Molecular characterization of four (4) accessions Sesame (*Sesamum indicum L.*) cultivated in different localities of Central Africa and the Pakréssaya variety (of Burkinabe origin).

ABSTRACT

Sesame (Sesamum indicum L) is probably the oldest oilseed crop known and used by humans. In the Central African Republic, sesame remains an "orphan crop" due to the lack of scientific data on the genetic diversity of local accessions. Genetic diversity is the extent of genetic variability measured at the level of individuals, a population, a metapopulation, a species or a group of species. It is ensured by the genetic variability between individuals within the species. It is in this sense that this present study is included, which tends to contribute to a better knowledge of the genetic characteristics of sesame accession in the Central African Republic. More specifically, it is a question of evaluating the genetic diversity of sesame according to agroecological zones; of determining the existence of a genetic structuring of sesame according to agroecological zones, of tracing the phylogeny of sesame in the Central African Republic. Total DNA was extracted using the Zymo Research Kit, taking for each accession or variety a pool of five seeds constituting an individual. The different analyses carried out are: analysis of genomic sequences, genetic polymorphism (with BioEdit software version 7.0.5.3), differentiation and genetic structuring (with Arlequin software version 3.5.2.2. Software), demographic evolution (DnaSP software version 5.10.01), construction of phylogenetic trees (MEGA software version 7.0.14, Mr Bayes version 3.2.6). The results obtained showed the effectiveness of the 28S gene for the differentiation of sesame accessions (Sesamum indicum) because it showed significant polymorphism between the different accessions studied and the Pakréssaya variety.

Keywords: Sesame, Pakréssaya variety, Central African accession, 28 S gene, Molecular characterization.

Commented [TP1]: *"It is in this sense that this present study is included..."*
➞ Awkward phrasing. Replace with something clearer, like, "This study aims to contribute to..."
"The different analyses carried out are: analysis of genomic sequences..."
➞ Better phrased as: "The analyses included genomic sequence analysis, genetic polymorphism..."
The abstract introduces a wide range of analyses without connecting them cohesively to the research objectives. For instance:
- Explicitly state how each analysis (e.g., phylogenetic tree construction) contributes to understanding genetic diversity.

The last sentence could summarize the broader significance of the findings.
Phrases like "genetic diversity is ensured by the genetic variability" are repetitive and can be streamlined.

Introduction

Sesame (*Sesamum indicum* L) is probably the oldest oil crop known and used by man. **[1]** . This plant belongs to the Pedaliaceae family and is important because of the resistance of its oil

Commented [TP2]: The historical and nutritional significance of sesame is mentioned repeatedly. Consider cond

Commented [TP3]: *"Sesame (Sesamum indicum L) is probably the oldest oil crop known and used by man."*
➞ Consider replacing "by man" with "by humans" for inclusivity.

to oxidation and rancidity. Its demand on the world market is increasingly increasing due to its multiple technological and food interests. In food, the oil is used in cooking, the seeds are used in the production of dough **[2]** . In industry, sesame is used for the manufacture of products such as perfumes, skin cosmetics, hair oils and soaps **[3]** . Native to Africa and India, it is cultivated for its seeds, which are rich in oil, vitamins and proteins **[4]** . In Africa, there are 17 species of Sesamum , two of which also grow in the wild. The species *S. indicum* is also known by the synonym *Sesamum oriental* L. Other lesser-known synonyms of *S. indicum can also be cited.* such as *S. edule* , *S. luteum* , *S. oleiferum* , *S. africanum* , *S. foetidum* . The genus *Sesamum* includes a large number of varieties that differ in size, shape, growth habit, flower color, seed size, color, and composition **[5]** . Some plants have a genetic potential that allows them to adapt to environmental conditions **[6]** .

Commented [TP4]: *"Its demand on the world market is increasingly increasing..."*
➞ Replace with "Its global demand is steadily rising..." to avoid redundancy.

However, in Central African Republic **,** sesame remains an "orphan crop" due to the lack of scientific data on the genetic diversity of local accessions . Genetic diversity is the extent of genetic variability measured at the scale of individuals, a population, a metapopulation, a species or a group of species. It is ensured by the genetic variability between individuals within the species. It expresses the property that organisms have of acquiring new characteristics through mutations and the effects of natural selection. Thanks to this variability and the limits of the species, individuals differ from each other for one or more characteristics **[7]** . During their evolution, cultivated plants acquire biological characteristics allowing them to adapt to new environments. All of these biological characteristics shaped by evolutionary processes have generated new characteristics constituting genetic diversity within the species. Genetic diversity of crops plays an important role in sustainable development and food security, as it allows the cultivation of plants in the presence of various biotic and abiotic stresses **[8]** . It is in this sense that this present study is included, which tends to contribute to a better knowledge of the genetic characteristics of sesame accession in Central Africa. More specifically, it is about :

Commented [TP5]: "Genetic diversity is the extent of genetic variability measured at the scale of individuals, a population, a metapopulation, a species or a group of species."

- Assess the genetic diversity of sesame according to agroecological zones;
- Determine the existence of a genetic structuring of sesame according to agroecological zones;
- Tracing the phylogeny of sesame in Central Africa

3.1 Materials and methods

3.1.1 Sample collection

This study was carried out on samples of sesame seeds (*Sesamum indicum)* collected in different geographical areas of Central Africa. The seeds used were obtained from plants at the fully mature stage. They come from two distinct categories of sesame populations: four local accessions of Central African origin and one variety (Pakréssaya) of Burkinabe origin. The accessions come from four different localities (Bambari, Bossembélé, Kemo, Kouango) of Central Africa. The variety Pakréssaya of Burkinabe origin was obtained from the Central African Institute for Agronomic Research (ICRA) via the Bakéré agricultural research station based in Bossembélé in Ombella M'poko. The samples were then sent to the genomics laboratory of the Faculty of Science and Technology of Cheikh Anta Diop University in Dakar for molecular study where they were recoded according to their locality of origin: ScBa (Central African-Bambari sesame); ScBo (Central African sesame-Bossembele); Scke (Central African sesame-Kémo); ScKo (Central African sesame-Kouango); SpBu (Pakréssaya sesame from Burkina Faso).

3.1.2 DNA extraction

Total DNA was extracted using the Zymo Research Kit, taking for each accession or variety a pool of five seeds constituting an individual . The seeds were ground using a porcelain mortar. Each ground material thus obtained was then transferred to a 1.5 ml Eppendorf tube. In each tube, 95 µl of water, 95 µl of buffer [solid tissue Buffer (Blue)] and 20 µl of proteinase K were added to break the bonds between the tissues in order to individualize the cells. This mixture was vortexed and incubated at 70 °C for three (3) hours during which the tubes were vortexed every 30 minutes until the end of the incubation time. 400 µl of Genomic Bending Buffer (lysis buffer) was then added to each tube and the mixture was immediately vortexed for 15 seconds to prevent precipitate formation. The supernatant was transferred to a Zymo-Spin column equipped with a silica membrane and a collection tube, then centrifuged at 13,000 rpm for one minute to fix the DNA on the silica membrane. The collection tube and its contents were then discarded. The silica column was then placed in a new collection tube so that the negatively charged DNA would bind to the positively charged silica membrane. On the other hand, proteins, lipids and polysaccharides were not retained by the membrane and passed into the collection tube which was discarded. The DNA thus fixed on the silica membrane is purified

Commented [TP6]: Some steps (e.g., DNA purification and gel electrophoresis) are described in excessive detail, making the section harder to follow. These could be summarized without losing critical information.
Many technical terms and steps are included but not adequately explained (e.g., "recoding" of samples, choice of 28S gene).
No mention of controls or replicates in PCR and sequencing processes, which are critical for validation.

Commented [TP7]: The accessions come from four different localities (Bambari, Bossembélé, Kemo, Kouango) of Central Africa."
- Issue: "Central Africa" is vague. Specify whether it refers to the Central African Republic or the broader region.

Commented [TP8]: 1."**The samples were then sent to the genomics laboratory of the Faculty of Science and Technology of Cheikh Anta Diop University in Dakar for molecular study where they were recoded according to their locality of origin."**
- Issue: The sentence is too long and lacks clarity. "Recoded according to their locality of origin" needs explanation—what does recoding mean in this context?

Commented [TP9]: "Taking for each accession or variety a pool of five seeds constituting an individual."
- Issue: The term "constituting an individual" is ambiguous. Clarify whether this represents a composite sample or a replicate.

Commented [TP10]: "This mixture was vortexed and incubated at 70 °C for three (3) hours..."
- Issue: The incubation time (3 hours) seems excessive compared to typical DNA extraction protocols. Was this step validated for the kit used?

by successive addition of 400 µl of pre-wash buffer (DNA Pre-Wash Buffer) which pass through the silica membrane and are collected in collection tubes after centrifugation for 1 minute at 13000 rpm. 700 and 200 µl of wash buffer (g-DNA Wash Buffer) are added respectively and collected in the collection tube after centrifugation at 13000 rpm for 1 min and 3 min respectively. The silica column was then placed on an Eppendorf tube. 50 µl of DNA Elution Buffer, previously heated to 70 °C in order to increase the extraction yield, were deposited thereon. The tubes were then incubated at room temperature for five minutes and centrifuged at 13,000 rpm for one minute to allow the DNA to detach from the silica membrane and end up in the tubes, which will then be stored at -20°C.

The quality of the DNA extracts was assessed after electrophoretic migration on a 2% agarose gel at 100 volts for 35 minutes. In each well, 7 µl of DNA extract were mixed with 2 µl of bromophenol blue. One of the wells contained 6 µl of molecular weight marker (100 bp DNA Ladder) to serve as a reference. Before casting the gel, 10 µl of SafeView was added to the gel to allow revelation under a blue light source.

3.1.3 PCR amplification and sequencing of the 28S gene

The 28S gene was chosen in this present study. The 28S gene does not code for proteins but its transcription gives an rRNA **(Figure 1)** which itself adopts a stable secondary and tertiary structure to perform catalytic functions in protein synthesis. For this gene, mutations are not conserved but cause a compensatory effect from one strand to another according to the hydrogen bonds between the bases **[9]** .

Commented [TP11]: "The 28S gene was chosen in this present study."
- Issue: No justification for why the 28S gene was chosen. Explain why this marker was preferred over others (e.g., ITS regions or chloroplast markers).

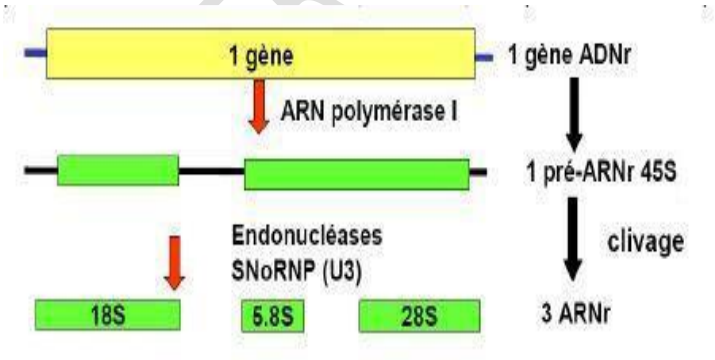

Figure 1: Descriptive diagram of the transcription of the ribosomal gene into 28S catalytic RNA or ribozyme **[9]** .

28S gene amplification was performed using the OneTaq® Quick-Load® 2X Master Mix with Standard Buffer kit, with a reaction volume of 25 µl containing: 2 µl of DNA extract; 9.5 µl of

ultrapure water; 12.5 µl of Master mix and 0.5 µl of each primer which are: F: 5'-TACCGTGAGGGAAAGTTGAAA-3' and R: 5'-AGACTCCTTGGTCCGTGTTT-3'.

This is a three-step PCR that was performed with an Eppendorf-type thermocycler, under the following conditions: initial denaturation at 94°C (3 min); 35 cycles [denaturation at 92°C (30 sec); annealing at 52°C (45 sec); elongation at 72°C (1 min)]; final elongation at 72°C (10 min). To see if the primers have stuck, an electrophoretic migration on 2% agarose gel at 100 volts for 30 minutes was performed after depositing 5 µl of PCR product mixed with 2 µl of loading blue in the wells of the gel. Six (6) µl of molecular weight marker were deposited in one of the wells to serve as a reference.

The 28S gene was sequenced using the forward primer. Sequencing reactions were performed in a MJ Research PTC-225 Peltier thermal cycler with ABIPRISM BigDye TM Terminator Cycle kits. Each sample was sequenced using the forward primer.

3.1.4. Analysis of genomic sequences

3.1.4.1 Genetic polymorphism

To highlight the similarities between the sequences but also to be able to carry out correctly phylogenetic analyses, sequences were carefully checked and corrected manually and by BioEdit software version 7.0.5.3 **[10]** , referring to the chromatograms . Sequence alignment was performed using the Clustal W multiple alignment algorithm **[11]** . Genetic diversity parameters are used to assess the degree of gene variability in the population. These parameters, including sample size, variable sites (V), invariable sites (C), non-informative variable sites (S), informative variable sites (Pi), total number of mutations, number of haplotypes, average number of nucleotide differences, as well as genetic diversity indices (haplotypic and nucleotide diversity) were highlighted by DNASP software version 5.10 **[12]** .

On the other hand, the nature of the mutations (transition or transversion) and the mutation rate were determined using MEGA software version 7.0.14 **[13]** . For genetic diversity indices, haplotype diversity and nucleotide diversity values of 0.5 and 0.05 respectively are considered high (strong) .

3.1.4.2 Differentiation and genetic structuring

In order to see if the gene variability is sufficient to create genetic differentiation between groups, but also within each group, two indices have been estimated: the genetic distance D **[14]** , and the genetic differentiation factor or F_{ST} **[15]** . When the genetic distance is large, the genetic similarity is lower and the divergence time greater **[16]** . It is intended to measure the average number of substitutions that occurred after the divergence of two populations, and is expected to increase linearly with time **[17]** .

Commented [TP12]: "To see if the primers have stuck..."
- Issue: Informal phrasing. Use precise terminology like "To confirm primer binding and amplification..."

Commented [TP13]: "Sequences were carefully checked and corrected manually and by BioEdit software..."
- Issue: Manual correction introduces potential bias if not standardized. Specify how corrections were validated.

Commented [TP14]: "Genetic diversity indices show a high hd (0.857) and a low Pi (0.15)."
- Issue: This statistical interpretation should be reserved for the Results section. It is out of place here.

Genetic distance D assumes that the rate of gene substitution per locus is uniform across loci and lineages.

According to Wright, the closer the F_{ST} is to 1, the more the populations are genetically structured among themselves. On the other hand, populations do not present allelic differences if the F_{ST} is zero. For each value of the F_{ST}, the P-value allows us to accept or reject it depending on whether it is respectively significant or not significant. A p-value less than 0.05 allows us to say that the F ST is significant.

Intra- and inter-tissue genetic distances were calculated with the MEGA software **[14]** . The intra- and inter-tissue $F_{ST\ values}$ were explained with the Arlequin program version 3.5.2.2 **[18]** . With this same program, the genetic structuring of the population according to the groups was studied by doing the analysis of molecular variance (AMOVA: Analysis of Molecular Variance). This procedure seeks to estimate the indices of genetic structure using information on the allelic content of the haplotypes, as well as their frequencies **[19]** . The significance of the covariance components associated with the different hierarchical levels (based on a priori groupings of the total population into subpopulations) is tested using the 10,000 permutations procedure.

3.1.7.3 Demographic evolution

To distinguish sequences whose evolution follows a neutral evolutionary model from those evolving according to a non-random process, the *mismatch distribution curve* , which is the graphical representation of the distribution of genetic distances existing between individuals in a population, was designed with the DnaSP software version 5.10.01 **[13]** and from a constant-size population model. A multimodal distribution is the signal of a stable population (of constant size) and a unimodal distribution represents an expanding population. Genetic diversity indices (haplotypic and nucleotide diversity) have been used to infer demographic evolution.

3.1.7.4 Construction of phylogenetic trees

Phylogeny attempts to reconstruct the evolutionary filiations leading to the sequences studied. In other words, it allows, from aligned sequences, the suggestion of a phylogenetic tree, an oriented graph that attempts to reconstruct the history of successive divergences during evolution, describing the relationships between a group of current taxa (sequences) and their hypothetical common ancestors.

The phylogenetic affinities between our sesame subpopulations were estimated by four methods, the first three of which were designed using MEGA software version 7.0.14 **[14]** and the last using Mr Bayes software version 3.2.6 **[20]** :

- *neighbor-joining* method , whose algorithm starts by constructing a distance matrix (by the p-distance method) between each pair of sequences. The sequences are then grouped according to their relative distance, in order to produce a tree.
- The maximum parsimony method searches all possible topologies in order to find the most parsimonious tree, that is, the one whose total number of changes to informative sites is minimized, in order to explain our dataset.

The maximum likelihood method, a probabilistic method, allows to test all the histories that could have generated the current dataset analyzed by maximizing the probability. It is estimated using the Kimura 2 parameter model. The Bayesian method, also probabilistic, uses the concept of posterior probabilities, where the probability is estimated on the basis of the model proposed by the software, with one million generations and four Markov chains. For this method, the creation of an input file is necessary: the latter is submitted to the software and the calculations are stopped if the frequency is less than or equal to 0.015. The tree is visualized with the Fig Tree software version 1.3.1 **[21].**

Node resolution in phylogenetic trees was assessed by repeating 1,000 *bootstraps* . A *bootstrap* is only considered significant if its value is greater than 70%.
Phylogenetic trees were rooted with an *outgroup sequence* from *Harpagophytum procumbens* (NCBI accession no. LS999881.1), a species in the same family [as] Sesamum *indicum* , because there were no 28S gene sequences from other species in the genus *Sesamum.* in the NCBI database.

3.2 Results

3.2.1. Revelation of DNA extracts

DNA was extracted for all individuals with fluorescent bands for all individuals. Smears were noted *for* all individuals which indicate DNA degradation (**Figure 2**).

Commented [TP15]: Interpretation Missing: While results are presented, there is little to no interpretation or connection to the study's objectives (e.g., what the genetic diversity means for sesame cultivation).
Figures and Tables: Figures and tables are mentioned but not sufficiently described or integrated into the narrative.
Conceptual Gaps: Key findings like Fst values and phylogenetic trees need a clearer explanation of their significance.

Commented [TP16]: "DNA was extracted for all individuals with fluorescent bands for all individuals. Smears were noted for all individuals which indicate DNA degradation."
- Issue: This sentence is repetitive and unclear. Smears typically indicate degradation, but no mention is made of steps taken to address or minimize it.

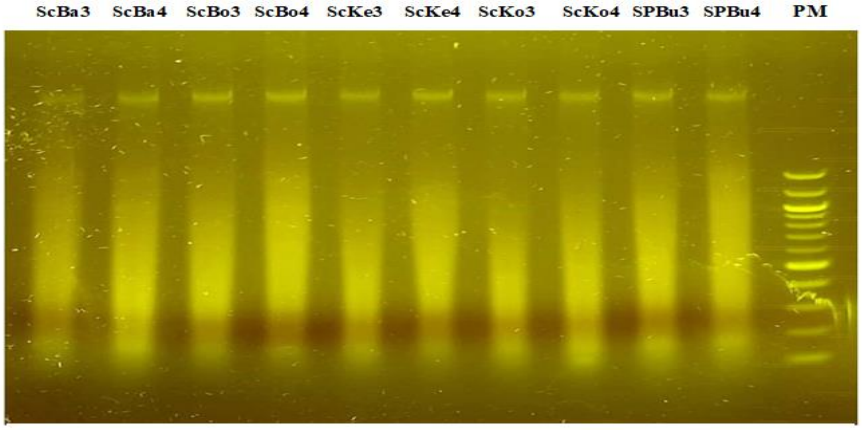

Figure 2: Electrophoretic migration profile of sesame seed extracts .

3.2.2. Control of PCR products

Figure **3** shows successful amplification of the 28S gene from four accessions and one sesame variety with clear bands at different positions showing non-specificity for primer binding.

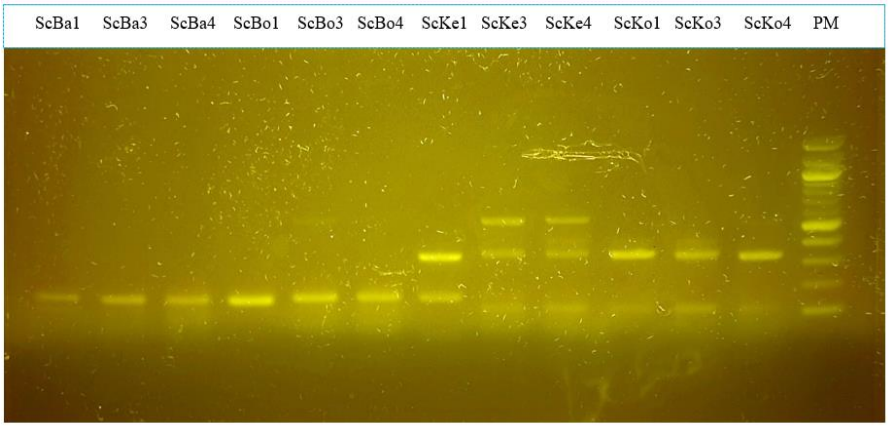

Figure 3: Electrophoretic migration profile of 28S gene amplicons from sesame seed DNA extracts

3.2.3 Alignment of 28S gene sequences

Sequencing yielded fourteen 28S gene sequences with three (3) individuals per locality, except for the group of individuals from Kémo where there are only two (2) correct sequences. The alignment of these sequences is highlighted in Figure 4.

Commented [TP17]: Figure 2:
- Issue: No description of the figure's relevance. How do the bands confirm DNA quality or quantity? Smears suggest degraded samples; was this considered in downstream analyses?

Commented [TP18]: "Clear bands at different positions showing non-specificity for primer binding."
- Issue: If non-specificity is noted, explain how this issue was resolved or its impact on the results.

Commented [TP19]: Figure 3:
- Issue: No explanation of how the figure validates successful PCR amplification. Include interpretation (e.g., expected band size).

Commented [TP20]: "Sequencing yielded fourteen 28S gene sequences with three (3) individuals per locality, except for the group of individuals from Kémo where there are only two (2) correct sequences."
- Issue: Lack of detail on why only two sequences were obtained for Kémo. Were the others of poor quality or contaminated? Explain.

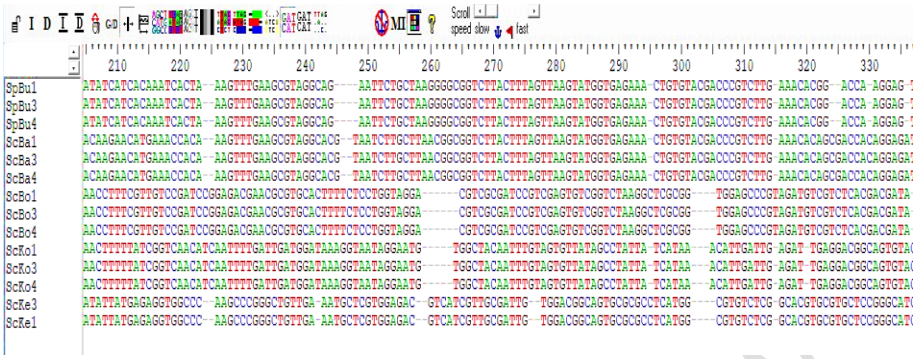

Figure 4: Aligned DNA sequences of the 28S gene of the five ecotypes from Central Africa and Burkina Faso (Take an area with variability).

3.2.4 Molecular characteristics of 28S

For all fourteen (14) sequences, the parameters of genetic variability are reported in **Table 1** . There are 39 sites with gaps out of the 420 sites representing sequence size or nucleotide count. Compared to the non-gap sites, 274 sites are conserved and 107 are variable, the latter being all informative variable sites. Examples of conserved or variable sites are highlighted in Figure **5** . The fourteen sequences are subdivided into five haplotypes, with 161 mutations in total, and a fairly high average number of nucleotide differences of 57.06. The sequences have more transversion (66.66%) than transition (33.33%).

Genetic diversity indices show a high hd (0.857) and a low Pi (0.15).

Table I : Parameters of genetic diversity in the global population

Settings	Global population
Sample size	14
Number of sites	420
Sites with gaps	39
Invariable sites (preserved) C	274
Variable sites (polymorphic) V	107
Non-informative variable sites (singleton) S	0
Variable informative sites	107
Total number of mutations (Eta)	161
Number of haplotypes (h)	5
Average number of nucleotide differences (K)	57.06
Transitions (%)	33.33
Transversion (%)	66.66
Mutation rate (R) transition/transversion	0.5
Haplotypic diversity (hd) ± variance	0.857 ± 0.002
Nucleotide diversity (Pi) ± variance	0.15±0

The genetic diversity parameters for each subpopulation are given in **Table II.** Of the 412 sites representing the sequence size and for all subpopulations—Bambari, Bossembélé, Kouango, Kémo, and Burkina—eight, 22, thirteen, twelve, and twelve sites with gaps are reported, respectively. Bambari sesame has the lowest rate among the sites with gaps (8), while Bossembélé sesame has the highest rate among the same sites (22). The number of variable sites is zero for all subpopulations, as are most of the genetic diversity parameters. The same percentages of transition and transversion are found for all groups. Each subpopulation is represented by one (1) haplotype.

Table II: Genetic diversity parameters for each subpopulation

Settings	Bambari	Bossembélé	Kouango	Kemo	Burkina
Sample size	3	3	3	2	3
Number of sites	412	412	412	412	412
Sites with gaps	8	22	13	12	12
Preserved sites (C)	412	398	407	408	408
Variable sites (V)	0	0	0	0	0
Singleton sites (S)	0	0	0	0	0
Information sites	0	0	0	0	0
Mutation number (Eta)	0	0	0	0	0
Number of haplotypes (h)	1	1	1	1	1
Average number of nucleotide differences (K)	0	0	0	0	0
Transitions (%)	33.33	33.33	33.33	33.33	33.33
Transversion (%)	66.66	66.66	66.66	66.66	66.66
Mutation rate (R) transition/transversion	0.5	0.5	0.5	0.5	0.5
Haplotypic diversity (hd) ± variance	0	0	0	0	0
Nucleotide diversity (Pi) ± variance	0	0	0	0	0

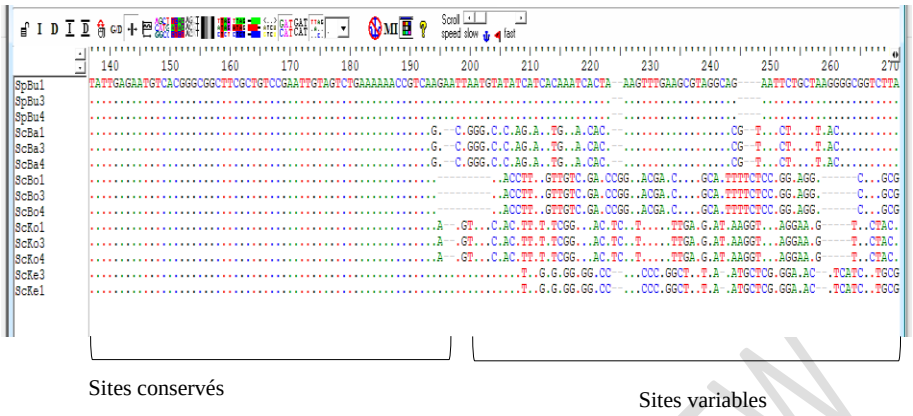

Figure 5: Homology sites of a 28S gene sequence alignment

3.2.5 Differentiation and genetic structuring

The intrapopulation genetic distance is highlighted in **Table III** , which shows zero genetic distances. On the other hand, the genetic distances are slightly high and quite close when comparing populations (**Table IV)** with values ranging from 0.286 to 0.208. However, a low genetic distance value (0.067) is found between the sequences of the pakréssaya variety and those of the Bambari accession.

Commented [TP21]: "The intrapopulation genetic distance is highlighted in Table III, which shows zero genetic distances."
- Issue: This finding is not interpreted. What does "zero genetic distance" imply about the populations?

Table III : Intrapopulation genetic distance (D)

Populations	Bambari	Bossembélé	Kemo	Kouango	Burkina Faso
Distance (D)	0	0	0	0	0

Table IV: Interpopulation genetic distance (D)

Bambari Bossembélé Kémo Kouango subpopulations				
Bambari				
Bossembélé	0.279±0.028			
Kemo	0.286±0.027	0.208±0.029		
Kouango	0.252±0.029	0.231±0.03	0.230±0.013	
Burkina Faso	0.067±0.028	0.250±0.026	0.218±0.026	0.229±0.031

The Fst values between sesame accession subpopulations from different localities are very high and identical when comparing all groups (Fst=1). However, the p-values *are* not significant (**Table V**).

Table VI , which highlights the source of molecular variance, indicates that all variability (100%) is explained by clustering of sesame subpopulations by locality of origin.

Commented [TP22]: "The Fst values between sesame accession subpopulations from different localities are very high and identical..."
- Issue: The phrase "identical" needs clarification. If Fst values are high (approaching 1), it indicates strong genetic differentiation—explain this finding.

Table V : Interpopulation genetic differentiation factors (Fst)

Fst (P-Value)				
Subpopulations	**Bambari**	**Bossembélé**	**Kemo**	**Kouango**
Bambari				
Bossembélé	1 (0.106)			
Kemo	1 (0.090)	1 (0.097)		
Kouango	1 (0.109)	1 (0.100)	1 (0.100)	
Burkina Faso	1 (0.092)	1 (0.113)	1 (0.091)	1 (0.096)

3.3.3 Demographic evolution

The analysis of the mismatch distribution , which corresponds to the measurement of nucleotide differences per pair of observed sites, shows a multimodal curve (in red in **Figure 6**).

This analysis, as well as the demogenetic indices inferring the evolution of a population, could not be carried out on the different sub-populations due to an absence of nucleotide differences between the sequences of individuals from the same locality.

Commented [TP23]: "The multimodal curve represented for this analysis is the signal of a stable or constant population, i.e. well structured and quite evolved."
- Issue: The interpretation of a multimodal curve as "quite evolved" is conceptually unclear. Multimodal curves typically indicate a stable population size, not necessarily "evolution."

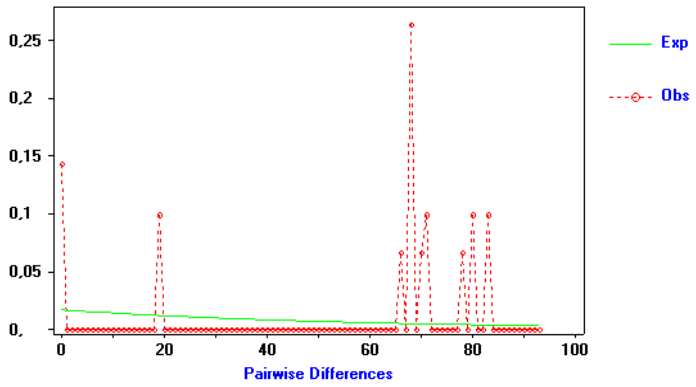

Figure 6: *Mismatch distribution* curve for the entire population.

3.3.4 Reconstruction of phylogenetic trees

At this level of analysis, the same findings emerge regardless of the method of constructing phylogenetic trees: *neighbor-joining method* (**figure 7**); maximum parsimony method (**figure 8**); maximum likelihood method (**figure 9**); Bayesian inference method (**figure 10**). Each subpopulation is classified into a subclade specific to it. However, the Bossembélé subpopulation clearly differentiated itself from the others from the start (clade in blue), a bit like the *outgroup* , which is another species. Unlike the Bossembélé group, the other subpopulations show a certain phylogenetic link and this link is more important between individuals from Burkina and those from Bambari (clade in red), which share a clade with a node (hypothetical ancestor) robust with a resolution of 100% on all trees — except that of *neighbor-joining* (96%).

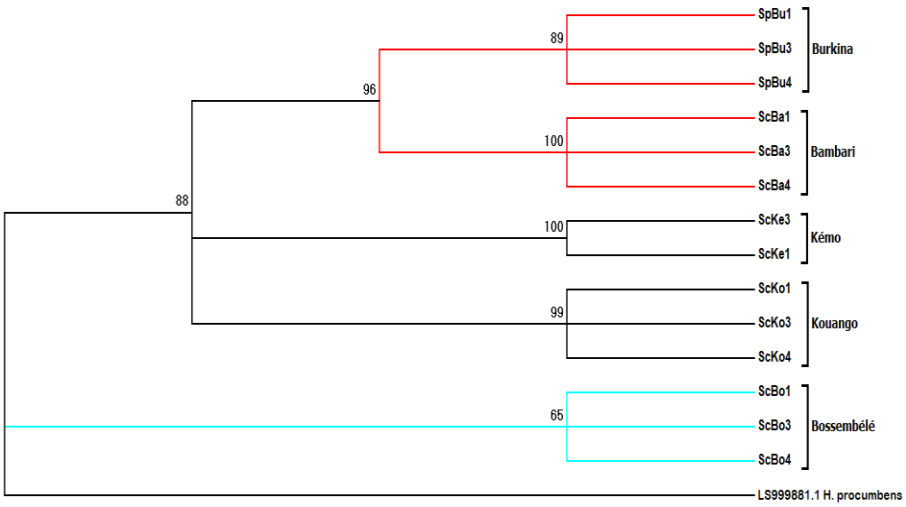

Figure 7: Phylogenetic tree constructed using the *neighbor-joining method*

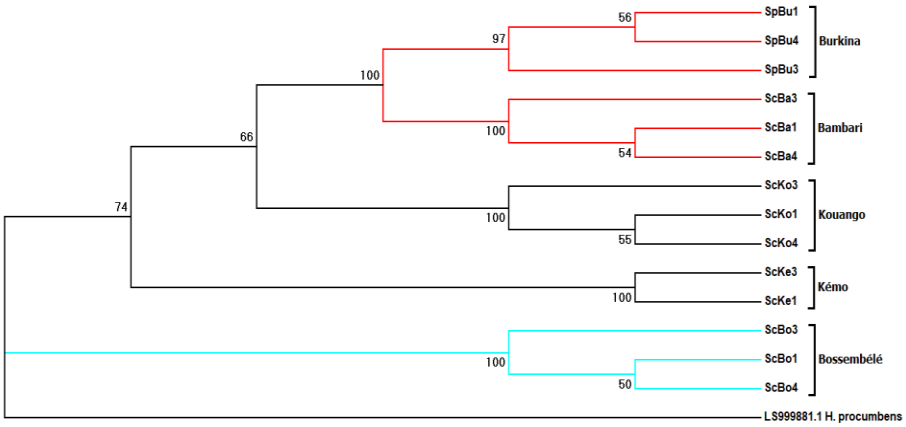

Figure 8: Phylogenetic tree constructed using the maximum parsimony method

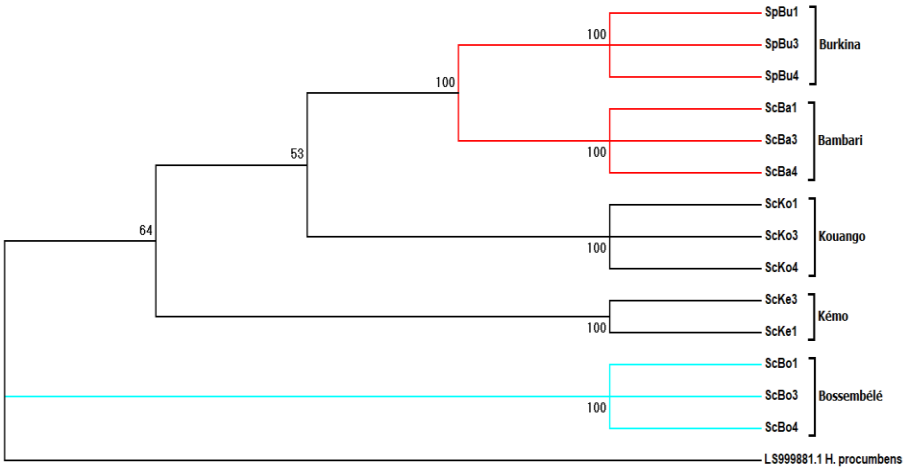

Figure 9: Phylogenetic tree constructed using the maximum likelihood method

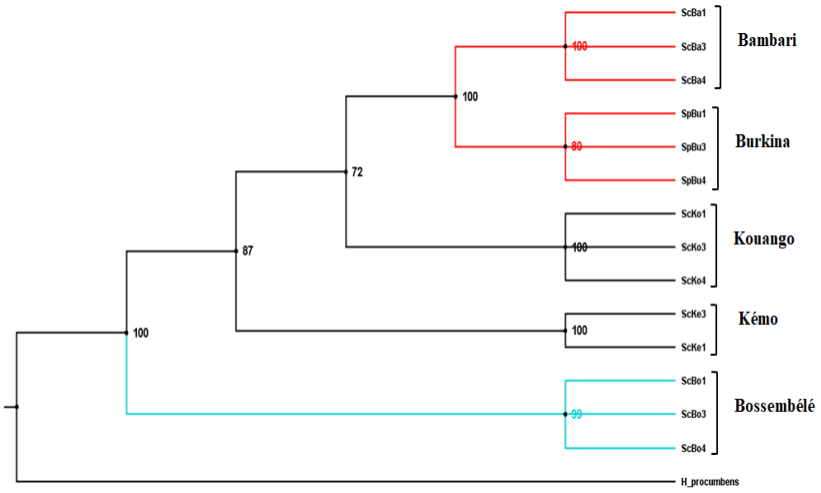

Figure 10: Phylogenetic tree constructed with Bayesian inference

3.3 Discussion

3.3.1 Genetic variability

Assessing the genetic diversity of crop species is essential for breeders to have good starting material for breeding purposes. Genetic diversity of crop species can be studied using various methods, including 28S gene sequencing.

Commented [TP24]: "Assessing the genetic diversity of crop species is essential for breeders to have good starting material for breeding purposes."

- Issue: Generic and repetitive; this point is made earlier in the paper. Focus on how the results specifically contribute to breeding efforts for sesame.

"The results also showed a transversion rate (66.66%) equal to double the transition rate (33.33%)."

- Issue: The biological implication of this ratio is not discussed. Why is the transversion rate high, and what might it indicate about evolutionary pressures or genome stability?

"These results are comparable to those reported by Ercan and his collaborators..."

- Issue: The comparison to Ercan's study lacks depth. Highlight specific similarities or differences, such as the magnitude of diversity or methods used.

The results of genetic variability show a significant polymorphism of the 28S gene in the sesame seed population. Indeed, this difference between individuals at the genetic level is highlighted by the presence of sites with gaps (39 in total out of 420 sites), which can mean insertions or deletions of nucleotides, the number of variable sites (107) which are all informative sites, a significant total number of mutations (Eta = 161) and a fairly high average number of nucleotide differences (K = 57.06) showing a large nucleotide variation. The results also showed a transversion rate (66.66%) equal to double the transition rate (33.33%). Transitions can be spontaneous mutations not resulting from the effect of evolutionary factors, unlike transversions which are much more complex mutations to perform. These results are comparable to those reported by Ercan and his collaborators on the analysis of genetic diversity in Turkish sesame populations **[22].** The number of haplotypes is five in total, as is the number of subpopulations. This is not a coincidence, since the analysis of the parameters of genetic variability within the subpopulations showed that sesame individuals from the same locality do not present any polymorphism, which is highlighted by the haplotype value which is 1 for all subpopulations. This is explained by the absence of variable sites (V = 0) acting so that all other parameters are also zero. Thus, no genetic diversity exists within the subpopulations. However, this diversity is found in the overall population.

3.3.2 Genetic structuring

The polymorphism (genetic diversity) of the 28S gene found through the analyses of genetic variability in the sesame population from different localities prompted us to seek confirmation of genetic differentiation between subpopulations. All values of intrapopulation genetic distance (within subpopulations) are zero. Analysis of molecular variance (AMOVA) allows comparing groups and evaluating them. This analysis of molecular variance (AMOVA) showed a percentage of zero variations within subpopulations. These results confirm the absence of variability between the sequences of the 28S gene of sesame seeds from the same locality. However, there is indeed genetic differentiation between subpopulations through the more or less high interpopulation genetic distance values (which vary from 0.286 to 0.208) and the maximum Fst values between subpopulations (Fst = 1). This genetic differentiation is 100% explained by the subdivision of the population, according to the locality of origin, by the results of AMOVA analysis, highlighting a genetic structuring of this population. The analyses of the genetic structuring thus showed that there was no resemblance between the sesame accessions of Central Africa, but also between these and the Pakréssaya variety of Burkina Faso. This interpopulation genetic differentiation is certainly linked to environmental conditions (i.e. climate). Unlike our study, which is based on the sequencing of the 28S gene of sesame, other

Commented [TP25]: "All values of intrapopulation genetic distance (within subpopulations) are zero."
- Issue: This statement is not sufficiently interpreted. Explain why there is no variation within subpopulations and its implications for genetic uniformity.

"The analyses of the genetic structuring thus showed that there was no resemblance between the sesame accessions of Central Africa, but also between these and the Pakréssaya variety of Burkina Faso."
- Issue: Contradictory to earlier findings where some genetic similarity between Bambari and Pakréssaya was noted. Provide a consistent interpretation.

"Climatic variations between localities, such as temperature, precipitation, and humidity, can influence natural and artificial selection of traits."
- Issue: This is speculative. No direct evidence is presented linking the observed genetic differentiation to climate.

studies based rather on molecular markers of sesame have also revealed detailed information on its genetic diversity **[23]** which are consistent with the present study. The varieties cultivated in the localities of Kouango, Kémo, Bambari, Bossembélé and Kouango have developed different traits between them and the Pakréssaya variety imported from Burkina Faso. Climatic variations between localities, such as temperature, precipitation, and humidity, can influence natural and artificial selection of traits. For example, sesame grown in a very hot and dry region might develop genetic traits that favor drought tolerance, while sesame grown in a more humid region might exhibit traits adapted to cooler conditions **[24].** However, even though the 28S gene in this case does not encode a protein, it still yields an rRNA that constitutes ribosomes that are the center of amino acid synthesis. To date, many applications combining several techniques have been used to assess genetic diversity and genetic relationships among plants **[25], [26], [27].** Our results revealed 100% genetic variability among the different sesame groups studied. In the same context, this polymorphism rate is higher than that reported by kumar and sharma (57%) when they explored the genetic diversity of Indian sesame **[28]** and among Korean sesame accessions (33%) **[29]** . These results are consistent with the previous study of Abate and co-workers, who reported that sequencing and the use of molecular markers are appropriate methods to identify high genetic diversity in sesame germplasm **[30]** .

Our results concluded that 28S gene sequencing is useful for assessing genetic diversity among Central African sesame populations. However, the lowest genetic distance value (D=0.067) was found between the sequences of the Pakréssaya variety and that of the Bambari accession. This resemblance is justified by the fact that the Bambari accession and the Pakréssaya variety would come from the same lineage and could have common ancestors. This reflects this significant genetic similarity, even if they are cultivated in different localities. Genetic variations inherited from ancestors can explain the similarity in some traits **[31]** . It should be noted, however, that if climatic conditions and cultivation practices are comparable in the two localities, varieties can also develop similar traits in response to similar environmental pressures. If two localities have similar climatic conditions, varieties can exhibit similar traits in response to these conditions. Soil composition and quality also influence plant growth. Taiz and colleagues confirmed that climatic conditions (temperature, humidity, sunshine) also play a crucial role in plant development and physiology. Similar soils in two different localities can therefore lead to similar characteristics between varieties **[32]** . The resemblance between the Pakréssaya variety and the sesame accession grown in Bambari could therefore result from a combination of genetic and environmental factors, as well as cultural practices.

3.3.3 Demographic evolution

Commented [TP26]: **"The multimodal curve represented for this analysis is the signal of a stable or constant population."**

- Issue: The explanation of the multimodal curve is overly simplified. Provide more context on why this pattern was observed (e.g., lack of recent population bottlenecks or expansions).

"The above analyses resulted in null or impossible results."

- Issue: This phrasing is unclear. Why were these analyses null or impossible, and how does it affect the study's conclusions?

For To highlight the evolutionary model of the sesame population, genetic diversity information was used for *mismatch distribution analyses* . The multimodal curve represented for this analysis is the signal of a stable or constant population, i.e. well structured and quite evolved. Similarly, the high values of haplotypic diversity (hd = 0.857) and nucleotide diversity (Pi = 0.15) that emerged from the genetic variability analyses reflect a stable population. These results further confirm the existence of genetic structuring in our sesame population.

However, due to the absence of genetic variability within the subpopulations, the above analyses resulted in null or impossible results. It can therefore be assumed that the sesame accessions from different localities (Bossembélé, Bambari, Kémo, Kouango), as well as the Pakréssaya variety from Burkina, are populations at the very beginning of demographic expansion or having undergone a selective sweep.

3.3.4 Phylogenetic trees

To establish a link between the different sesame accessions from Central Africa (Bambari, Bossembélé, Kouango and Kémo) and the Pakréssaya variety from Burkina Faso, four phylogenetic trees were created using different methods (the *neighbor-joining method* , the maximum parsimony method, the maximum likelihood method and Bayesian inference). The results obtained confirm everything that has been found on the polymorphism and genetic structuring of sesame. Indeed, each subpopulation is represented by a subclade specific to it, i.e. no other sequence from another group appears there. Similarly, the result of the lowest genetic distance found between the Bambari accession and the Pakréssaya variety was represented at the level of the phylogenetic trees by the same clade composing the two groups with a largely robust node, exceeding 90%.

The different trees also highlight a progressive differentiation between the subpopulations. The phylogenetic trees therefore confirm that interpopulation genetic diversity is linked to geographical isolation, which reinforces the thesis that the more distant the localities are from each other, the greater the polymorphism. This confirms the work of Zhivotovsky and Krauss, who mention that, for essentially self-pollinating species, phylogenetic trees allow a better estimation of genetic diversity **[33]** .

Indeed, the Kouango sequences seem genetically close to the Burkina-Bambari subclade, but at the same time differentiate themselves from this group. Similarly, the Kemo sequences differentiate themselves from the other aforementioned groups since they form a subclade independent of the others. However, the accession of Bossembélé according to the different

Commented [TP27]: "The results obtained confirm everything that has been found on the polymorphism and genetic structuring of sesame."

- Issue: Overly vague and uninformative. Explain which specific findings (e.g., geographic isolation or genetic clustering) are supported by the trees.

"This could be justified by its geographical position in the north of the Central African Republic, a region dominated by savannah and therefore with a low rainfall rate."

- Issue: Speculative reasoning. There is no data presented in the study directly linking genetic differentiation to savannah or rainfall.

phylogenetic trees is found to be distant from the other subclades and has differentiated itself from the start. This could be justified by its geographical position in the north of the Central African Republic, a region dominated by savannah and therefore with a low rainfall rate. Although the 28S gene is one of the most reliable means to conduct this type of study, its use for the characterization of sesame seeds is limited. This is the reason why we do not have enough scientific data related to the species studied (Sesamum *indicum*).

Conclusion

The 28S gene sequencing was effective in differentiating sesame (*Sesamum indicum*) accessions , as it showed significant polymorphism between the different accessions studied and the Pakréssaya variety. The 28S gene sequencing, combined with other analysis techniques such as phylogenetic tree construction methods (maximum parsimony, maximum likelihood, *neighbor-joining* , Bayesian inference), genetic differentiation and structuring, and demographic evolution, to name a few, represents a robust tool for identifying genetic structure in populations of species. The intrapopulation genetic distance is zero. The analysis of the genetic diversity of Central African sesame allowed us to lift the veil on the different sesame accessions present in the Central African territory. What we take away from this study is the uniformity of seeds belonging to the same locality, which means that all accessions from the same locality look similar. This could be related to the distances between localities, which do not allow for seed exchanges between farmers.

However, interpopulation genetic diversity is much greater than intrapopulation diversity. This polymorphism would be related to geographical distances. By comparing sesame accessions

from Central Africa with the Pakréssaya variety, the genetic distance values show considerable genetic variability. The results obtained confirm that 28S gene sequencing is effective in detecting polymorphism within and between sesame accessions collected in geographical locations in Central Africa, but also in comparison with foreign genetic materials (Pakréssaya variety). Phylogenetic trees revealed a similarity between the Bambari accession and the Pakréssaya variety. This similarity has no link with the geographical diversity of the accession collection locations. These results may have agronomic implications in order to improve the production and development of sesame cultivation throughout the Central African territory.

.

REFERENCES.

1. Bedigian, D., 2003. Evolution of sesame revisited: domestication, diversity and prospects. Genet Resour Crop Evol. 50:779-787.
2. Abou-Gharbia HA, Shehata AAY, Shahidi F. 2000. Effect of processing on oxidative stability and lipid classes of sesame oil. Food Research International, 33: 331–340.
3. Anilakumar K, Pal A, Khanum F, Bawa A. 2010. Nutritional, Medicinal and Industrial Uses of Sesame (Sesamum indicum L.) Seeds – An Overview. Agriculturae Conspectus Scientificus. 75(4): 159-168. DOI: https://hrcak.srce.hr/66001 .
4. Walter, O.-A., Patrick, R., Paul, G., & Patrick, O. 2016. Genotype by environment interaction in sesame (Sesamum indicum L.) cultivars in European Scientific Journal May 2020 edition Vol.16, No .15 ISSN: 1857-7881 (Print) e - ISSN 1857-7431 211 Uganda. African Journal of Plant Science, 10(10), 189 -202. https://doi.org/10.5897/AJPS2016.1426 .
5. Demol J, Baudoin JP, Louant BP, Maréchal R, Mergeai G, Otoul E. 2002. Plant breeding: Application to the main species cultivated in tropical regions. Les Presses Agronomiques de Gembloux: Gembloux; 581 p.

6. Diallo, B., Samba, S., Sane, D., & Diop, T. 2014. Effect of sodium chloride on the germination of Ricinus communis L. seeds. International Journal of Biological and Chemical Sciences, 7(4), 1534. https://doi.org/10.4314/ijbcs.v7i4.10 .
7. Freeland, J.R., 2005. Molecular Ecology. West Sussex, John Wiley and Sons Ltd.
8. Esquinas-Alcazar, J., 2005. Protecting crop genetic diversity for food security: political, ethical and technical challenges. Nat Rev Genet, 6(12), 946-953.
9. Hancock, JM, Tautz , D. & Dover GA 1988. –Evolution of secondary structures and compassory mutations of the ribosomal RNAs of *Drosophyla melanogaster* . *Mol. Biol. Evol.* , **5** (4): 393-414.
10. Hancock, JM, Tautz, D. & Dover GA 1988. –Evolution of secondary structures and compassory mutations of the ribosomal RNAs of *Drosophyla melanogaster* . *Mol. Biol. Evol.* , **5** (4) 393-414.
11. Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* , 41: 95 – 98.
12. Thompson JD, Higgins DG, Gibson TJ (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". *Nucleic Acids Research* ; 22: 4673 – 4680. DOI: 10.1093/nar/22.22.4673.
13. Librado P., & Rozas J., 2009. DnaSP version 5: software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* , 25:1451–1452.
14. Kumar S., Stecher G., & Koichiro T. 2016. "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." *Molecular Biology and Evolution* ; 33 (7); p. 1870 -74. DOI: https://doi.org/10.1093/molbev/msw054 .
15. Nei M. 1987. "Molecular Evolutionary Genetics". Columbia University press, New York, 506p.
16. Wright S. 1950. "The Genetical Structure of Populations." *Annals of Eugenics* ; 15 (1)), p. 323 -54. DOI: https://doi.org/10.1111/j.1469-1809.1949.tb02451.x .
17. Avise JC 1994. "Molecular Markers, Natural History and Evolution. »Chapman and Hall, New York. DOI: http://dx.doi.org/10.1007/978-1-4615-2381-9 .
18. Takezaki N. & Nei M. 1996. "Genetic Distances and Reconstruction of Phylogenetic Trees From Microsatellite DNA". *Genetics* ; 144 (1), p. 389 -99. DOI: https://doi.org/10.1093/genetics/144.1.389 .

19. Excoffier L., Lischer HEL, 2010. Arlequin suite version 3.5: a new series of programs to perform population genetics analyzes under Linux and Windows. *Molecular Ecology Resources* , 10:564–567.

20. Excoffier L., Smouse P., & Quatro J., (1992). "Analysis of molecular variance inferred from metric distance among DNA haplotypes: Application to human mitochondrial DNA restriction data". *Genetics* ; 131, p. 479-491. DOI: 10.1093/genetics/131.2.479

21. C. Ling, A. Luo and J. Gao, " Mr Bayes 3.2.6 on Tianhe-1A: A High Performance and Distributed Implementation of Phylogenetic Analysis," 2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS), Wuhan, China, 2016, pp. 1181-1186, doi: 10.1109/ICPADS.2016.0156.

22. Ambaut A., 2014 : Figtree, a graphical viewer of phylogenetic trees. V.1.4.2.

23. Ercan, AG, Taskin, M. Turgut, K. (2004). Analysis of genetic diversity in Turkish sesame (Sesamum indicum L.) populations using RAPD markers ⋆ . Genetic Resources and Crop Evolution, 51(6), 599-607.

24. Kim, DH, Zur, G., Danin-Poleg, Y., Lee, SW, Shim, KB, Kang, CW, Kashi, Y. (2002). Genetic relationships of sesame germplasm collection as revealed by inter simple sequence repeats. Plant Breeding. 121(3), 259-262.

25. Smith, J., & Johnson, A. (2020). Environmental Impact on Sesame Cultivation: Climate, Soil, and Agricultural Practices. Journal of Agricultural Sciences, 45(3), 235-248.

26. Antony JJJ, Shamshir RA, Poobathy R., Chew B., Subramaniam S. (2015). Somaclonal variations were not induced by cryopreservation: Somaclonal variation levels of in vitro and thawed protocorms of Dendrobium Bobby Messina analyzed by SCoT and TRAP DNA markers. South Afr. J. Bot. 100, 148–157. 10.1016/j.sajb.2015.04.012 [CrossRef] [Google Scholar].

27. Thakur J., Dwivedi MD, Singh N., Uniyal PL, Goel S., Pandey AK (2021). Applicability of start codon targeted (SCoT) and inter simple sequence repeat (ISSR) markers in assessing genetic diversity in crepidium acuminatum (D. Don) szlach. J. Appl. Res. Med. Aromat. Plants 23, 100 310. 10.1016/j.jarmap.2021.100310 [CrossRef] [Google Scholar]

28. Tikendra L., Potshangbam AM, Amom T., Dey A., Nongdam P. (2021). Understanding the genetic diversity and population structure of Dendrobium chrysotoxum Lindl.- An

endangered medicinal orchid and implications for its conservation. South Afr. J. Bot. 138, 364–376. 10.1016/j.sajb.2021.01.002 [CrossRef] [Google Scholar]

29. Kumar V., Sharma SN (2011). Comparative potential of phenotypic markers, ISSR and SSR for characterization of sesame (Sesamum indicum L.) varieties from India. J. Crop Sci. Biotechnol. 14, 163–171. 10.1007/s 12892-010-0102-z [CrossRef] [Google Scholar]
30. Kim DH, Zur G., Danin-Poleg Y., Lee SW, Shim KB, Kang CW, et al. (2002). Genetic relationships of sesame germplasm collection revealed by simple sequence repeats among them. Plant Breed. 121, 259–262. 10.1046/j.1439-0523. 2002.00700.x [CrossRef] [Google Scholar].
31. Abate M., Mekbib F., Ayana A., Nigussie M. (2015). Assessment of genetic diversity of Ethiopian sesame (Sesamum indicum L.) germplasm using ISSR markers. Brother. Biotechnology. J. 8, 1-13. 10.9734/BBJ/2015/18481 [CrossRef] [Google Scholar].
32. Bedigian, Dorothea. (2016). Botany and Cultivation of Sesame, 4 (3): 24-34. Source: Plant Science Today.
33. Taiz, Lincoln, and Zeiger, Eduardo, Environmental Physiology of Plants 3rd Edition Publisher: Sinauer Associates 2010: 345-390.