Results of the Nested clade phylogeographic analysis (NCPA).

Demographic evolution

The results of the mismatch analysis of molecular markers, particularly the Cyt-B and concatenated matrix, displayed a mixed pattern for the global population as well for the bio geographical zones with curves showing multimodal distribution followed by unimodal distribution (Figure 11); with not significant SSD and raggedness values; which suggests that the global population had undergone a recent population expansion. These inferences are in accordance with the corresponding Tajima's D and Fu's Fs tests. Indeed, the P-values of these demogenetic tests were negatively no significant; and those of the R2 were positively no significant (Table 6).

Discussion

The objective of this study is to characterize the genetic variability and structure and to retrace the phylogenetic affinities on the one hand and another hand to define the phylogeographic relations between populations of the West and Central African subregion by DNA sequences. It is important to note that a major challenge in the study of the spatio-temporal dynamics of differentiation reside on detection of barriers to dispersal, on identification of environmental constraints that regulate gene flows between populations and their evolution in space and time [71] Moreover, combination of spatial and temporal methods makes it possible to interpret manner more exhaustive the structuring of genetic diversity observed. Indeed, one of the crucial steps in integrated pest management, an alternative to the chemical pathway (dangerous and costly), is the knowledge of the status and characteristics of the implicated populations [72].
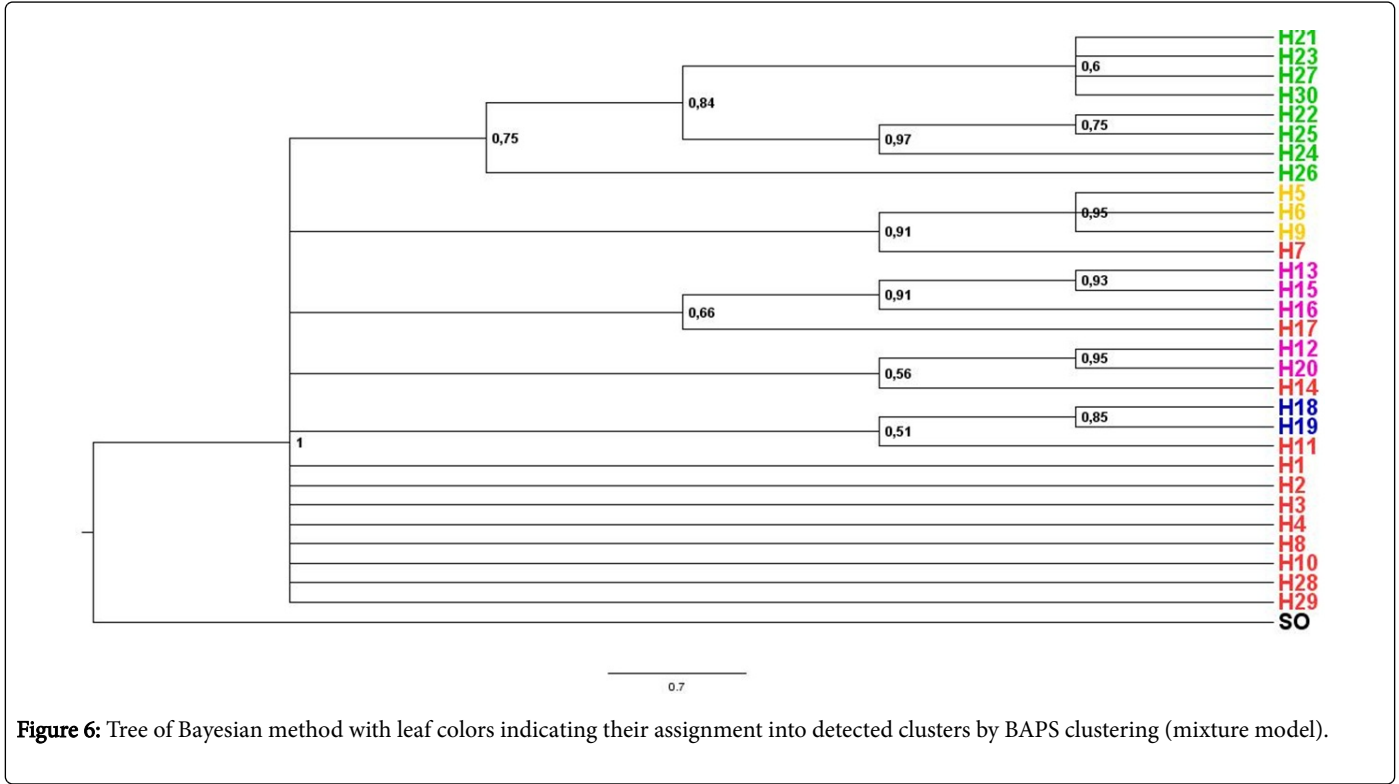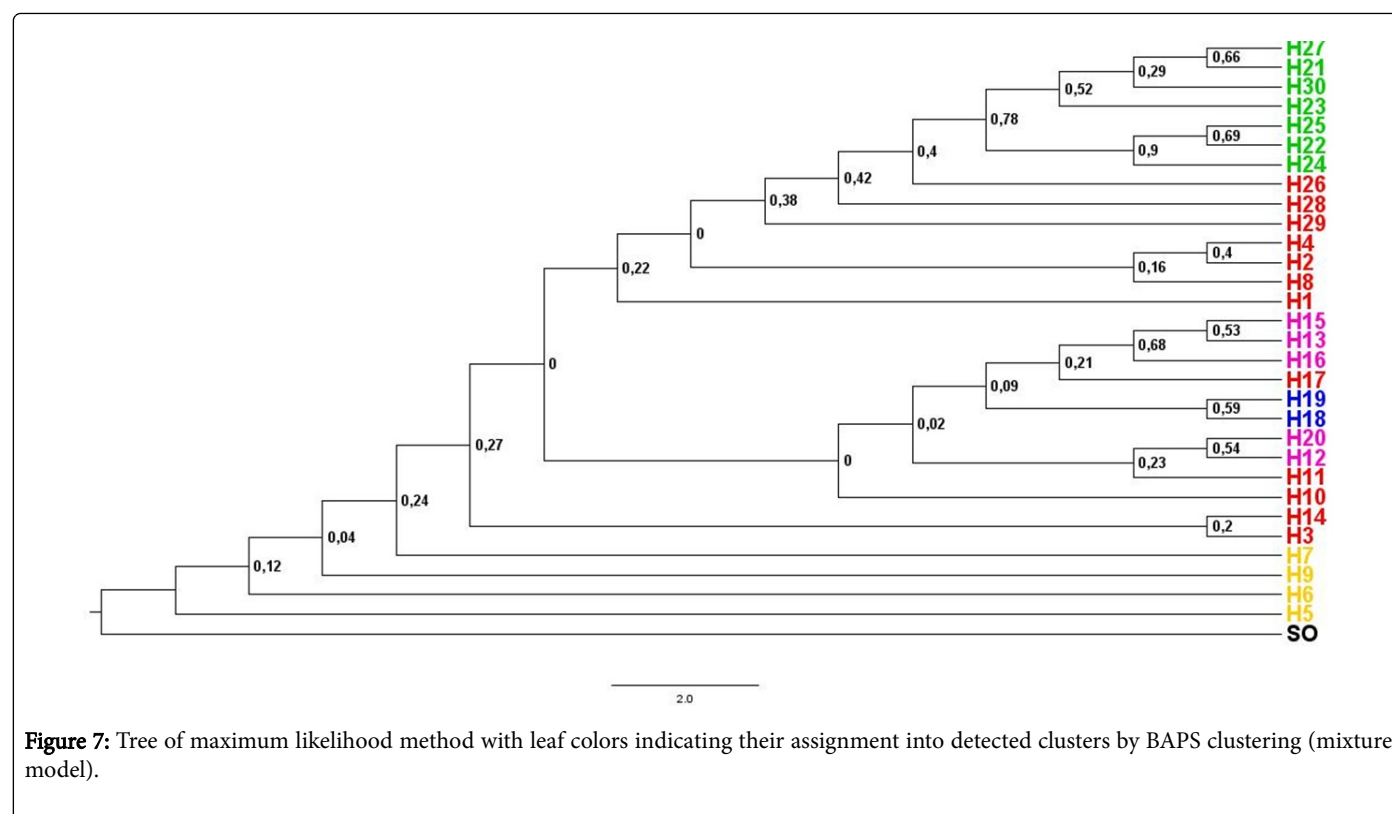

Figure 6: Tree of Bayesian method with leaf colors indicating their assignment into detected clusters by BAPS clustering (mixture model).

Figure 7: Tree of maximum likelihood method with leaf colors indicating their assignment into detected clusters by BAPS clustering (mixture model).

Variability and genetic diversity of populations

The values of the characteristic parameters of the genetic diversity (S, K and π; Table 1) of this study are in the ranges to those observed in others studies carried out in some biological models, such as *Drosophila melanogaster* [73]; and on insect pest such as *Busseola fusca* [74]; *Callosobruchus maculatus* [75]; *Tribolium castaneum* [76]. There is a high A-T content as expected. Indeed, in insects, especially at third codon positions, there are high levels of A+T saturation in mitochondrial sequences (>75%) [77,78]. Both in the global population and subpopulations, the synonymous substitutions are superior to non-synonymous substitutions. The values of global omega (dN/dS ratio) over the entire *Sitophilus zeamais* open reading frame characterizes a purifying selection (ω ≤ 0.3) for the COI and a neutral selection (0.3 < ω < 1) for the Cyt-B and concatenated matrix according to γ-MYN (Modified Yang Nielsen) method [79,80]. This value of omega smaller than 1 could be due to avoidance of deleterious functions, rather than the loss of function. Thus, if a gene is expressed, codon usage, nucleotide bias and other factors (protein toxicity) will generate some purifying selection even if the gene might not have such a function. Vast majority of mutations are either neutral (i.e. have no phenotypic effect), or deleterious. Advantageous mutations are very rare. Thus, usually only parts of a protein are under positive selection while others are under strong purifying selection to maintain the basic structure and function of a protein according to site models; and moreover, usually positive selection acts only during a certain evolutionary time window according to branch models.

The ecological parameters (rainfall, temperature, humidity, etc.) suggested an effect on the nucleotide diversity of maize weevil between populations of the Sahelo-sudanian zone and ones of the Guineo-Congolian regions according the Chi 2 of π, Hd and K. Moreover, the principal components analysis of standard indices of the genetic variability shows that this divergence is mainly due to the differentiation of the populations of Burkina Faso and Niger compared to the other populations. These two populations have, indeed, characteristics typically Sahelian.

Genetic and phylogeographic structure

The results reveal the existence of genetic structure and the lack of phylogeographic signal of *Sitophilus zeamais* in West Africa. Indeed, the AMOVA analysis indicates that the molecular variance due to differentiation between individuals constituting a population is, on average, between 60-70% (P<0,001) compared with 30-40% between populations. While between population groups according to the biogeographical regions; it is between 8 to 14% (P<0,001) depending on the molecular marker. The AMOVA analysis confirms the results of Chi2 on the nucleotide diversity and also the analysis of the matrices of divergence and genetic differentiation between populations. It suggests that the source of the molecular variance was mostly related to individual characteristics des populations and somewhat to the ecological parameters of the two natural regions. Indeed, in the latter case, the fundamental difference between the two biogeographical regions being defined essentially according to their isohyets; the rainfall is being more substantial in the South than in the North. According to Wright et al. [81], the twice the rate of DNA substitutions of tropical trees species compared to species nesting in the highest latitudes would be associated with climate that would influence the metabolism.
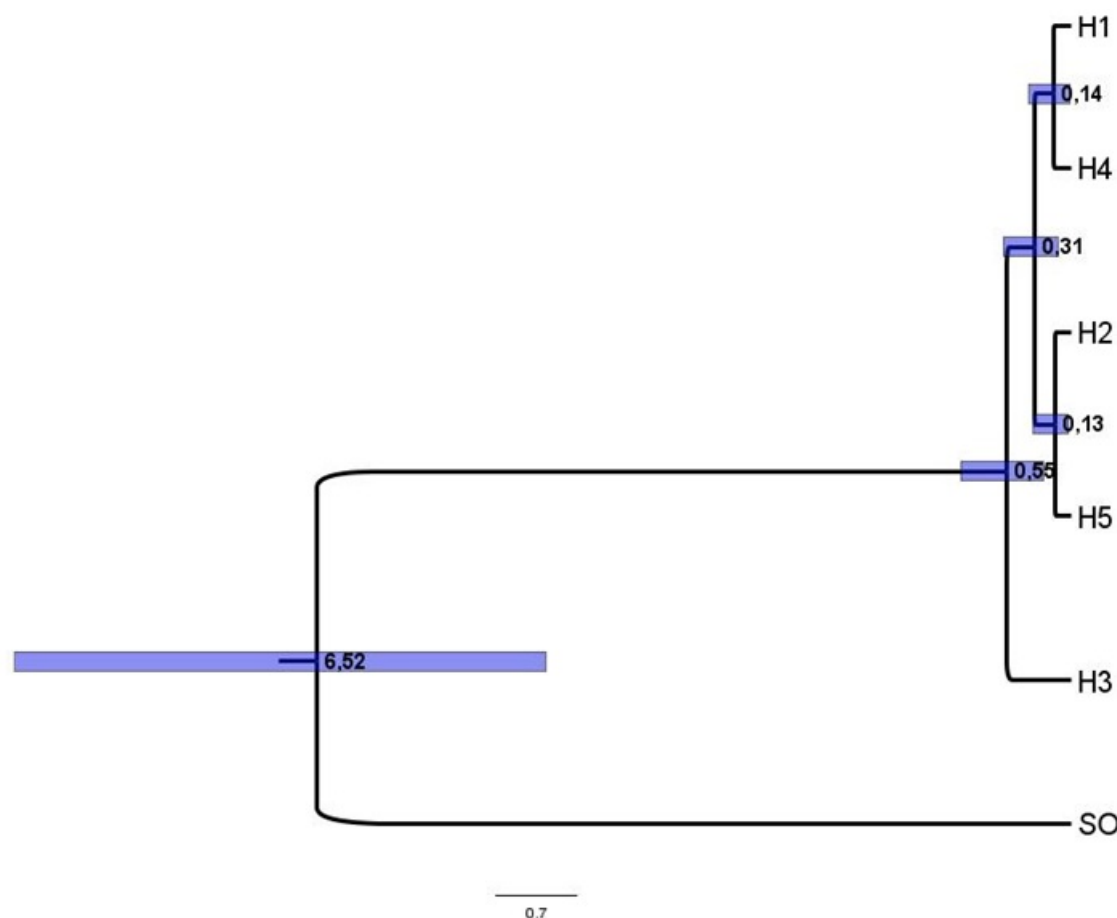
Figure 8: Estimation of times divergence with COI data: the final haplotypes' trees with node bars representing the age 95% HPD intervals of each clade labeled on the nodes.

Thus, this suggests the climate would play a decisive role in biology or even genetic configuration of natural populations. These results also would mostly evoke the implication of the particular characteristics to each population (countries) such as agricultural systems, intensity of commercial activities in the genetic structure of insect pests' populations. Therefore, the genetic structure of populations would also involve to anthropogenic activities. In addition, the analysis of SAMOVA [32], used to highlight genetic barriers between the populations, revealed that there are no genetic barriers between these ones (P<0.05: between Ghana and the other populations (Cyt-B and concatenated matrix) on the one hand; on the other hand, between Mali and the others (COI matrix)). Whereas, the BAPS hierarchical and Bayesian clustering analysis [33] revealed 5, 6 and 4 clusters of haplotypes (haplo-groups) by the mixture method; against 4, 3 and 2 clusters with method of admixture respectively with concatenated matrix, Cyt-B and COI. These results corroborate the presence of a genetic structure but also the existence of a restricted gene flow between populations of *Sitophilus zeamais* in the West African subregion, between biogeographical regions on the one hand and between West Africa and Central Africa where the majority haplotype is the only one present. This gene flow can be explained by the commercial dynamism of the West African subregion. Indeed, the economic importance of cereal crops would have led to the importation of grain by rural populations. This transfer of seeds from one country to another could be accompanied by the transfer of larvae, cocoons, eggs or even adults. And thus a new installation of insects pest in areas not previously affected by these post-harvest losses [5]. This tends to attenuate any genetic differentiation that would lead isolation.

Moreover, the values of NST were not significantly greater than these of GST and the values of both parameters were no null. Indeed, according to Pons and Petit [31] there is a phylogeographic signal when distinct haplotypes mixed in the same population are, on average, more closely related than haplotypes from another population. These comparative values suggested that there is no correspondence between haplotypes similarities and their geographic distribution in *S. zeamais*. Thus, there is no phylogeographical structure. In addition, these results confirmed the presence of a genetic structure. They also correlated with the Mantel tests [24] whose results indicated that Wright's genetic isolation by geographical distance (IBD) [23] was not significant. This showed that genetic differentiation is not explained by isolation. According to Bossart and Prowell [82], genetic isolation may be due to another factor than geographic distance between populations. Indeed, ecological heterogeneity of a habitat of a species often involves barriers to dispersal.
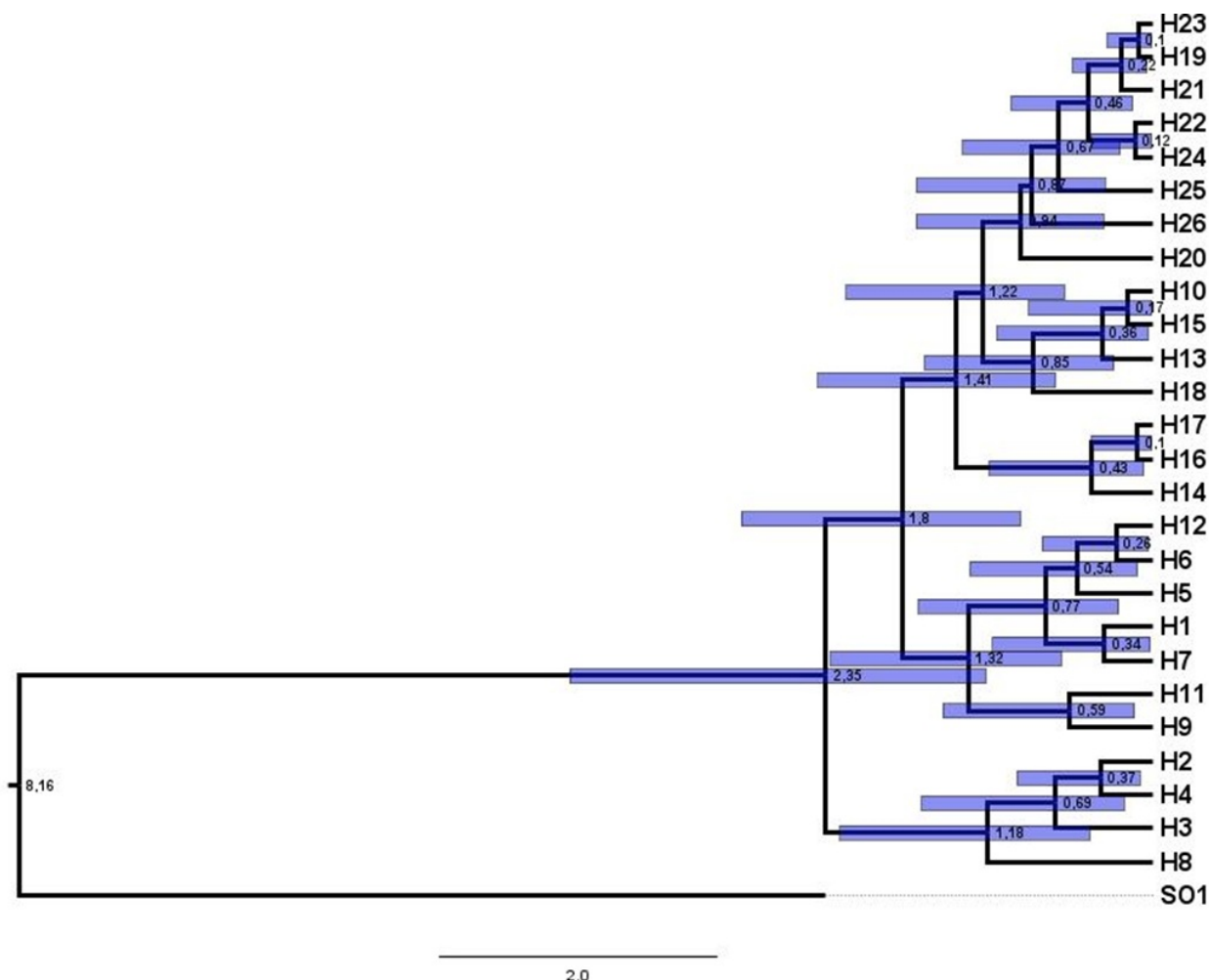
Figure 9: Estimation of times divergence with Cyt-B data: the final haplotypes' trees with node bars representing the age 95% HPD intervals of each clade labeled on the nodes.

Phylogenetic affinities, biogeography and life history evolution

The populations studied showed an important phylogenetic signal. According to Kartavtsev [83,84], effects of saturation and homoplasy on Cyt-B and COI data are small or negligible. The haplotype network revealed a more or less star-shaped form; which suggests a demographic expansion. The majority haplotype could be considered as sub-regional or even interregional haplotype because it were located until in Central Africa where it is essentially the only haplotype encountered in the 2 populations sampled. After the analyses of phylogenetic constructions of concatenated matrix and the evaluation of the topological contractions of the various methods by using the likelihood-maps of Strimmer and Von Haeseler [54], our results suggest that individuals sampled form together 5 phylogenetic groups spread out in West and Central African. Among these 5 haplo-groups identified, one regroups individuals come from in Senegal and Republic of Guinea on the one hand; and on the other hand, one haplo-group put together individuals sampled in Burkina Faso and Niger. Moreover, for Niger and Ghana populations, each one had, in addition, an endemic haplo-group.

The haplo-groups identified in this study showed no correlations with the distinct biogeographic regions. These monophyletic clusters would have diverged to the Pleistocene about 1 million years ago. This period is characterized by climatic instability in Sub-Saharan Africa with alternating hot and humid; cold and dry periods [85]. These Pleistocene climatic oscillations have played an important role in the contemporary diversity of many species and communities [86,87]. Variations of humidity, in particular, have contributed to changes in the distribution of tropical forests [88] and would lead to the existence of refuge areas in the wettest areas [89] for animals. The presence of *Sitophilus zeamais* therefore would predate domestication of maize in the subregion.

Figure 10: Estimation of times divergence with concatenated matrix data: the final haplotypes' trees with node bars representing the age 95% HPD intervals of each clade labeled on the nodes.
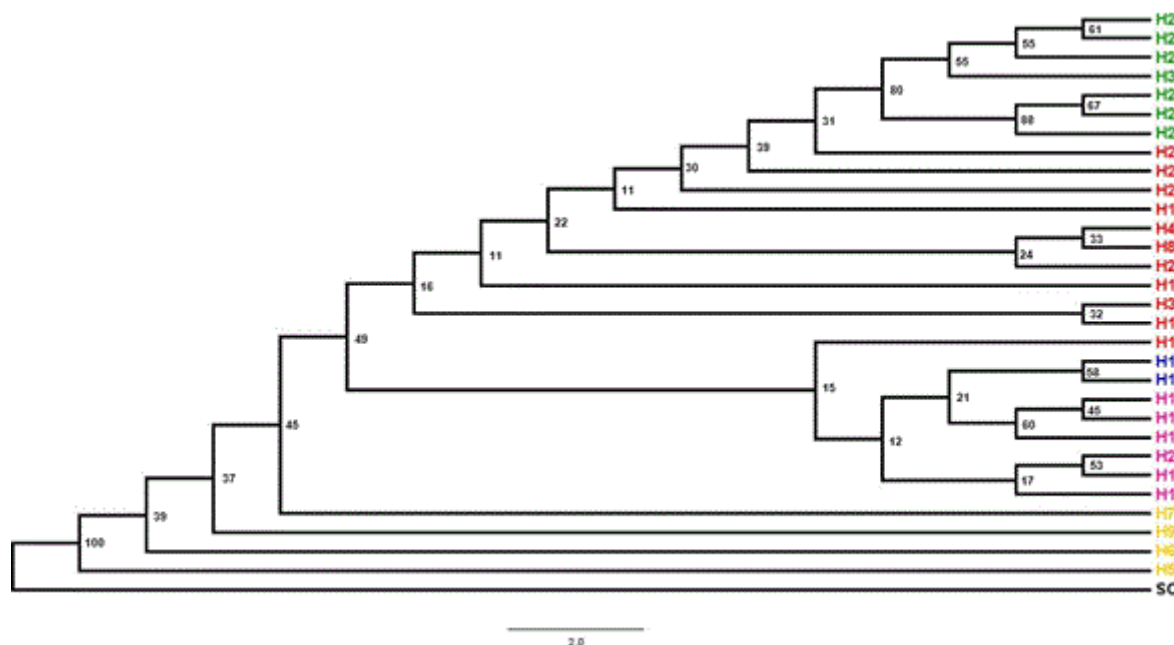

Figure 11: Tree of Neighbor joining method with leaf colors indicating their assignment into detected clusters by BAPS clustering (mixture model).

These inferences are corroborated with results of Sezonlin et al. [74] on a lepidopteran insect, the maize stalk borer (*Busseola fusca*) of which the genetic structure is due to these events of the same period. This same characteristic is also observed in rodents in the subregion [90,91]. It is important to note that even if the configuration of the clusters predates domestication of corn, trade of grains had played a significant role in this configuration. Indeed, in addition to the presence of endemic haplotypes, the populations are also characterized

by the sharing of haplotypes in common, like the majority haplotype. The existence of several endemic haplotypes of *S. zeamais* in biogeographical zones could be explained by a gene flow which could be reduced or even stopped over time, particularly during the Pleistocene events.

Parameters	Geographic groups	N	Demogenetic tests (P-value)				Mismatch distribution tests (P-value)		Expansion coefficient
			Fay & Wu's H	Tajima's D	Fu's Fs	R2	SSD	Rg	S/K
Cyt-B	GSA	60	0.5031	-1.835	-5.7384	0.1007 (0.5170)	0.7493	0.027	10.22
			-0.352	(0.0110)*	-0.034		(0.0000)*	-1	
	GGC	52	0.1221	-0.3139	1.84322	0.1063 (0.5350)	0.04557	0.36355	5.06
			-0.303	-0.396	-0.806		-0.1	-0.44	
	Global Population	112	0.3158	-1.0744	-1.9476	0.1089 (0.5600)	0.39741	0.1953	14.66
			-0.371	(0,2030)	-0.42		-0.05	-0.72	
COI	GSA	55	0.0232 (0.3100)	-0.723	-0.3794	0.1089	0.0234	0.19413	6.78
				-0.288	-0.389	-0.56	-0.07	-0.03	
	GGC	54	0.0704 (0.3120)	-1.8285	-0.2619	0.1083 (0.5190)	0.0023	0.424	16.84
				(0.0100)*	-0.34		-0.42	-0.61	
	Global Population	109	0.0887 (0.3340)	-1.2757	-0.3207	0.0917 (0.5350)	0.01282	0.30905	17.53
				-0.149	-0.3645		-0.245	-0.32	
Concatenated matrix	GSA	50	0.4245 (0.3620)	-1.3668	-7.8354	0.1058 (0.5360)	0.0154	0.0234	7.46
				-0.067	(0.0110)*		-0.95	-0.74	
	GGC	49	0.2519 (0.3470)	-1.8974	0.1722	0.1044 (0.5790)	0.327	0.1362	6.23
				-0.179	-0.581		(0.0000) **	-1	
	Globale Population	99	0.6323 (0.3600)	-1.1321	-3.8316	0.0922 (0.5450)	0.1712	0.0798	12.17
				-0.123	-0.296		-0.475	-0.87	

Table 6: Statistics of past demography. N: number of individuals by population: GSA: Group of Sahelo-sudanian populations; GGC: Group of Guineo-congolian populations; SSD (sum of squared deviation under expansion model); Rg (Harpending's raggedness index); D (Tajima's D test statistic); Fs (Fu's Fs test statistic).

This fact is confirmed by the NCA results which suggested that historical events that led to the current populations' structure of *Sitophilus zeamais* would be result of restricted gene flow between populations and dispersal events. Indeed, history constitutes an element of investigation in determining composition of a community [92] or even current structure of genetic diversity.

The indices of diversity would indicate that the infestation in West Africa would be much older than in Central Africa subregion. Indeed, taking into account the nucleotide diversity which integrates an evolutionary character and the relative high haplotypic divergences [57,93] within zones; the period of evolutionary divergence within group of Sahelo-sudanian populations was longer than the evolutionary period in group of Guineo-Congolian populations.

The most important values of these indices were found in the Sahelo-sudanian zone, particularly in Niger. Therefore, this would evoke that the origin of the infestation be in the North. These results were in accordance with those of Sézonlin et al. [74] and Corrêa et al. [70] Nevertheless, according to Petit et al. [94], Dick et al. [89] and Ley et al. [95] these greater diversities (haplotype and nucleotide) can also reveal both refuge and/or crossing zone of recolonization routes i.e., secondary contact zone. However, this origin and dispersion from North to South would be supported by many statements. The first one was the fact that the majority haplotype, with more of 43% of individuals sampled, was present in this population. Indeed, there was a direct probabilistic relationship between the frequency of an allele and its age [96,97] because the most frequent haplotype would be on average the oldest. The second statement was the expansion which has often led to a population structure in which genetic diversity decreases with distance from the original population [86]. There were also the results of the analysis of clades nesting with clade 1-1 which highlights the model of "restricted but recurrent gene flow" which predicts that

the oldest clades (namely, interior clades) should be more geographically widespread than terminal clades [35]. However, this hypothesis was not in accordance with postulated refugia of humid savannahs in the South (Guineo-Congolian region) [11,98,99] where a small number of lineages would survive and would take refuge during unfavorable Pleistocene (aridity, glaciation) periods in confined areas; and then would colonize other territories during wet (favorable) periods [86,87]. In the West African subregion, for example, there are rodents [100,91]; ungulates [101]; etc. However, during the unfavourable periods of the Pleistocene, there were so species that took refuge under the mountains or rocky massifs instead of the wetlands considered as "safe areas". This was the example of a rodent of the genus Acomys (Rodentia, Muridae) [100] known for its preference for rocky habitat [102].

Hence necessity for more in-depth studies to identify, with greater probability, the origin of the infestation and routes of dispersal of lineages after a return to environmental conditions conducive to expansion. In this sense, interdisciplinary collaboration between, for example, genetics of insect pest populations and archaeo-entomology would be a synergistic effect in order to give a scientific answer to this question about origin of populations of *Sitophilus zeamais* in Africa.

Demographic evolution

Low intraspecific divergences (<1%) and negative values of Tajima's D defined the populations. These results combined with the analysis of the mismatch distributions and demographic indices (SSD and Rg) as well as the diversity indices (Hd and π) and the demogenetic tests (R^2, and Fs de Fu) showed a global population that would have experienced, originally, a bottleneck (during Pleistocene) followed by moderate demographic expansion. This recent population explosion may be due to the colonization of new territories after the last glacial maximum, as has been demonstrated in many species [86,103]. Moreover, the configuration of phylogenetic trees confirmed the demographic expansion of the global population [104,105]. According to the inferences of Grant and Bowen [57] and Avise [58], this rapid expansion was based on a low ancestral population where time is not sufficient to regain a strong diversity between haplotypes. Indeed, according to Lefèvre et al. [106], these phases of severe bottleneck have a sustainable impact on genetic diversity by accentuating genetic drift resulting in the disappearance of many variants which will be regenerated slowly by mutation at the end of this phenomenon. According to these authors, genetic diversity of species would reflect the date and intensity of its last bottleneck and not its maximum abundance at population equilibrium. This demographic characteristic of the global population is also observed, specifically, in the populations of the Sahelo-sudanian zone. Whereas for populations from the Guineo-Congolian zone, Cameroon, CAR and Mali populations where there was little genetic diversity, this was the signal of a recent demographic bottleneck or the result of a founding effect colonization by a small number of lineages [57,58]. The identification of this recent colonization of insect pests of Southern of West African countries such as Cameroon was highlighted by Sézonlin et al. [74] for the Maize Stalk Borer.

The demographic indices (SSD and Rg) of the 3 markers confirmed the demographic expansion of the global population. However the mismatch curves (Cy-B and COI) showed a mixed curve (multimodal and unimodal). These multimodal peaks may appear in the mismatch distribution, despite the expansion of the population, if a genetic structure was noted in a population from which these discrepancies

appear [107], which was the case of our studied population. In addition, it is also important to note that the demogenetic tests are similar in their mechanisms, but differ in what they actually measured and compared. Thus, for example, there may be significantly negative values for Fu's Fs and not Tajima D's. While in an ideal case, all tests should agree. The existence of these discrepancies may also be due, in addition, to population reduction, recent bottleneck or migration that has resulted in secondary contact after differentiation of previously separated groups [108,109]. From these results, there are two threats: an expansion of insects, therefore a mass depreciation of cereal stocks, and a very probable emergence of strains resistant to control methods, especially chemical control. This emergence would be favoured by the existence of this high genetic variability within the populations. Indeed, Nei's indices [26,110] showed that the contribution of mean genetic diversity within population (HS) to total genetic diversity (HT) is higher than that one of inter-population genetic diversity (DST), both between populations and between biogeographical regions, whatever molecular marker. These values of the genetic diversity would evoke a multidirectional gene flow and would be a consequence of trade of seeds. Indeed, importation of haplotypes from other regions would explain the presence of the excessive number of haplotypes infrequent and somewhat divergent. According to Frankham [111], Nei's genetic diversity is positively correlated with population size. A strong genetic diversity is correlated to life history traits of R-strategy species [112,113] which produce many descendants, such as insects. Moreover, according to Kimura [50], in mitochondrial DNA, transitions can represent 90% of the mutations. However, our results showed higher transversions rates than transitions. Which would may evoke an acquisition of resistance by a changement of configuration (Cyt-B and COI) of molecule, because this type of mutations leads to a change in the configuration of molecule by the replacement of a purine molecule with a pyrimidine molecule, and vice versa?

Populations of *Sitophilus zeamais* can therefore be considered as having a very important genetic pool, which provides considerable potential for adaptation and survival and would be a disadvantage for food security. Because there are an interspecific competition for vegetable resources between insect pests and humans.

Conclusion

In managing natural populations, identifying of a weakness in the evolutionary biology of insect pests through genetic variability is fundamental. Indeed, genetic variability represents the fundamental link of biodiversity because the destiny of populations, their appearance, as well as the extinction of species, depend on it. This study revealed that populations of *Sitophilus zeamais* are composed of several common haplotypes but they are also characterized by endemic haplotypes, which would explain the differentiation and genetic structure that exists between them. The genetic divergence between individuals of a population concentrates the essential of the molecular variance, followed by those between populations in a bio-geographical region; then that one between natural regions. It is also noted the absence of a Phylogeographic signal. There is high haplotype diversity, with an excessive number of haplotypes that are infrequent and somewhat divergent. Therefore, ecological parameters, anthropogenic activities (trade) and characteristics of the agricultural systems of the countries would play an important role in the expansion process of the maize weevil populations. This strong genetic variability within populations and the higher transversions rates than transitions in mitochondrial genome. It would be noted the existence of 2 threats: an

expansion of insects, thus massive depreciation of stocks of cereals; and a very probable emergence of resistant strains. Hence, the need to define policies for the monitoring of standards of trade and cooperation on modalities of commercial exchange of seeds to attenuate or even eliminate infestation of crops and stored foodstuffs. Moreover, the center of infestation and dispersal would be in the Sahelo-sudanian zone, more precisely in Niger since the Pleistocene.

The vast majority of molecular studies concern American and European populations and few are devoted to African ones. Our study provides new insights into the genetic structure and the phylogeography of *Sitophilus zeamais* on Africa sub-Saharan, therefore constitutes a contribution in the knowledge of repartition of ecotypes of the sub-Saharan entomofauna.

In the prospects, a study with wider sampling in Sub-Saharan Africa, combining several markers (morphological, genetic) would contribute better to characterize populations.

Acknowledgements

References

1. Nukenine EN, Monglo B, Awason I, Ngamo LST, Thuenguem FFN, et al. (2002) Farmer's perception on some aspects of maize production and infestation levels of stored maize by *Sitophilus zeamais* in the Ngaoundéré region of Cameroon. Cam J Biol Biochem Sci 12: 18-30.

2. Macauley H, Ramadjita T (2015) Cereal crops: rice, maize, millet, sorghum and wheat. Available on https://www.afdb.org/fileadmin/uploads/afdb/Documents/Events/DakAgri2015/.

3. FAOSTAT (2015) Food and Agriculture Organization of the United Nations, Available on http://Faostat.fao.org.

4. Delobel A (1995) The shift of *Caryedon serratus* from wild *Caesalpiniaceae* to groundnuts took place in West Africa (*Coleoptera:Bruchidae*). J Stored Prod Res 31: 101-102.

5. Ndiaye (2014) Genetic characterization, evaluation of gene flow for different ecotypes of *caryedon serratus*, peanut pest. Doctoral thesis, Cheikh Anta Diop, University of Dakar, Senegal.

6. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, et al. (2001) Terrestrial Eco regions of the world: a new map of life on earth. Biosci 51: 933-938.

7. White F (1983) The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. Natural Resources Research 20, UNESCO, Paris.

8. Arbonnier M (2004) Trees, Shrubs and Lianas of West African Dry Zones CIRAD, Margraf Publishers.

9. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meteorol Z 15: 259-263.

10. Linder HP, Klerk HM de, Born J, Burgess ND, Fjeldsa J, et al. (2012) The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. J Biogeogr 75: 1189-1205.

11. Jean M (1996) The African rain forest: main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary: In-Alexander IJ, Swaine MD, Watling R: Essays on the ecology of the Guinea-Congo rain forest. Proc R Soc Edinb 104: 31-73.

12. Epule TE, Ford JD, Lwasa Sh, Lepage L (2017): Climate change adaptation in the Sahel. Env Sci 75: 121-137.

13. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, et al. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87: 651-702.

14. Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98.

15. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

16. Kumar S, Stecher G, Tamura K (2016): MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33: 1870-1874.

17. Librado P, Rozas J (2009) DnaSP ver 5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

18. Fox J (2005) The R Commander: A Basic Statistics Graphical User Interface to R. J Stat Softw 14: 1-42.

19. R Core Team (2015) R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

20. Pond SK, Muse SV (2005) Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22: 2375-2385.

21. Pond SK, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679.

22. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, et al. (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol 30: 1196-1205.

23. Wright S (1943) Isolation by distance. Genet 28: 114-138.

24. Mantel N (1967) Detection of disease clustering and a generalized regression approach. Cancer Res 27: 209-220.

25. Addinsoft (2015) XLSTAT software, Paris, France.

26. Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70: 3321-3323.

27. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evol 38: 1358-1370.

28. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genet 131: 479-491.

29. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10: 564-567.

30. Suzuki R, Shimodaira H (2014) pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 1.2-2, URL http://CRAN.R-project.org/package=pvclust.

31. Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144: 1237-1245.

32. Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11: 2571-2581.

33. Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. Math Biosci 205: 19-31.

34. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37-48.

35. Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genet 140: 767-782.

36. Posada D, Crandall KA, Templeton AR (2000) Geodis : A program for the cladistic nested analysis of the geographic distribution of genetic haplotypes. Mol Ecol 9: 487-488.

37. Templeton AR (2011) Inference key for the nested haplotype tree analysis of geographical distances. Available on http://darwin.uvigo.es/download/geodisKey_06Jan11.pdf.

38. Hey J, Machado CA (2003) The study of structured populations-new hope for a difficult and divided science. Nature Rev Genet 4: 535-543.

39. Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. Mol Ecol 13: 789-809.

40. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstruction of phylogenetic trees. Mol Biol Evol 4: 406-425.

41. Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17: 368-376.

42. Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17: 754-755.

43. Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. Biometrika 76: 297-307.

44. Darriba D, Taboada GL, Doallo R, Posada D (2012) JModel Test 2: more models, new heuristics and parallel computing. Nat Methods 9: 772.

45. Steel MA, Lockhart PJ, Penny D (1993) Confidence in evolutionary trees from biological sequence data. Nature 364: 440-442.

46. Xia X (2013) DAMBE 5: A comprehensive software package for data analysis in molecular biology and evolution. Mol Biol Evol 30: 1720-1728.

47. Lyons-Weiler J, Hoelzer GA, Tausch RJ (1998) Optimal outgroup analysis. Biol J Linnean Soc 64: 493-511.

48. Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE in Philippe Lemey, Marco Salemi and Anne-Mieke Vandamme, eds. The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. 2nd edition Cambridge University Press.

49. Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) - An index of substitution saturation and its application. Mol Phylogenet Evol 26: 1-7.

50. Kimura M (1980): A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111-120.

51. Felsenstein J (2005) PHYLIP (phylogeny inference package) version 3.6 Distributed by Author. Department of Genome Sciences, University of Washington, Seattle, Available from : http://evolution.genetics.washington.edu/phylip.html.

52. Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704.

53. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

54. Strimmer K, von Haeseler A (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 94: 6815-6819.

55. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.

56. Ramos-Onsins SE, Rozas J (2000) Statistical properties of new neutrality tests against population growth. Mol Biol Evol 19: 2092-2100.

57. Grant WS, Bowen BW (1998) Shallow Population Histories in Deep Evolutionary Lineages of Marines Fishes: Insights from Sardines and Anchovies and Lessons for Conservation. The Am Genet Assoc 89: 415-426.

58. Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA.

59. Peck DR, Congdon DC (2004) Reconciling historical processes and population structure in the sooty tern *Sterna fuscata*. J Avian Biol 35: 327-335.

60. Harpending HC (1994) - Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum Biol 66: 591-600.

61. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genet 123: 585-595.

62. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genet 147: 915-925.

63. Drummond AJ, Rambaut A, Suchard MA (2015) BEAST: Bayesian Evolutionary Analysis by Sampling Trees. V 1.8.2, Available on: http://beast.bio.ed.ac.uk/.

64. Brower AVZ (1994): Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc Natl Acad Sci USA 91: 6491-6495.

65. Farrell BD, Sequeira AS, O'Meara BC, Normark BB, Chung JH, et al. (2001) The evolution of agriculture in beetles (*Curculionidae*: *Scolytinae* and *Platypodinae*). Evol 55: 2011-2027.

66. Rambaut A (2014) Figtree: a graphical viewer of phylogenetic trees. V 1.4.2, Available on: http://tree.bio.ed.ac.uk/software/figtree.

67. Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci 17: 57-86.

68. Rodriguez F, Oliver J, Marin A, Medina J (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142: 485-501.

69. Posada D, Baxevanis AD, Davison DB, Page RDM, Petsko GA, et al. (2003) Using Modeltest and PAUP* to select a model of nucleotide substitution, Current Protocols in Bioinformatics, New York: John Wiley & Sons.

70. Corrêa AS, Vinson CC, Braga LS, Guedes RNC, de Oliveira LO (2016) Ancient origin and recent range expansion of the maize weevil *Sitophilus zeamais*, and its genealogical relationship to the rice weevil S. oryzae. Bull Entomol Res 107: 9-20.

71. Fontaine MC (2008) Structure and evolution of *Harbor Porpoise* populations in Western Palearctic waters. Doctoral thesis, Faculty of Sciences DSGE, Oceanology, University of Liège.

72. Sembène M (2000) Variability of Transcribed Internal Spacer (ITS) of ribosomal DNA and polymorphism of microsatellite loci in cedar *Caryedon serratus* (Olivier, 1790): differentiation in host races and peanut infestation in Senegal. PhD thesis at the Sciences Cheikh Anta Diop, University of Dakar, Senegal.

73. Andolfatto P (2001) Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol 18: 279-290.

74. Sézonlin M, Dupas S, Le Rü B, Le Gall P, Moyal P, et al. (2006) Phylogeography and population genetics of the maize stalk borer *Busseola fusca* (*Lepidoptera*, *Noctuidae*) in sub-Saharan Africa. Mol Ecol 15: 407-420.

75. Kébé K, Alvarez N, Espíndola A, Justy F, Olivieri I, et al. (2016) Insights into the genetic structure of the cowpea pest *Callosobruchus maculatus* in Africa. J Pest Sci 89: 449.

76. Diome T, Thiaw C, Sarr M, Ndong A, Kane M, et al. (2013) Genetic diversity of *Tribolium castaneum* (Herbst) population in storage infrastructures and agro ecological zones in Senegal. Int J Biosci 9: 248-258.

77. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41: 353-355.

78. Pedersen BV (2002) European bumblebees (Hymenoptera: Bombini): Phylogenetic relationships inferred from DNA sequences. Insect Syst Evol 33: 361-386.

79. Wang DP, Wan HL, Zhang S, Yu J (2009) Gamma-MYN: a new algorithm for estimating Ka and Ks with consideration of variable substitution rates. Biol Direct 4: 20.

80. Peters AE, Bavishi A, Cho H, Choudhary M (2012) Evolutionary constraints and expression analysis of gene duplications in *Rhodobacter sphaeroides*. BMC Res Notes 5: 192.

81. Wright S, Keeling J, Gillman L (2006) The road from Santa Rosalia: A faster tempo of evolution in tropical climates. J Evol 93: 7718-7722.

82. Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitation, lessons, and new directions. Trends Ecol Evol 13: 202-206.

83. Kartavtsev YP (2011) Sequence divergence at mitochondrial genes in animals: applicability of DNA data in genetics of speciation and molecular phylogenetics. Mar Genomics 4: 71-81.

84. Kartavtsev YP (2011) Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. Mitochondrial DNA 22: 55-65.

85. Wagner T (2002) Late Cretaceous to early Quaternary organic sedimentation in the eastern Equatorial Atlantic. Palaeogeogr Palaeoclimatol Palaeoecol 179: 113-114.

86. Hewitt GH (2000) The genetic legacy of Quaternary ice ages. Nature 405: 907-913.

87. Hewitt GH (2004) Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci 359: 183-195.

88. Hammond DS (2005) Tropical forests of the guiana shield Ancient forest in a modern world: Biophysical features of the Guiana Shield. Ed Hammond DS 15-194.

89. Dick CW, Abdul Salim K, Bermingham E (2003) Molecular Systematic Analysis Reveals Cryptic Tertiary Diversification of a Widespread Tropical Rain Forest Tree. Am Nat 162: 691-703.

90. Quérouil S, Verheyen E, Dillen M, Colyn M (2003) Patterns of diversification in two African forest shrews: *Sylvisorex johnstoni* and *Sylvisorex ollula* (*Soricidae*, *Insectivora*) in relation to paleo-environmental changes. Mol Phylogenet Evol 28: 24-37.

91. Bryja J, Granjon L, Dobigny G, Patzenhauerová H, Konečný A, et al. (2010) Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (*Rodentia*): the environment/geography/genetic interplay. Mol Ecol 19: 4783-4799.

92. Ricklefs RE (1987) Community diversity: Relative Roles of Local and Regional Processes. Science 235: 167-171.

93. Oremus M, Garrigue C, Cleguer C (2011) Isolation and genetic diversity of New Caledonia *dugongs*. Unpublished report for the Direction of the Environment of the South Province, New Caledonia.

94. Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, et al. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300: 1563-1565.

95. Ley AC, Dauby G, Köhler J, Wypior C, Röser M (2014) Comparative phylogeography of eight herbs and lianas (*Marantaceae*) in central African rainforests. Front Genet 5: 403.

96. Watterson GA, Guess HA (1977) Is the most frequent allele the oldest? Theor Popul Biol 11: 141-160.

97. Donnelly P, Tavaré S (1986) The ages of alleles and a coalescent. Adv Appl Prob 18: 1-19.

98. Maley J (2001) The impact of arid phases on the African rain forest through geological history: In: African rain forest ecology and conservation (eds Weber W, White L, Vedder A, Naughton-Treves L) 68-87, Yale University Press, New Haven, Connecticut.

99. Léal ME (2004) The African rain forest during the Last Glacial Maximum, an archipelago of forests in a sea of grass. PhD thesis, Wageningen University, Wageningen.

100. Nicolas V, Granjon L, Duplantier JM, Cruaud C, Dobigny G (2009) Phylogeography of spiny mice (genus *Acomys*, *Rodentia*: *Muridae*) from the south-western margin of the Sahara with taxonomic implications. Biol J Linnean Soc 98: 29-46.

101. Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. Mol Ecol 21: 3656-3670.

102. Volobouev VT, Aniskin VM, Lecompte E, Ducroz JF (2002). Patterns of karyotype evolution in complexes of sibling species within three genera of African murid rodents inferred from the comparison of cytogenetic and molecular data. Cytogenet Genome Res 96: 261-275.

103. Maley J (1991) The African rain forest vegetation and palaeo-environments during late Quaternary. Clim Change 19: 79-98.

104. Ramirez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. Genetics 179: 555-567.

105. Girod (2010) Genetic consequences of Quaternary climatic variations and distribution of Neotropical forest species: The example of the palm tree Astrocaryum sciophilum. Ecol Environ, Pierre and Marie Curie University, Paris.

106. Lefevre Th, Raymond M, Thomas F (2016) Evolutionary biology. 2nd edition.

107. Maltagliati F, Di Giuseppe G, Barbieri M, Castelli A, Dini F (2010) Phylogeography and genetic structure of the edible sea urchin *Paracentrotus lividus* (*Echinodermata*: *Echinoidea*) inferred from the mitochondrial cytochrome b gene. Biol J Linnean Soc 100: 910-923.

108. Maruyama T, Fuerst PA (1984) Population bottlenecks and non-equilibrium models in population genetics I Allele numbers when populations evolve from zero variability. Genet 108: 745-763.

109. Maruyama T, Fuerst PA (1985) Population bottlenecks and non-equilibrium models in population genetics II Number of alleles in a small population that was formed by a recent bottleneck. Genet 111: 675-689.

110. Nei M (1987) Molecular evolutionary genetics, Columbia University Press, New York. NY, USA.

111. Frankham R (1996) Relationship of Genetic Variation to Population Size in Wildlife. Conservation Biology 10: 1500-1508.

112. MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton.

113. Galtier N, Romiguier J (2015) Parental investment predicts genetic diversity in animal species. Med Sci 31: 17-19.