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

**NDIAYE Mama Racky**,1,2 THIAW Cheikh1,3 and Mbaké Sembène1,2

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

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

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

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

**Receiving date:** Oct 15, 2017; **Acceptance date:** Oct 24, 2017; **Publication date:** Oct 30, 2017

**Copyright:** © 2017 Racky NM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**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 phylogenetic 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), haplotype network, phylogenetic relationships, demographic tests and Nested Clade phylogenic relationships and Phylogeographic structuring of maize weevil populations in West and Central Africa. The latter 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*) 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 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 groups 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 semi-arid 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-GGATCCACTGATAATAGCATTCCC-3) and CB2 (5-CCTGGAATTTAAATATAATTTACTTC-3) for Cytochrome B gene (Cyt-B); Ron II (5-TATGTACTACCATGAGGACAAATATC-3) and Nancy (5-ATTACACCTCCTAATTTATTAGGAAT-3) for Cytochrome oxidase 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 MgCl₂, 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 nonsynonymous (dN) and synonymous (ds) substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny as implemented in Hy Phy, 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 diversifications 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 used rainfall parameters as structuring factors. A hierarchic 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 NCPS**

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 NCPS (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 NCPS 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*.
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).

<table>
<thead>
<tr>
<th>Geographical groups</th>
<th>Climate</th>
<th>N</th>
<th>Nµ</th>
<th>S</th>
<th>SIP</th>
<th>SS</th>
<th>K</th>
<th>HT</th>
<th>HP</th>
<th>%HP</th>
<th>Hd (± SD)</th>
<th>Π (± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>58</td>
<td>51</td>
<td>28</td>
<td>23</td>
<td>4.9898</td>
<td>22</td>
<td>20</td>
<td>90.91</td>
<td>0.8203</td>
<td>0.01129</td>
</tr>
<tr>
<td>semi-arid to semi-humid</td>
<td></td>
<td>55</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.5899</td>
<td>4</td>
<td>2</td>
<td>50</td>
<td>0.5165</td>
<td>0.00142</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>48</td>
<td>43</td>
<td>29</td>
<td>14</td>
<td>5.763</td>
<td>24</td>
<td>21</td>
<td>87.5</td>
<td>0.922</td>
<td>0.00673</td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>11</td>
<td>2.3718</td>
<td>6</td>
<td>4</td>
<td>66.66</td>
<td>0.3424</td>
<td>0.00537</td>
</tr>
<tr>
<td>GGC</td>
<td></td>
<td>54</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>0.3564</td>
<td>3</td>
<td>1</td>
<td>33.33</td>
<td>0.2047</td>
<td>0.00086</td>
</tr>
<tr>
<td>humid to subhumid</td>
<td></td>
<td>49</td>
<td>18</td>
<td>18</td>
<td>12</td>
<td>6</td>
<td>2.889</td>
<td>9</td>
<td>6</td>
<td>66.66</td>
<td>0.512</td>
<td>0.00337</td>
</tr>
<tr>
<td></td>
<td></td>
<td>112</td>
<td>68</td>
<td>58</td>
<td>36</td>
<td>22</td>
<td>3.9574</td>
<td>26</td>
<td></td>
<td></td>
<td>0.6448</td>
<td>0.00895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>109</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>8</td>
<td>0.5134</td>
<td>5</td>
<td></td>
<td></td>
<td>0.3986</td>
<td>0.00124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99</td>
<td>63</td>
<td>56</td>
<td>37</td>
<td>19</td>
<td>4.602</td>
<td>30</td>
<td></td>
<td></td>
<td>0.789</td>
<td>0.00538</td>
</tr>
<tr>
<td>Chi2 test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.039</td>
<td>-0.00082</td>
</tr>
<tr>
<td>(GSA vs. GGC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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; Nµ: number of mutations; S: number of segregating sites; SIP: Parsimony informative sites; SS: singleton sites; K: mean pair-wise nucleotide; HT: number of total haplotypes; HP: number of private haplotypes; %HP: 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).
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 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**
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 (GS) and the populations group of the Guinea-Congolese zone (GCC), 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 Guinea-Congolese one by identifying 5 (mixture model) or 4 (admixture model) clusters (Figure 4).

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)
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).
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 not 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 respectively (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].
Table 4: Indices of signal phylogeographic, Nei (1973;1987) & flow gene between populations and biogeographic zones. Estimates of average gene diversity within populations (Hs), total gene diversity (Ht), inter-population differentiation (GST), and number of substitution types (NST) (mean ± SE (Standard Error) in parentheses); Number of migrants (Nm); 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 | Inferred Pattern |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade 1-1</td>
<td>1-2-3-4: NO</td>
<td>Restricted Gene Flow with Isolation by Distance</td>
</tr>
<tr>
<td>Clade 1-2</td>
<td>1-2-11-17: NO</td>
<td>Inconclusive outcome</td>
</tr>
<tr>
<td>Clade 1-3</td>
<td>1-2-11-17: NO</td>
<td>Inconclusive outcome</td>
</tr>
<tr>
<td>Clade 1-4</td>
<td>1-2-11-17: NO</td>
<td>Inconclusive outcome</td>
</tr>
<tr>
<td>Clade 1-5</td>
<td>1-2-11-17: NO</td>
<td>Inconclusive outcome</td>
</tr>
<tr>
<td>Clade 2-1</td>
<td>1-2-3-5-6-7: YES</td>
<td>Restricted Gene Flow/Dispersal but with some Long Distance Dispersal</td>
</tr>
<tr>
<td>Total cladogram</td>
<td>1-2-3-4: NO</td>
<td>Restricted Gene Flow with Isolation by Distance</td>
</tr>
</tbody>
</table>

Table 5: Results of the Nested clad 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 dispersion, 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].
Variability and genetic diversity of populations

The values of the characteristic parameters of the genetic diversity $(S, K$ and $\pi$; 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 ($\omega \leq 0.3$) for the COI and a neutral selection ($0.3 < \omega < 1$) for the Cyt-B and concatenated matrix according to $\gamma$-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 $\pi$, 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.
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.
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 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.
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.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Demogenetic tests</th>
<th>Mismatch distribution tests</th>
<th>Expansion coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(P-value)</td>
<td>(P-value)</td>
<td></td>
</tr>
<tr>
<td>Geographic groups</td>
<td>N</td>
<td>Fay &amp; Wu’s H</td>
<td>Tajima’s D</td>
</tr>
<tr>
<td>GSA</td>
<td>60</td>
<td>0.5031</td>
<td>-1.835</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.352</td>
<td>(0.0110)*</td>
</tr>
<tr>
<td>Cyt-B</td>
<td>52</td>
<td>0.1221</td>
<td>-0.3139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.303</td>
<td>-0.396</td>
</tr>
<tr>
<td>Global Population</td>
<td>112</td>
<td>0.3158</td>
<td>-1.0744</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.371</td>
<td>(0.2030)</td>
</tr>
<tr>
<td>GSA</td>
<td>55</td>
<td>0.0232 (0.3100)</td>
<td>-0.723</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.288</td>
<td>-0.389</td>
</tr>
<tr>
<td>COI</td>
<td>54</td>
<td>0.0704 (0.3120)</td>
<td>-1.8285</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0100)*</td>
<td>-0.34</td>
</tr>
<tr>
<td>Global Population</td>
<td>109</td>
<td>0.0887 (0.3340)</td>
<td>-1.2757</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.149</td>
<td>-0.3645</td>
</tr>
<tr>
<td>GSA</td>
<td>50</td>
<td>0.4245 (0.3620)</td>
<td>-1.3668</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.067</td>
<td>(0.0110)*</td>
</tr>
<tr>
<td>Concatenated matrix</td>
<td>GGC</td>
<td>49</td>
<td>0.2519 (0.3470)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.179</td>
<td>-0.581</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1.1321</td>
<td>-3.316</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.123</td>
<td>-0.296</td>
</tr>
<tr>
<td>Global Population</td>
<td>99</td>
<td>0.6323 (0.3600)</td>
<td>-1.1312</td>
</tr>
</tbody>
</table>

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 would 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 savannas 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 θ) and the demogenetic tests (R², 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 phylogeographic 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 experienced, originally, a bottleneck (during Pleistocene) followed by maximum, as has been demonstrated in many species [86,103].

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 caused by a colonization event 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

This publication was produced with support from the project ERA (Education and Research in Agriculture) of U.S. government agency: USAID and from the WAAPP ("West Africa Agricultural Productivity Program") for the project "Insect pests of grain". We also thank the Francophone University Association (AUF). Acknowledgements are also due to Drs Penda NDIAYE, Arame NDIAYE, Toffène DIOME and Fatimata MBAYE for help in the production and analysis of data; and for reviewing.

References


