Barcoding: Comparison of Variation Degree of COI and Cytochrome b Mitochondrial Markers in Two Species Primary Maize Pests (*Sitophilus zeamais* and *Sitophilus oryzae*)

Assane NDONG[a]*, Cheikh THIAW[b], Babacar Diallo[c], Malick SARR[d], Toffène DIOME[e], Mamadou KANE[f], Mbacké SEMBENE[g]

[a, c, d,e,g] *Faculty of sciences and Technology, Department of Animal Biology, University Cheikh Anta DIOP, PO Box 5005 Dakar, Senegal*

[b,d]*Senegalese Institute of Agricultural Research; Regional Center for the Study on Improvement of Adaptation to Drought (ISRA/CERAAS), P.O. Box 3320 Thies Escale, Senegal.*

[a,f,g]*BIOPASS Laboratory, UMR 022 IRD-CBGP, Bel-Air, PO Box 1386, Dakar, Senegal*

[a]*Email: assanendong85@yahoo.fr*
[b]*Email: thiacheikh@hotmail.com*
[c]*Email :babspalpalis@gmail.com*
[d]*Email: malsarr777@yahoo.fr*
[e]*Email: eltof@yahoo.fr*
[f]*Email: mamadou.kane@ird.fr*
[g]*Email: mbacke.sembene@ird.fr*

Abstract

The phylogeny, taxonomy and identification of insects were originally based upon shared or derived morphological and anatomy criteria. However, these processes are confronted with the barrier of the existence of sympatric or cryptic species.

----

* Corresponding author.

E-mail address: assanendong85@yahoo.fr

Thus, genetic analyses have more recently played an increasingly important role in confirming existing or establishing often radically different insects' groupings and phylogenies. This study provides a detailed comparison of the effectiveness of the two genes COI and Cytochrome B in intra-specific study in two species of *Sitophilus* genus and their identification on barcoding. The cyt-b chows more variability at the intra-specific level and allows a better estimation of the studied parameters. COI gives better resolution when separating species based on sequence data. The barcodes sequences of the two markers also allow a good discrimination of the two species.

1. Introduction

Maize *Zea mays* is now classified as major agricultural sector in West Africa to ensure food security with an estimated extra value over €2 billion and without counting its use in the textile and pharmaceutical industry, biodegradable plastics production and biofuels [1,2]. The source of income for 85% of Senegalese depends mainly on agriculture and its derivatives particularly on maize cultivation that is practiced by 37.7% of households [3]. Yet already limited by drought and poor soil fertility, lack of varieties adapted to soil and climatic conditions and outdated agricultural equipment [4,5], production and storage of maize in West Africa are also confronted with the pest insects. Species of the *Sitophilus* genus, commonly called weevils, are among the greatest devastating maize stock. *Sitophilus zeamais* Weevils Motschulsky 1855 and *Sitophilus oryzae* Linnaeus 1763, two species very similar in appearance, are pests that usually have a wide distribution area due to unintentional introductions related to the transportation of grain they infest. The identification of the species in question and a better understanding of the dynamics of their populations are prerequisites for the development of integrated and environmentally friendly strategies of struggles. As a result, phylogeny and taxonomy of these insects increasingly interested researchers but they were generally conducted from morphological and/or anatomical criteria. Indeed, the application of these methods is confronted with the barrier of the existence of twin or cryptic species. Thus, the genetic analyzes have been recently used to contribute to the confirmation of the existence of different predefined taxonomic groups. Recent advances in the preparation, identification and sequencing of DNA, RNA and proteins, storage and analysis of large-scale data have revolutionized our understanding of biodiversity. The diversity of organisms being the result of changes in DNA sequences and environmental effects, genetic variation is considerable and every individual of a species, with the exception of monozygotic twins, has a unique DNA sequence [6]. The use of DNA sequences is therefore a better way for the identification, characterization, and following the evolution of predatory species. Particularly insects are good models of genetic studies for understanding many mechanisms that characterize the living world. In the *Sitophilus* genus, some authors [7] propose an identification of immune system genes and their expression in the bacteriome tissue. The authors in [8] have used the "RNAi (RNA interference)" technique for determining the interactions of knockdown gene system in the bacteriome tissue. Gafishi and his colleagues [9] studied the transmission of resistance in the *Sitophilus* genus and its impact in the crop yield.

Mitochondrial genes are by far the most used in these phylogeny, population genetics or phylogeography studies. This is justified by their high variability [10], their significant contribution in understanding the evolution of phytophagy mechanisms and co-evolutionary relations between host plant and parasite [11,12] and their ability to provide interesting tracks on the history of the geographic expansion of pests [13]. Today, many genetic studies are carried out not only by processing the data with different methods, but by combining data from two or more genes according to the concept of concordance between different loci [12]. COI genes and Cyt b handsets are more discriminating than nuclear genes in bar-coding because they give a strong inter-specific variability while remaining sufficiently polymorphic to intra-specific level [6]. The DNA bar-coding, used in pest controls, predatory species and Food Security [14], made the Linnaean taxonomic system more accessible by allowing rapid and automatic identification of species by using short sequences and standardized genetic markers. It is becoming more and more efficient method, essential in the identification and description of species, thus contributing to the development of taxonomy and biodiversity research [15]. The two most frequently used genetic markers are cytochrome oxidase I (COI) and cytochrome b (cyt b). Nevertheless, to the constant evolution of DNA technology, the choice of markers and their specificity based on the objectives of studies remains controversial. The question that arises is: what marker should be used for a specific genetic study? Thus, the overall objective of this study is to test the reliability of two mitochondrial markers (COI and Cyt-b) in the genetic characterization of *Sitophilus spp* species. It is specifically to compare the genetic variability of the two markers in intra and inter specific level, the choice of one or the other gene in the species distinction studies to test their ability to be used for studies of barcoding.

2. Materials and Methods

2.1. Sample collection and geographical location

Insects were collected directly from a mass rearing performed on already infested maize from the field. Samples of the two species (*S. oryzae* and *S. zeamais*) were collected in 6 regions of Senegal. The study was extended in Labe, a town belonging to the maize growing natural region of Middle Guinea. **Figure 1A** shows the location of these places and **Figure 1B**, agro-ecological zones.

The different sampling parameters and the different agro-ecological zones to which belongs each location are recorded in the **table 1**. Sampled individuals are encoded using the first capitalized letter of the genus name followed by the two first letters of the place of origin (the first capitalized letter and the second lower case) and finally a serial number. Eg SBa1 code represents a *Sitophilus* individual from the locality of Bambey and as order number 1.

2.2. Genetic study

2.2.1. Genomic DNA extraction S. oryzae and S. Zeamais

DNA extraction is performed with the Dneasy® Tissue Kit (Qiagen GmbH, Germany) according to the following extraction protocol: After dissecting insects, head, prothorax and legs are used for extraction. These parts are ground and then digested in 1.5 ml tubes supplied with the kit with 180µl of ATL buffer and 20µl of proteinase K and incubated at 55°C for 3h. Cell lysis is done with 200µl of AL Buffer. The mixture is

homogenized by vortexing and then incubated at 70°C for 10 min. 200µl of 96-100% ethanol were then added to the mixture for purification. Solutions are poured into columns previously placed in 2 ml collection tubes and centrifuged at 13,000 rpm (revolutions per minute) for 1 min. The successive addition of 500µl of AW1 buffer, followed by centrifugation for 1 min and AW2, followed by centrifugation for 3 minutes allows to fix the DNA to the silica membrane and to dry it. The DNA is then dropped to the silica membrane by adding 40µl of AE Buffer previously incubated at 70 ° C. The extraction product is stored at -20°C.
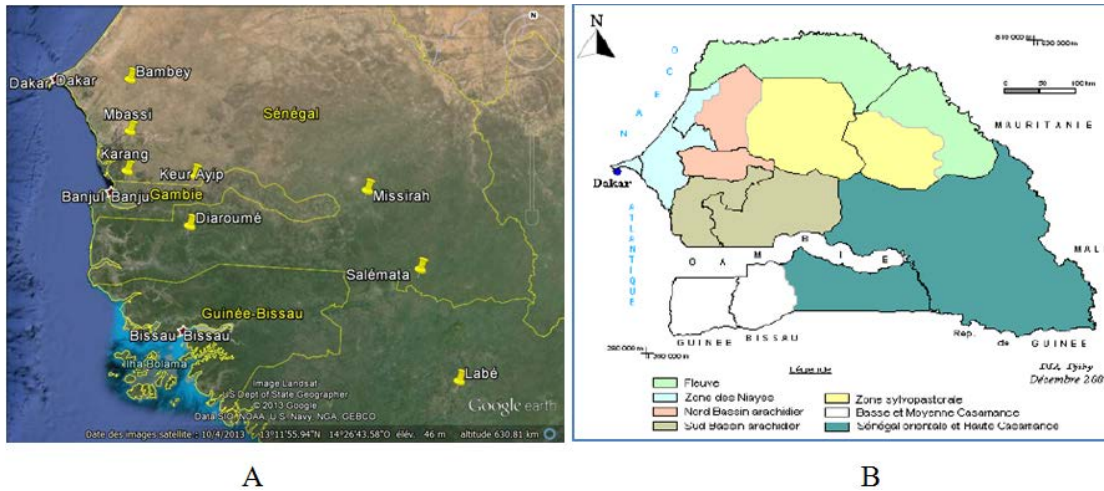

Figure 1: Location of the different localities (A) and agro-ecological sampled zones (B [1]).

Table 1: Summary of characteristics of the collected samples

Country	Regions	Localities	Agro-ecological zones	Storage Infrastructures	Codes
	Diourbel	Bambey	Nord Bassin Arachidier	Hut	SBa
	Kaolack	Keur Ayip		Stores	SKe
		Karang	Sud Bassin Arachidier	Stores	SKa
	Fatick	Mbassis		Field	SMb
Senegal	Tambacounda	Missirah	Haute Casamance /Senegal Oriental	Attic	SMi
	Kédougou	Salémata		Stores	SSa
	Sedhiou	Diaroume	Basse et Moyenne Casamance	Bed room	SDi
G. Conakry	F. Djalon	Labe	Moyenne Guinea	kitchen	SLa

2.2.2. Gene amplification or PCR

A thermo cycler of Master Cycler (Eppendorf) kind was used with the following parameters: an initial denaturation step of 3 min at 94 ° C then 35 cycles of denaturation (1 min at 94 ° C), annealing (1 min at 47 ° C) and elongation (1 min at 72 ° C), followed by a final elongation stage of 10 min at 72 ° C. For each marker, the following forward and reverse primers were used: for COI primers are C1-N2191 (5'-CCAGGTAAAATTAATATAAACTTC-3') and C1-J-1751 (5'- GGATCACCTGATATAGCATTCCC- 3') and for Cyt b the primers CB1 (5'-TAT GTA CTA CCA TGA GGA CAA ATA TC-3') and CB2 (5'-ATT ACA CCT CCT AAT TTA TTA GGA AT-3') were used. The Sequencing allowing the determination of the nucleotide sequence of a DNA fragment was carried out by Macrogen in South Korea.

2.3. Genetic analysis

In this study, a population is considered to be all the individuals of a species belonging to the same agro ecological zone or the same locality. AMOVA (molecular variance analysis) statistical tests and the Mantel test were not performed for the *S. oryzae* species because the number of sampled locations for this species (2 locations) is less than the number allowed by the software (3 minimum).

2.3.1. Cleaning and sequence alignment

Cleaning, adjustment and alignment of sequences were manually performed; using BioEdit software Version 7.2.0 [16]. Sequence alignment allows seeing sites homology and highlighting variations in the obtained sequences.

2.3.2. Polymorphism and genetic diversity

The nucleotide composition of sequences for each marker, for each species and the frequencies of substitutions were generated with MEGA Version 5.05 software. The percentages were calculated based on the total number of sites. For both markers, different indices were calculated for each population standard diversity indices with the S variable sites (singletons, informative parsimonies), the number of haplotypes (a single nucleotide sequence, which may be common to several individuals, but which differs from other haplotypes by one or more nucleotide substitutions) for each gene and for each species were identified with the software Dnasp version 5.10.01 [17].

2.3.3. Differentiation and genetic structuring

The indices of genetic differentiation: the Fst, is calculated from the software Arlequin version 3.1 [18] and genetic distances were obtained with the software MEGA version 5.05 [19]. The study of these parameters allows accounting for the structuring and the intra and inter-specific genetic differentiation.

The Mantel test was performed with the XLSTAT software [20]. The test is to see whether the geographical distances affect the genetic differentiation of studied populations. The statistical r (r (AB)) given by this test represents the correlation coefficient between two matrices A and B, the corresponding two-tailed P value is obtained after 10,000 permutations. A p-value less than the significance level alpha = 0.05 to reject the null hypothesis H0 that the matrices are uncorrelated and accept the alternative hypothesis that the matrices are correlated. Conversely, a higher p-value at the significance level used to accept the null hypothesis and reject the alternative hypothesis.

2.3.4. Differentiation and genetic structuring

A comparative analysis between the haplotype diversity Hd [21] and nucleotide Pi [22,21] allows to speculate on demographic trends of the studied populations. Such diversity indices were determined with the software Dnasp Version 5.10.01 [17]. Different statistical tests such as D Tajima, Fs Fu [22,23] and R2 Ramos generated from the Arlequin version 3.1 software [18] were used to test the difference to neutrality assumption. Fu Fs of the test compares the average number of differences taken in pairs ($\Theta\pi$) with the number of haplotypes (k) in the population. D Tajima is based on the difference between the average number of differences taken in pairs ($\Theta\pi$) and the number of polymorphic sites (S). In a population of constant size, D Tajima and Fs Fu is 0. Conversely, a demographic expansion leads to negative values of D and Fs [18]. SSD (sum of deviation squares), the index of raggedness (rg) were determined with the Arlequin version 3.1 software [18].

The demographic history of populations was also explored with the analysis of Mismatch distribution or distribution disparities (ie distribution of the differences between the sequences taken in pairs). This is to establish the distribution of the number of differences between individuals within a population, taken in pairs. The curves of Mismatch distribution [24] were performed with the Dnasp software Version 5.10.01 [17], under the assumption of constant population.

2.3.5. Phylogenetic analysis
2.3.5.1. Networks of haplotypes

To show graphically the relationship between the analyzed haplotypes were constructed haplotype networks with Network software version 4.6.1.1 using the method of the median joining. For each marker, sequences of both species were combined. The haplotype network is a pictorial support of the genetic structure of the studied populations.

2.3.5.2. Phylogenetic reconstructions

Phylogeny of the populations for both genes was investigated by constructing phylogenetic trees using different methods: Method of Neighbor-Joining (NJ), maximum parsimony (MP) and the method of maximum likelihood (ML). These trees were generated from the software MEGA Version 5.05 [19]. The validity of the trees was tested by the procedure called bootstrap [25,26]. The value of the bootstrap corresponds to each node, the number of times the one found among the different sub trees the same connection [27]. Finally, the last trees are constructed with the method of Bayesian inference from mrbayes Version 3.1.2 software. Phylogenetic trees for

this method were visualized from Fig tree version 1.3.1 software. The phylogenetic tree analysis was conducted in connection with the robustness of bootstraps or the value of posterior probabilities for Bayesian inference.

2.3.6. Analyses of DNA barcoding

The barcode data of the DNA sequences of the two markers were obtained in two ways: The haplotypes were processed in barcode sequence considering only the variable sites between the different haplotypes, the conserved sites are removed. This procedure was performed manually, using the software Version 7.2.0 Bioedit 2005 [16]. Then the phylogeny barcodes sequences were studied using the methods of phylogenetic reconstruction previously cited.

3. Results and discussion

3.1. Results

3.1.1. Polymorphism and genetic diversity

In Cyt b for *S. oryzae*, 13 sequences of 480 bp were analyzed with 386 conserved sites 80% and 17 singletons representing 3.5%. No parsimony informative site was identified. Nucleotide frequencies are 29.47% adenine (A), 34.46% for thymine (T), cytosine (C) and 21.68% Guanine (G) 14.82%. For *S. zeamais*, 34 sequences of 480 bp were analyzed with 95% conserved sites, 6 variables sites: 0.6% informative parsimonies and 0.6% singletons (**Table 2**). The nucleotide frequencies are 28.34% for A, 36.65% for T, 20.63% for C and 14.37% for G.

COI with *S. oryzae* 24 sequences of 410 bp have been analyzed with 386 (94%) conserved sites, 24 variables sites whose 19 informative parsimonies (4.6%) and 5 Singletons (1.21%). For *S. zeamais* 49 sequences of 410 bp were analyzed with 380 or 92% of conserved sites, 30 variable sites, 28 informative parsimonies site (6.8%) and 2 singletons sites (0.4%) (**Table 2**). Nucleotide frequencies are relatively close between the two species (A = 39.76%, T / U = 28.9%, C = 11.53% and G = 19.81% for *S. oryzae*; A = 39.37 %, T / U = 29.95%, C = 11.23% and G = 19.45% for *S. zeamais*).

Table 2: Variability of sequences of the two markers for both species

Markers	Species	Number of sites	Invariables sites	Variables sites	
				Singletons	Parsimonies informative
COI	*S. oryzae*	480	386→80%	17→3,5%	0
	S. zeamais	480	460→95%	3→0,6%	3→0,6%
Cyt b	*S. oryzae*	410	386→94%	5→1,2%	19→4,6%
	S. zeamais	410	380→92%	2→0,4%	28→6,8%

The COI gene has very different substitution rate values between the two species. For *S. oryzae*, mutations are mainly of transversion type (to 99.29% in total), the transitions are very low (0.39% in total) and are made only

between pyrimidine bases (CT) (bold Schedule 1). Whereas the reverse was observed for *S. zeamais*. For this species transitions are considerable (45.77%) and highest among purine bases (AG) with a rate of 45.46% against 0.31% for the transitions between pyrimidine bases.

For the cyt b, there is no significant difference in the frequency of substitution between the two species. The frequency of transition mutations is 53.35% for *S. oryzae* and 53.74% for *S. zeamais* and transversion mutations are respectively 46.64% and 45.82%. The transitions are more common between purine bases (49.83% and 48.42% for *S. oryzae* and S. *zeamais* respectively) between pyrimidine bases (3.52% and 5.32% for *S. oryzae* and *S. zeamais* respectively).

3.1.2. Genetic diversity of populations
3.1.2.1. Haplotypes of S. oryzae and S. zeamais species

For the COI, only 3 haplotypes of which 2 individual haplotypes are identified to the *S. oryzae* species on 13 sequenced individuals. The haplotype 1 is a majority with 11 individuals (84.6% of total population) distributed in the localities of Diaroumé and Salémata. *S. zeamais* has 10 haplotypes with 7 individual haplotypes. The H1 haplotype is a majority with 21 individuals (61.7% of the total population) distributed throughout the sampled localities.

The analysis of cyt b sequences helped to define 16 haplotypes out of the 24 sequenced individuals for the *S. oryzae* species including 14 individual haplotypes (Annex). The haplotype 2 prevails with 8 individuals (33% of the total population) distributed in the two sampled localities for this species (Diaroumé and Salémata). For *S. zeamais*, 18 haplotypes were defined out of the 49 sampled individuals including 11 individual haplotypes. For this case, the haplotype 2 is a majority with 22 individuals (44% of the total population) met on all sampled localities.

3.1.2.2. Nucleotide and haplotype diversity

Haplotype and nucleotide diversities calculated for both species were higher with the marker cyt b (**Figure 2**). A comparison of haplotype diversity of both species for both markers shows that with the cyt b haplotype diversity is higher for *S. oryzae*. As against the COI, haplotype diversity is higher for the *S. zeamais* species.
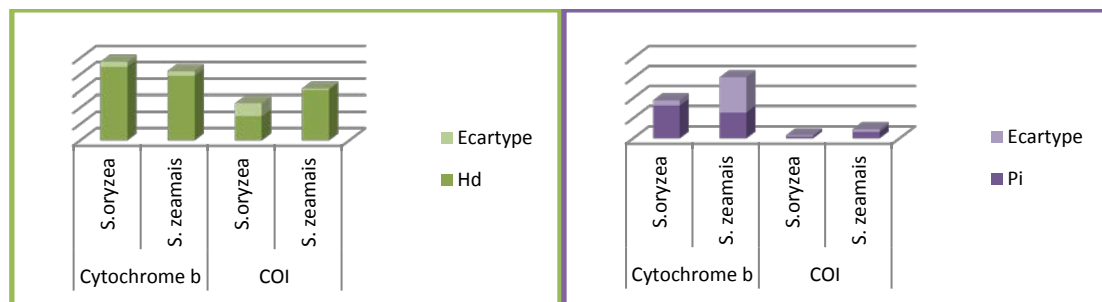

Figure 2: Histograms of haplotype diversity (A) and nucleotide (B) COI and Cyt b

3.1.3. Differentiation and genetic structuring

3.1.3.1. Genetic distances

At the inter-specific level, in absolute value, the calculated genetic distance is higher for the COI (0.165) than cyt b (0.12). At the intra-specific level, genetic distances between haplotypes obtained with the cyt b (*S. oryzae*: 0.017 and *S. zeamais*: 0.013) are much higher than those of the COI always expressed very low genetic distances (*S. oryzae*: 0.001, *S . zeamais*: 0.003).

3.1.3.1.1. Intra agro-ecological zones' genetic distances

Agro-ecological zones at the intra-level, genetic distances calculated with the COI are very low for both species, however, a greater distance (0.018) is obtained for the Northern Groundnut Basin population NBA (H2: SBa4) of *S. zeamais*. The shortest intra-zone distance (0.001) is obtained at HCSO as well for *S. zeamais* as for *S. oryzae*. The cyt b shows agro ecological zones' intra-genetic distance values much higher than those distances calculated with the COI. *S. zeamais* genetic distances vary between 0.007 in MG and 0.015 in SBA while they are between 0.015 and 0.019 for *S. oryzae*.

3.1.3.1.2. Genetic distances inter agro-ecological zones

For both species, genetic distances between agro-ecological zones are significantly higher with the cyt b. For this marker, *S. oryzae* has a distance of 0.017 between BMG areas and HCSO, while between these two areas the distance calculated with the COI is 0.001. Also, for *S. zeamais*, a maximum value (0.016) was obtained between the NBA and SBA areas and the lowest distance value (0.010) was obtained between the NBA and MG for cyt b. The distances calculated with the COI are even lower with a maximum of 0.010 between the NBA and the remaining localities.

3.1.3.2. Genetic differentiation (Fst)

3.1.3.2.1. Inter-specific genetic differentiation

The indices of differentiation based on all sequences show a high genetic differentiation between the two species (Fst = 0.98; P-value = 0 with the COI and Fst = 0.90; P-value = 0 with cyt b). The observed differentiation is however higher with the COI.

3.1.3.2.2. Genetic differentiation between agro-ecological zone

With the COI, the *S. oryzae* species shows no differentiation between the two agro-ecological zones (BMC and HCSO) where it was sampled. For *S. zeamais*, only haplotypes NBA and SBA areas (Fst = 0.39) showed significant differentiation.

With the Cyt b, *S. oryzae* does also admit no genetic differentiation between the two agro-ecological zones. As for *S. zeamais* only the NBA and MG zones (Fst = 0.11), which have significant genetic differentiation (**Table 3**).

Table 3: Fst inter agro ecological zones (significant values are indicated by a star).

Species	Agro-ecological zones		Markers	
			COI	Cyt b
S. oryzae	BMC	HCSO	0,01615	0.01067
S. zeamais	NBA	SBA	0.39509*	0.11525
	NBA	HCSO	0.2619	-0.02787
	NBA	MG	0.21844	0.11326*
	SBA	HCSO	0.01300	-0.00398
	SBA	MG	-0.04925	0.09573
	MG	HCSO	-0.05222	0.04911

3.1.4. Mantel test

The test shows for the COI marker a correlation coefficient between 0 and -1 (r = -0.368) with a p-value (p-value <0.0001) lower than alpha. The null hypothesis H0 is rejected and the alternative hypothesis Ha the matrices that are correlated is accepted. The risk of rejecting H0 when it is true is less than 0.01%. Therefore, there's a correlation between genetic differentiation and geographic distance.

To the cyt b, the correlation coefficient r is between 0 and 1 (r = 0.192) with a corresponding p-value greater than alpha (p-value = 0.5909). The null hypothesis H0 cannot be rejected. The risk of rejecting H0 when it is true is 59.09%. Therefore, with the cyt b marker, genetic differentiation is not correlated to the geographic distance.
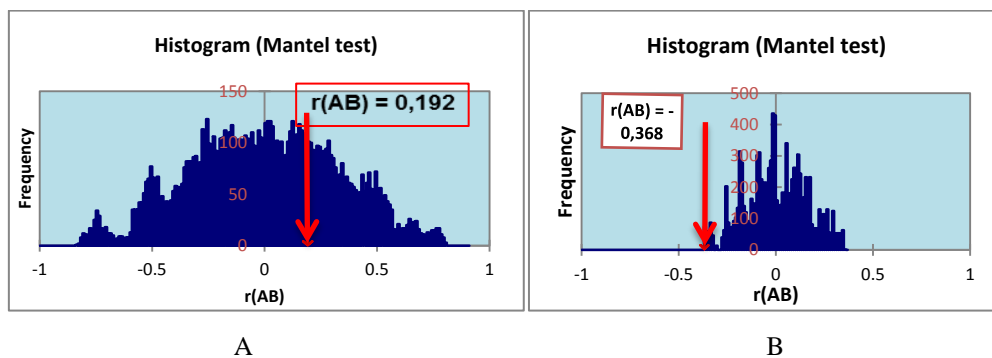

Figure 3: Histograms of Mantel tests for the *S. zeamais* species with Cyt b gene (A) and COI(B).

Graphically, the correlation coefficient is from the center (**Figure 3A**) with cyt b marker confirming the lack of correlation between the geographical distance and the genetic differentiation, while it is close with the COI (**Figure 3B**) always for this marker indicating a correlation between genetic differentiation and geographic distance.

3.1.5. Analysis of molecular variance (AMOVA)

The AMOVA test for the COI shows no significant differentiation niether between communities nor between haplotypes. Only agro ecological zones have a positive (18% variation) but which is still not significant (**Table 4**).

For the cyt b, there is no differentiation between agro ecological zones. A significant percentage of variation of 20.7% is estimated between populations within each agro-ecological zone (localities). The indicated variation within populations is significantly higher (89%) (**Table 4**), this variation is the percentage of genetic difference between haplotypes.

Table 4: Analysis of molecular variance of the populations of *S. zeamais* with COI and cyt-b markers (significant values are indicated by a star)

Source of variation	Sum of squares		Components of Variance		Percentage of variation	
	COI	Cyt b	COI	Cyt b	COI	Cyt b
Between agro-ecological zones	4.971	2.060	0.15798	-0.03956 Va	18.73256	-9.90845
Intra agro-ecological zone	1.460	1.206	-0.05689	0.08265 Vb	-6.74563	20.70097*
Intra populations	20.040	15.672	0.74224	0.35619 Vc	88.01307	89.20747*

3.1.6. Demographic evolution of populations

With the COI, polymorphism analysis revealed high values of haplotype diversity (Hd = 0.295 for *S. oryzae* and Hd = 0.613 for *S. zeamais*, Hd> 0.05) and low nucleotide diversity values (Pi = 0.00096 for *S. oryzae* and Pi = 0.00323 for *S. zeamais*; P <0.005) (Annex). These results are supported by the significant negative values of Tajima statistical tests D (D = -1.65231, P-value = 0.034 for *S. oryzae* and D = -1.65231, P-value = 0.002 for *S. zeamais*) and Fs Fu (Fs = -0.68877, P-value = 0.107 for *S. oryzae* and Fs = -0.68877, P-value = 0.013 for *S. zeamais*), significant values are indicated by stars (**Table 5**). SSDs determined with this marker as well as indices of raggedness (rg) are not significant. Ramos R2 are also significant for both species (P-value = 0).

The observed curves of mismatch distribution (**Figure 4**) are almost superimposed to the expected curves whereas the latter are built on the assumption of an expanding population.

Table 5: Demographic evolution indices for both species with both markers (significant values are indicated by a star).

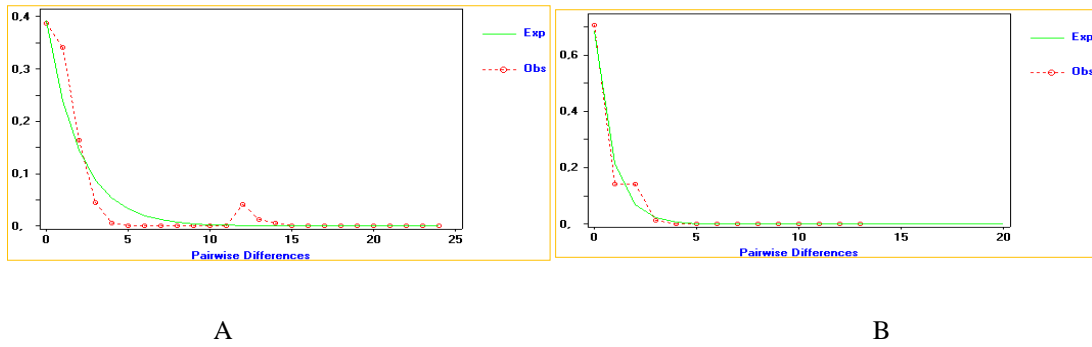| Markers	Species	Demographic Parameters				
		D of Tajima	Fs of Fu	SSD	rg	Ramos R2
COI	*S. oryzae*	-1.65231*	-0.68877	0.01085	0.33481	0,16041*
	S.zeamais	-2.32321*	-4.02737*	0.00250	0.05169	0,16185*
Cyt b	*S. oryzae*	0.15499	-4.26955*	0.01418	0.04199	0,16129*
	S.zeamais	-0.63798	-3.07178	0.70024*	0.04681	0,16343*

Figure 4: Curves of Mismatch distribution of the COI in *Sitophilus zeamais* (A) and *Sitophilus oryzae* (B).

With the cyt b, both species have high values of Hd (Hd = Hd = 0.895 and 0.782 respectively for *S. oryzae* and *S. zeamais*) and Pi (Pi = 0.016 for *S. oryzae* and P = 0.012 for *S. zeamais*). Tajima's D calculated are not significant (R = 0.15499, P-value = 0.621 for *S. oryzae*, D = -0.63798, P-value = 0.298 for *S. zeamais*. Unlike, SSDs are significantly positive. The Fs Fu are significant for *S. oryzae* (Fs = -4.26955, P-value = 0.04300). The curves of mismatch distribution for both species (**Figure 5**) have multimodal style. Ramos R2 is also significant.
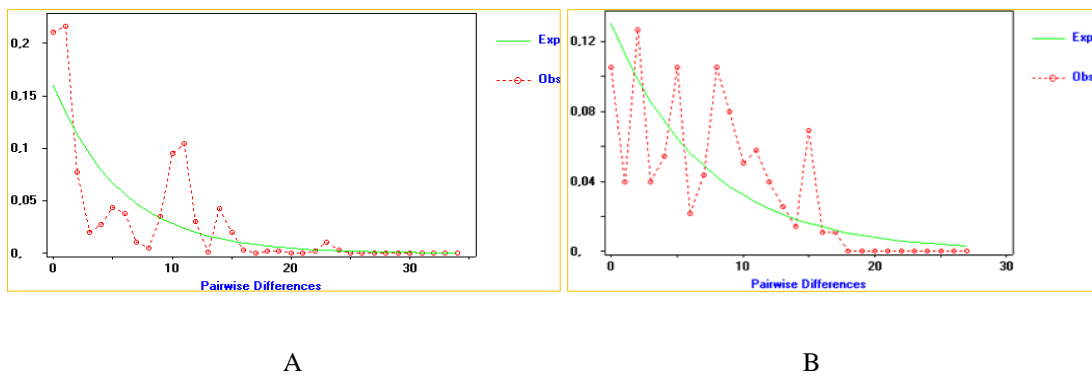

Figure 5: Curves of Mismatch distribution of the COI in *Sitophilus zeamais* (A) and *Sitophilus oryzae* (B).

3.1.7. Phylogeny of the S. oryzae and S. zeamais species
3.1.7.1. Networks of haplotypes

For *S. zeamais*, there was obtained with the COI (**Figure 6**) a representation in form of a star in which it appears a majority frequency haplotype including other derived haplotypes and it is separated by one or two mutational steps. Only the H2 haplotype of the locality of Bambey, represented in the network by the individual SBa4 has high differentiation. *S. oryzae* presents for this marker 3 haplotypes separated by one or two mutational steps.

In contrast with the cyt b (**Figure 7**) the mutational steps between haplotypes are on average high for both species, but especially with *S. oryzae*. However, the haplotypes of the two species are separated by the haplotype networks of the two markers. The distribution found in the networks did not show a clear clustering of haplotypes in relation to agro-ecological zones.

Figure 6: Haplotypes network of populations of *S. oryzae* (green and blue) and *S. zeamais* for the COI marker. Legend: agro-ecological zones are represented in color and proportion for each haplotype. The used color code is as follows: 🟢 **BMC**   🟡 **NBA**   🔴 **SBA**   🔵 **HCSO**   ⚫ **MG**
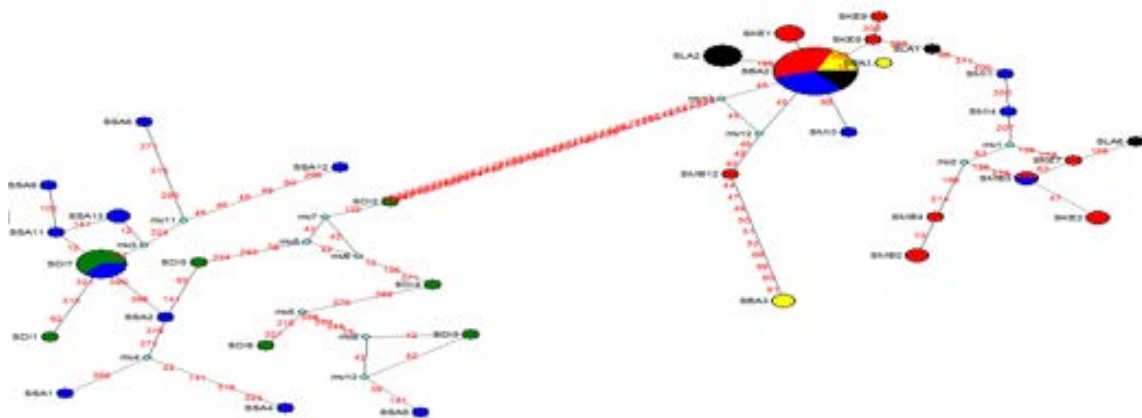

Figure 7: Haplotypes network of populations of *S. oryzae* (green and blue) and *S. zeamais* for the marker cyt b.

3.1.7.2. Phylogenetic trees

With the COI (**Figure 8A**) haplotypes of both species (red for *S. zeamais* and blue for *S. oryzae*) are separated by the phylogenetic tree. They are grouped into two clades supported by posterior probability values equal to 1. For this marker No particular grouping is noted between haplotypes except haplotypes H8 and H13 of *S. zeamais* species that form a haplogroup supported by posterior probability of 0.89. These haplotypes correspond to haplotypes H5 shown in the haplotype network by the individual SLa4 from Labe and H6 represented by the individual SMi3 from Missirah. These two locations are in different agro-ecological zones. The haplotypes H5 of the *S. zeamais* species differentiates itself. It corresponds to the H2 haplotype (SBa4) which is an individual haplotype of the locality of Bambey represented by the individual SBa4 on the haplotype network. Haplotypes from both species are also clearly separated by the tree of cyt b (**Figure 8B**). They fall into two clades (clade of the *S. oryzae* species red and that of *S. zeamais* in blue) supported by posterior probability values equal to 1. No

particular grouping was found between haplotypes for *S. oryzae*. The population of *S. zeamais* is better structured with a group into sub-clade, in group and haplogroup. H3 haplotypes (localities of Mbassis and Missirah) and H12 (the locality of Mbassis) of the *S. zeamais* species are differentiate from the group. This differentiation emerged by the COI marker at the level of network haplotype and the level of the phylogenetic tree. Haplotypes H8, H4 and H7 of the *S. zeamais* species also form a haplogroup. They all belong to the locality of Keur ayip, a locality of the SBA.
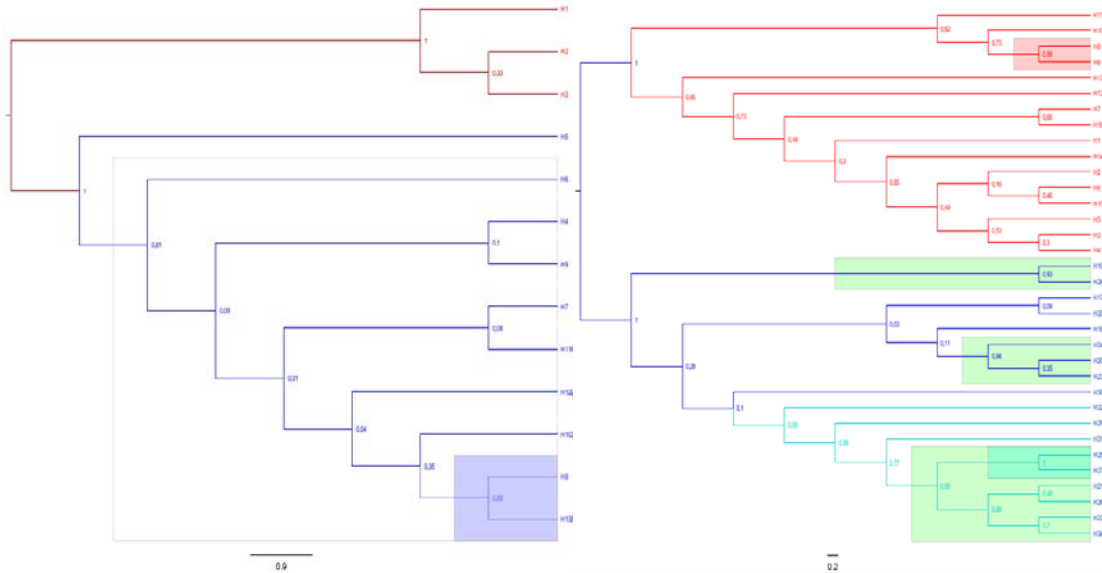

Figure 8: Phylogenetic tree of *S. oryzae*'s haplotypes (in red) and *S. zeamais* (blue) for the COI (A) and Cyt b (B) via the method of Bayesian inference.

3.1.8. Analyses of DNA barcoding
3.1.8.1. Barcode sequences

Barcode sequences of the two species for both markers were obtained by transforming the sequences with the software BioEdit, only the variable sites between species have been considered, the homologous sites were deleted. Barcode sequences of 89 bp for COI and 88 bp for the cyt b were finally obtained.

3.1.8.2. Phylogeny of barcode sequences of the two species

Maximum parsimony trees and maximum likelihood are not resolute for barcode sequences of COI. However, the method of the Bayesian inference allows for this marker optimal separating of the two species with clusters supported by the maximum posterior probability values equal to 1 (**Figure 9A**). This last method allows to separate the sequences of two species with the cyt b marker (**Figure 9B**). However, it is noted that the posterior probability values are not very high with the cyt b.
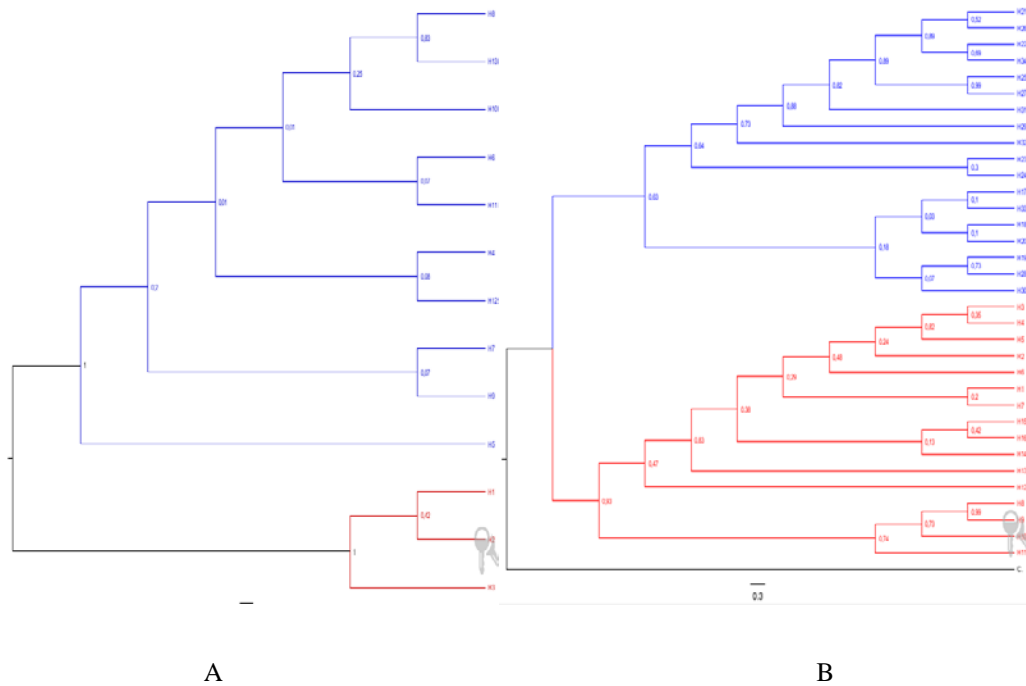
A	B

Figure 9: Phylogenetic tree barcode sequences of *S. oryzae* (red) and *S. zeamais* (blue) for the COI (A) and Cyt b (B) via the method of Bayesian inference.

3.2. Discussion

This study aims to test the reliability of the two mitochondrial markers (COI and Cyt b) in the study of genetic variability at intra specific level, defining the choice of one or the other gene in the studies of species' distinction and their ability to be used for studies of bar coding.

The estimated parameters for the variability in all DNA sequences show that the Cyt b is more variable, when sequences of the same species are considered. The variable sites are higher for the aligned sequences of the Cyt b marker. In contrast, between species the COI expresses more difference in the variability of sequences. The estimated substitution rate for this marker for example showed that mutations are made with opposite trends between the two species, while the Cyt b shows no significant difference at this level.

The various diversity indices used in our study show greater genetic diversity within populations with the Cyt b marker. This allows a better evaluation of study parameters, both between haplotype and between agro ecological zones. The estimated genetic diversity with the Cyt b is characterized by high values of haplotype and nucleotide diversity for both species. This marker could have a particularly high mutation rate for these species.

The used indices of differentiation show, with genetic distances from intra specific level, the estimated values for Cyt b are always higher than those obtained for the COI. This difference is expressed with the benefit calculation of genetic distances in intra- and inter- agro ecological zone with ever higher values for the Cyt b marker. Fst between agro-ecological zones are better estimated with the Cyt b compared to the COI for, it shows

no significant population differentiation. The analysis of molecular variance (AMOVA) also confirms this fact. Indeed, this statistic showed genetic differentiation of populations better reflected by the Cyt b marker. At the population level, for Cyt b, the essential differentiation is made between haplotype. Between localities, genetic differentiation is significant even though the estimated percentage at this level is very much lower than that of the differentiation between haplotypes. No differentiation is observed between agro ecological zones.

Mantel test's results for their part show opposite trends between the two markers. Indeed, for the COI, geographic distance affects the genetic differentiation whereas with the Cyt b there is no correlation between these two parameters. Therefore, there's no concordance between the two loci.

The more variability in the cyt b is illustrated by the haplotype networks where, for the cyt b marker; haplotypes for each species are separated by mutational steps particularly high with *S. oryzae* that presents more differentiation between haplotype. It must be remembered that Seck [28] reported that *S. oryzae* has a predilection for small grains such as rice, while according to [3], significant variations in the availability of this cereal up to good or bad availability depending on the periods, are recorded in localities where *S. oryzae* was sampled. Thereby forcing individuals of this species infest other cereals and particularly corn. This diversification of food in the same localities could cause genetic variations in populations of this species in the image of biotypes populations of *C. serratus* explained by their polyphagia according to [29].

Inversely, the COI marker has low genetic diversity of populations of *S. oryzae* with only 3 haplotypes. The calculated genetic distances with this marker for this species are low and the Fst are not significant. This shows that the study parameters measured at intra specific level with the COI marker are hardly correlated with elucidated characteristics in these localities. These results contrast with studies of [6] which suggest that despite its low variability, COI remains sufficiently polymorphic to intra-specific level to study the variability of a species. [30] also demonstrated that the Cyt b shows more variability at intra specific level and the results are better correlated with data on the studies of intra specific variation of certain characteristics (anatomy, taxonomy, ecology, ethology, food, etc.) in several species of rodents, but it is nonetheless the data of the COI show some degree of correlation with these characteristics. This controversy, however, could be lifted if one considers the number of sampled individuals in the case of our study is negligible compared to previous cited studies. This can be supported by the analysis of the results of AMOVA test that shows trends for the COI to the differentiation between agro-ecological zone and between haplotype.

We can see that at intra-specific level, the COI has low variability. However, between species, estimated values for the set of indices are always higher with this marker. This confirms the studies of [30] that, in the mitochondrial genome, sequences of certain genes are unique in expressing low variability at intra specific level but show enough inter specific variation to distinguish species, estimate their degree of differentiation or divergence duration via molecular clock calibrations.

In the estimation of demographic change for the Cyt b, the nucleotide and haplotype diversities are high. A great diversity at both nucleotidic and haplotypic, may correspond to the existence of well-differentiated lineages coexisting after a secondary contact between separated populations for a long time, or populations of large

effective size and stable for a long period of their evolutionary history until today [31]. This finding, confirmed by the analysis of the curves of mismatch distribution which presents for both species multimodal gaits, is not in disagreement with the haplotype networks that do for this marker also having no particular structure show populations in balance. However, Ramos R2's more sensitive for small populations are significantly positive indicating unlike other indices of expanding populations, insignificant D Tajima does not allow to infer on the demographic population trends. The above described demographic trends could be questioned.

The COI marker indicates a demographic expansion of populations in almost all the estimated parameters (nucleotide diversity (Hd) and haplotype (Pi), D Tajima and Fu's Fs). Ramos R2 more sensitive for small populations are significantly positive, showing thereby expanding populations. The structure of haplotype networks of *S. zeamais* for the COI, star-shaped is also well correlated with the trend of population growth.

So, there is a mismatch between loci in the estimation of the demographic evolution of the total population of both species, the two markers showing different evolutionary scenarios.

However, the lack of observed congruence between different demographic parameters determined with the Cyt b might suggest that this marker does not allow a better estimation of the demographic evolution of the total population of the studied species. Moreover, during the 2000s, the Senegalese government has launched a large crop campaign for maize. This pursuant to a considerable growth in the agricultural yield for this cereal which could contribute to an expansion of the infestation, more in line with data from the COI, inside the country and sub-regional level due to the increase in export flows.

The found distribution in the haplotypes networks for both markers did not show a clear grouping of individuals in relation to agro-ecological zones. Indeed, these representations unveiled for both species the presence of many unique haplotype variants, as well as common haplotypes and present in all or almost all agro-ecological zones. The star-shaped network structure for the COI still supports the idea of an expanding population of these pest insects.

Phylogenetic trees allowed seeing the phylogenetic relationships among the different studied populations and also the relationships between the different haplotypes of each species. It was clearly assumed that the populations of both species are better structured with Cyt b marker. For example, the haplotypes of *S. oryzae* showed no particular grouping with the tree of the Cyt b. Indeed, the populations of this species do not exhibit differentiation on all indices estimated with this marker. They form in fact a single unique population and recalling Ndiaye's studies [32]. Conversely, the phylogeny of haplotypes of the COI does not comment on the phylogenetic relationships between different haplotypes. This fact justifies the choice of the Cyt b in many genetic studies of insects.

Phylogenetic reconstruction of barcodes sequences of the two species for both markers helped distinguish them and also to see the susceptibility of these sequences for phylogenetic studies at the intra specific level.

With the COI, it was found that the trees of the methods of maximum parsimony and maximum likelihood are not limiting. Barcode sequences of the two species are better distinguished using the method of the Bayesian

inference in which they are grouped into two clades supported by the posterior probability values maximum equal to 1. However, the maximum parsimony methods and maximum likelihood for the Cyt b are more limiting. Therefore, the observed lack of congruence between the different reconstruction methods used for phylogenetic analyzes of barcode sequences of the two species, hence the importance of markers' combined studies.

Indeed, barcodes sequences were used in many phylogenetic studies to distinguish species. Most of them, the COI was the preferred marker. According to [33], this choice is related to its high variability in inter-specific level and its low variability in intra-specific one, which can greatly minimize the risk of introducing bias in the identification of individuals. The identification of the two species of the Sitophilus genus is no exception to the rule. As for Cyt b, as suggested [6], its high intra specific variability can be a handicap for technical analysis of bar coding. However in the case of our study, we found that with the Cyt b marker, the two species are distinguished with barcodes sequences. Also, for the same species, the topology of the phylogenetic tree of barcode sequences is similar to that of the tree of haplotypes indicating a good susceptibility of barcodes sequences of this marker to trace the phylogeny at the intra specific level. This however does not mean that the sequences of the COI barcodes cannot be used for phylogenetic studies at the intra specific level. As previously noted, the sample size is crucial for genetic studies, especially when using markers such as the COI, which is a little variable at the intra specific level.

4. Conclusion

The marker Cytochrome b shows greater genetic variability that the COI within species. Therefore it makes it more the diversity and genetic structure of populations at intraspecific level. However, the COI data are better correlated with results of previous studies and this marker appears to have more reliability.

Although, both markers have good discriminatory power of the two species. The highest variability of the COI, in interspecies, gives it an advantage in identification. Barcodes sequences allowed separating the two species but also, as has been found with the Cyt b, they recount well the phylogeny at the intraspecific level.

Acknowledgements

References

[1] C. A. Fall (2008). Etat des ressources phytogénétiques pour l'alimentation et l'agriculture dans le monde, Second Rapport National du Sénégal. ISRA, Ministère de l'agriculture, République du Sénégal.

[2] S. Guèye. "Activité insecticide des extraits et Huiles essentielles de *Lantana camara* L. et d'*Annona senegalensis* Pers. contre *Caryedon serratus* Ol., ravageur des stocks d'arachide ". Thèse de 3ème cycle, Université Cheikh Anta DIOP, Dakar, pp.123, 2008.

[3] ESASU. "Senegal: Enquiry about Food security, in Emergency situation, in Rural area". pp.15, 2008.

[4] A. Boyé. "Situation des ressources génétiques forestières du Sénégal". Atelier sousregional FAO/IPGRI/CIRAF sur la conservation, la gestion, l'utilisation durable et la mise en valeur des ressources génétiques forestières de la zone sahélienne, Ouagadougou, 22-24 sept. 1998, 2001.

[5] PANA. "Plan d'action nationale pour l'adaptation aux changements climatiques. Ministere de l'environnement et de la protection de la nature" – République du Sénégal. pp.84, 2006.

[6] P. Borsa, S. Lemer, W.J. Chen, A. Collet, A.Carassou, D. Ponton. "Identification par barcode (séquences nucléotidiques d'un fragment du gène ducytochrome *b*) des larves de Lethrinidae capturées en baies de Dumbéa et Ouano (Nouvelle-Calédonie)". Programme d'évaluation des ressources marines de la zone économique de Nouvelle-Calédonie, IRD, UR 128, Nouméa, Nouvelle-Calédonie, 2009.

[7] C. Anselme, V. Pérez-Brocal, A. Vallier, C. Vincent-Monegat, D. Charif, A. Latorre, A. Moya, A. Heddi. "Identification of the Weevil immune genes and their expression in the bacteriome tissue". *BMC Biology* : 10.1186/1741-7007-6-43, 2008.

[8] A. Vallier, C. Vincent-Monégat, A. Laurençon, A. Heddi. "RNAi in the cereal weevil *Sitophilus* spp: Systemic gene knockdownin the bacteriome tissue". *BMC Biotechnology*, pp.1472-6750, 2009.

[9] M.K. Gafishi, J.K. Tumutegyereize, G. Asea. "Inheritance of resistance to maize weevil *Sitophilus zeamais* (Motschulsky) in maize inbred lines and determination of their heterotic groups". *Second RUFORUM Biennial Meeting 20 - 24, Entebbe, Uganda,* 2010.

[10] W.J.O. Ballard, M.C. Whitlock. "The incomplete natural history of mitochondria". *Molecular Ecology,* vol.13, pp.729–744, 2004.

[11] D. Detwiler, J. Janovy. "The role of phylogeny and ecology in experimental host specificity : insights from a eugregarine-host system". *J. Parasitol.,* vol.94(1), pp.7‑12, 2008.

[12] J.C. Avise. "Phylogeography: retrospect and prospect". *Journal of Biogeography,* vol.36, pp.3–15, 2009.

[13] D. Muths, E. Tessier, J. Bourjea. "Contribution de la génétique des populations à l'étude de la connectivité régionale de certaines espèces de poissons commerciaux présents dans la Réserve Naturelle Marine de La Réunion : étude de faisabilité". Parc Marin, Réunion. Pp.23, 2008.

[14] P.D.N. Hebert, T.R. Gregory. "The Promise of DNA Barcoding for Taxonomy". Department of Integrative Biology, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, N1G 2W1, Canada. *Systematic Biology.* Vol.54(5), pp.852–859, 2005.

[15] J. Pons, G. Timothy, T.G. Barraclough, J. Gomez-Zurita, A. Cardoso, D.P. Duran, S. Hazell, P. Kamoun, W.D. Sumlin, A.P. Vogler. "Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects". *Syst. Biol.* Vol.55(4), pp.595–609, 2006.

[16] T.A. Hall. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". *Nucl. Acids. Symp. Ser.* Vol.41, pp.95-98, 1999.

[17] J. Rozas, P. Librado, J.C. Sánchez-Del Barrio, X.Messeguer, R. Rozas. "DnaSP Version 5 Help Contents [Help File]", 2010. Available with the program at http://www.ub.edu/dnasp/.

[18] L. Excoffier, G. Laval, S. Schneider. "Arlequin ver. 3.0: An integrated software package for population genetics data analysis". *Evolutionary Bioinformatics Online*. Vol.1, pp.47–50, 2005.

[19] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M.K.S. Nei. "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods". *Mol. Biol. Evol.* Vol.28 (10), pp.2731-2739, 2011.

[20] "Addinsoft. XLSTAT software", Paris, France, 2012.

[21] M. Nei. "Molecular Evolutionary Genetics". Columbia University Press, New York, NY, USA, 1987.

[22] F. Tajima. "Evolutionary relationship of DNA sequences in finite populations". *Genetics*, vol.105, pp.437-460, 1983.

 [23] Y.X. Fu. "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection". *Genetics*, vol.147, pp.915-925, 1997.

[24] A. R. Rogers, H. Harpending. "Population growth makes waves in the distribution of pairwise genetic differences". *Molecular Biology and Evolution*, vol.9, pp.552-569, 1992.

[25] B. Efron. "The jackknife, the bootstrap and other resampling plans. Standford university, department of statistics". *National Science Foundation Grand MCS* 77-16974, 1982.

[26] J. Felsenstein. "Confidence limits on phylogenies: an approach using the bootstrap". *Evolution,* vol.39, pp.783–791, 1985.

[27] M. Harry. "Génétique moléculaire et évolution". *Collection sciences fondamentale*. pp.326, 2001.

[28] D. Seck. "Stockage, conservation et transformation des récoltes en zones tropicales.Notes de cours", pp.31, 2009.

[29] M. Sembène, J.Y. Rasplus, J.F. Silvain, A. Delobel, "Genetic differentiation in sympatric populations of the groundnut seed beetle *Caryedon serratus* (Coleoptera: Chrysomelidae): new insights from molecular and ecological data". *Int. J. Trop. Insect Sci.,* vol.28 (3), pp.168-177, 2008.

[30] S.S. Tobe, A.C. Kitchener, A.M.T. Linacre. "Reconstructing Mammalian Phylogenies: A Detailed Comparison of the Cytochrome b and Cytochrome Oxidase Subunit I Mitochondrial Genes". *PLoS ONE,* vol.5(11): e14156. doi:10.1371/journal.pone.0014156, 2010.

[31] W.S. Grant, B.W. Bowen. "Shallow populatiion histories in deep evolutionary lineages of marine fishes: Insight from sardines and anchovies and lessons". *The Journal of Heredity*: vol.89 (5), 1998.

[32] M.R. N'diaye "Impact des zones agro-écologiques et agro-climatiques sénégalaises et Guinéennes sur la Diversité et la structuration génétiques de *Sitophilus* spp (coléoptèra, Curculionidae) ravageurs du maïs stocké". Mémoire de Master II. Université Cheikhe Anta Diop de Dakar, pp.50, 2012.

[33] K. Bala, G.P. Robideau, A.W.A.M. Cock, N. Desaulniers, C.A. Lévesque. "DNA Barcodes for the discovery and identification of new species of Pythium". *Agriculture and Agri-Food,* Canada. Carleton University, 2009.