

## Research Article

# Site-Specific Insertion Polymorphism of the MITE *Alex-1* in the Genus *Coffea* Suggests Interspecific Gene Flow

Christine Dubreuil-Tranchant,<sup>1</sup> Romain Guyot,<sup>1</sup> Amira Guellim,<sup>1</sup> Caroline Duret,<sup>1,2</sup> Marion de la Mare,<sup>1</sup> Norosoa Razafinarivo,<sup>1,3</sup> Valérie Poncet,<sup>1</sup> Serge Hamon,<sup>1</sup> Perla Hamon,<sup>1</sup> and Alexandre de Kochko<sup>1</sup>

<sup>1</sup>IRD, UMR DIADE, Centre IRD de Montpellier, BP 64501, 34394 Montpellier Cedex 5, France

<sup>2</sup>UMR ECOFOG INRA, BP 709, 97387 Kourou Cedex, Guyane Française, France

<sup>3</sup>FOFIFA, BP 1444, Ambatobe, Antananarivo 101, Madagascar

Correspondence should be addressed to Alexandre de Kochko, dekochko@ird.fr

Received 1 March 2011; Accepted 16 July 2011

Academic Editor: A. Rus Hoelzel

Copyright © 2011 Christine Dubreuil-Tranchant et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Miniature Inverted-repeat Transposable Elements (MITEs) are small nonautonomous class-II transposable elements distributed throughout eukaryotic genomes. We identified a novel family of MITEs (named *Alex*) in the *Coffea canephora* genome often associated with expressed sequences. The *Alex-1* element is inserted in an intron of a gene at the *CcEIN4* locus. Its mobility was demonstrated by sequencing the insertion site in *C. canephora* accessions and *Coffea* species. Analysis of the insertion polymorphism of *Alex-1* at this locus in *Coffea* species and in *C. canephora* showed that there was no relationship between the geographical distribution of the species, their phylogenetic relationships, and insertion polymorphism. The intraspecific distribution of *C. canephora* revealed an original situation within the E diversity group. These results suggest possibly greater gene flow between species than previously thought. This MITE family will enable the study of the *C. canephora* genome evolution, phylogenetic relationships, and possible gene flows within the *Coffea* genus.

## 1. Introduction

Recently, a new type of molecular marker based on the insertion polymorphism of transposable elements (TEs) was shown to be particularly effective for plant diversity studies [1–4]. Because of their repeated nature and, in some cases, their great number [5, 6], these mobile genetic elements may be inserted at different loci in the genome where they lead to mutations or chromosomal rearrangements. Their activity is responsible for considerable natural polymorphism that can be used to study within and between species diversity and to identify possible population genetic structure and phylogenetic relationships [7, 8].

Among these mobile genetic elements, MITEs (Miniature Inverted-repeat Transposable Elements) form a particular group. MITEs are short (<600 bp) nonautonomous type II transposable elements that are often quite widely distributed in eukaryote genomes but at the same time are highly

conserved, within a genome, in size and sequence, indicating that they might originate from a limited number of progenitors [9]. Their even distribution throughout plant genomes makes them an ideal tool for the study of genome evolution and genetic relationships [10–12].

Such elements could help to solve problematic phylogenetic relationships among species including those in the genus *Coffea*, which comprises 103 species originating from Africa, Madagascar, and several islands in the Indian Ocean [13]. A phylogenetic tree was constructed based on four plastid sequences: *trnL* intron, *trnL-F* IGS, *rpl16*, and *accD-psaI* IGS and one nuclear repeated sequence: rDNA ITS [13]. The tree contains valuable information but also many un-solved relationships concerning the evolution of the genus and the speciation process, especially in Madagascar. Several approaches using TE could be used to solve this problem such as SSAP [14] and REMAP [15].

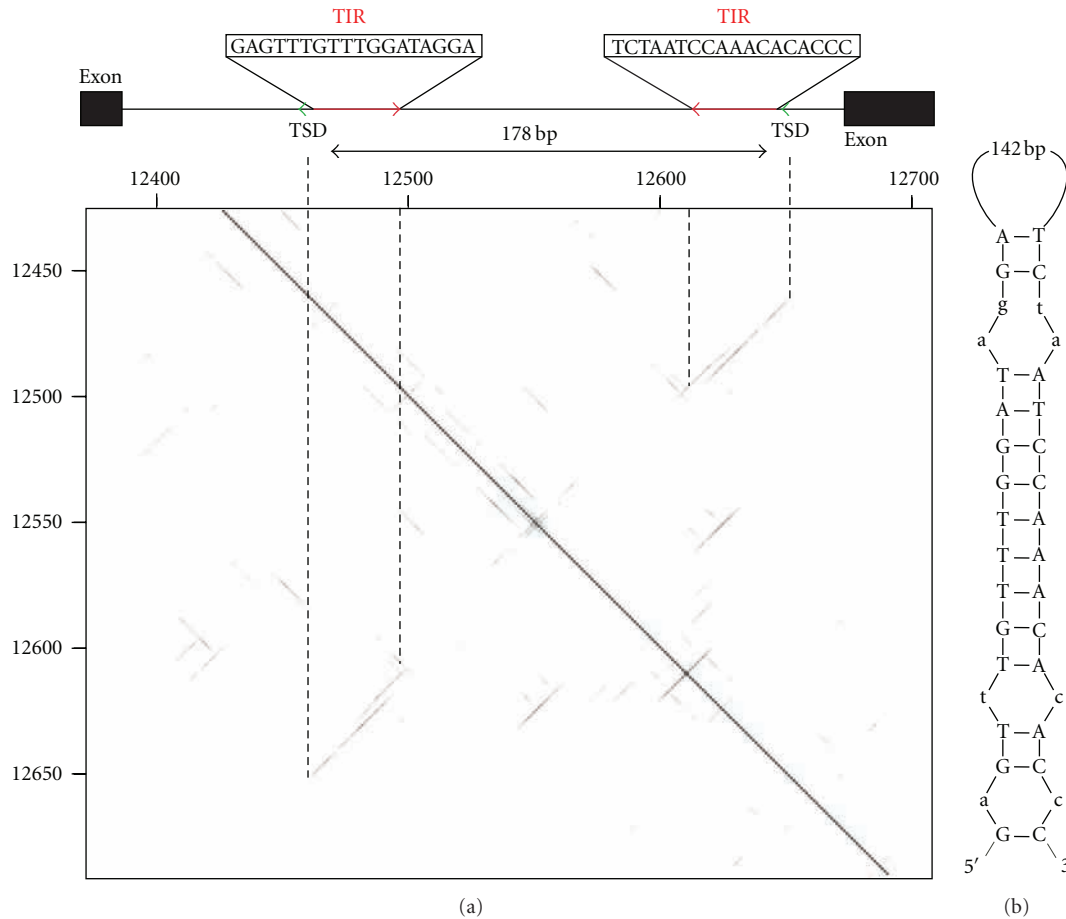


FIGURE 1: Structural characterization of the *Alex-1* MITE. (a) Dot plot of the MITE against itself allowed the identification of the 18 bp terminal inverted repeats (in red). (b) Folding of *Alex-1* revealed the typical stem loop structure of MITE elements.

Both approaches enable estimation of the genome-wide TE distribution. A third experimental procedure reveals site-specific insertion polymorphism [16]. This method requires the identification of information on flanking sequences to facilitate the design of primers to detect polymorphism by PCR. Sequencing of a *C. canephora* BAC clone (46C02, accession no. EU164537) enabled identification of a new MITE, named *Alex-1* (Figure 1), in the 12th intron of a gene (*g3*) of this BAC clone [17].

In this paper, we characterize this novel MITE family in the *C. canephora* genome. We also report the results of a study on the insertion polymorphism of the MITE *Alex-1* at the *g3* locus using PCR approaches on a representative set of *Coffea* species and a representative set of *C. canephora* diversity groups [18–21]. Our results revealed high insertion polymorphism of the *Alex* MITE at the *g3* locus, which was not linked to the phylogenetic relationships of the *Coffea* species studied here. Taken together, these results suggest greater gene flow between species than previously thought.

## 2. Material and Methods

**2.1. Plant Material.** Twenty-eight *Coffea* species grown in tropical greenhouses at the IRD center in Montpellier

(France) were used in this study. They represent the natural ecogeographical distribution of the genus. One plant from a related genus, *Psilanthus*, was included in the survey. One to 11 plants (genotypes) per species were analyzed (depending on the number of samples available in the collection) except for *C. canephora* for which 71 accessions were included (see Supporting Material available at doi: 10.4061/2011/3584122). For the latter species, DNAs from 12 plants from the Ugandan diversity group were kindly provided by P. Musoli from NARO/COREC (Uganda) and T. Leroy from CIRAD (France).

**2.2. DNA Isolation.** DNA was extracted and purified using Qiagen DNeasy mini Kits (Hilden, Germany) according to the manufacturer's instructions. DNA quantification was performed on a NanoDrop TM 1000 Spectrophotometer (LabTech, France).

**2.3. Identification of the MITE *Alex*.** *De novo* identification of the MITE was performed using dot-plot alignments (DOTTER software, [22]) based on the presence of inverted repeats at both ends and target site duplication. Evaluation of the redundancy of the MITE in nucleotide sequences was

conducted by BLAST searches using MITE *Alex* as a query against public databases of *C. canephora* (56,231 Expressed Sequence Tags (EST) (with an *e*-value cut-off of  $10e^{-4}$ ).

**2.4. PCR Amplifications.** Two primers were designed in conserved regions according to the 46C02 BAC sequence (accession no. EU164537) to check the presence/absence of the MITE *Alex-1* at the *g3* locus in the different accessions used in this study. The forward primer was designed in the 11th exon of that gene and the reverse primer in the 13th exon (see Supporting Material 1). The primer sequences used were:

G3F: 5' GTT-TGG-TTG-CTG-GGT-CTC-AT 3' and G3R: 5' CGA-CAA-GAG-GAA-AGC-CTC-AC 3'. The expected amplicon is 1093 bp long when the MITE *Alex-1* is present and 916 bp when it is absent (see supporting Material 1).

The PCR conditions were 94°C for 1 min. followed by 35 cycles at 94°C for 1 min, 58°C for 30 sec, 72°C for 45 sec, and a final elongation period at 72°C for 4 min. PCR products were observed by electrophoresis in 1% agarose gel after staining with ethidium bromide.

**2.5. PCR Product Sequencing.** In order to check the absence of insertion—as oppose to excision—of a former inserted MITE *Alex-1* at the *g3* locus, several PCR products from different species were sequenced. After electrophoresis, the bands at 1093 or 916 bp were excised from the gel using a razor blade. DNA was purified using a Quiagen PCR purification kit according to the manufacturer's recommendations and sent for sequencing to Eurofins-MWG (Ebersberg, Germany). Sequences were aligned using ClustalW software.

### 3. Results

**3.1. Identification of *Alex*, a Novel MITE Family, in *C. canephora*.** A BAC clone at the *CcEIN4* locus in *C. canephora* (BAC clone 46C02, accession no. EU164537) was recently sequenced [17]. This represented the first complete BAC clone ever sequenced in the *Coffea* genus. The typical structural features of a Miniature Inverted-repeat Transposable Element (MITE) were detected in the 12th intron of the *g3* gene, encoding a putative protein (nucleotides 12468...12645 of the BAC clone). This element, named *Alex-1*, was flanked by the 3 bp direct repeat AGT, generated upon the insertion of the element (Target Site Duplication, TIR) and had 18 bp Terminal Inverted Repeats (TIR) at both ends of the element. The small sequence size (178 bp), rich A/T composition (73.4%), and the ability to form secondary structures characterized *Alex-1* (Figure 1).

To further characterize this element, the nucleotide sequence of *Alex-1* was compared with the public nucleotide sequences of *C. canephora* comprising 56,231 Expressed Sequence Tags. BLAST searches produced 42 significant hits, suggesting that *Alex-1* belongs to a large family of MITE elements frequently associated with transcribed sequences. A BLASTN search performed on nonredundant (nr) public

libraries did not produce any significant hit (length, percentage, identity, and *e*-value), except in *Coffea* genomic sequences.

**3.2. Analysis of Genomic Polymorphism Associated with the MITE *Alex-1* at the *g3* Locus.** Table 1(a) shows the presence or absence of the *Alex-1* MITE at the *g3* locus in the different plants analyzed. Within-species polymorphism in *C. canephora* is presented in Table 1(b) (see Supporting Material 3).

The majority of the species (18/28) displayed total absence of the MITE at the *g3* locus, whereas 7/28 were homozygous for its presence and 3/28 displayed heterogeneous patterns. Only *C. liberica* var. *devewrei* and *C. canephora* displayed all three genotypes, homozygous +/+, homozygous -/-, and heterozygous +/-.

Two species or taxons showed a majority of homozygous genotypes, +/+ (*C. sp* N'Koumbala) or -/- (*C. humilis*), few heterozygotes +/-, but not the reciprocal homozygous, -/- (*C. sp* N'Koumbala) or +/+ (*C. humilis*).

At the sequence level, a closer look at the locus of insertion for the presence of *Alex-1* in both positive and negative plants showed that only one of the negative individuals, *C. anthonyi*, displayed the remnant of a TSD (Target Site Duplication) sequence that indicates the former presence of a MITE and thus its transposition (Figure 2). This process of excision was precise since the whole element was removed from the site of insertion, and no large deletion occurred in the flanking regions.

Interestingly, in the African species, no link was found between the geographical origin of the species and the presence of the *Alex-1* MITE. Indeed, *Alex-1* was present in East, West and Central African species (Mozambicoffea and Eucoffea, resp.). However, all the Mascarocoffea species originating from islands in the Indian Ocean (Madagascar, Mauritius, and Comoros) lacked the MITE at the *g3* locus. Most *Coffea* species appear to be homozygous since only seven plants out of 129 were heterozygous.

The situation of *C. canephora* is particularly interesting since the only diversity group E, originating from the Congo/Cameroon region, contains homozygous -/- genotypes (Table 1(b) and see Supporting Material 4). The hetero-zygotes detected in groups D, A, and C were previously identified as being intergroup hybrids all with a group E genotype in their pedigree [19]. Similarly, heterozygotes in group E turned out to be hybrids between group D and group E genitors [19]. The homozygous individual in group E was collected in RCI (Ivory Coast), far from the place of origin of that diversity group, which is in the Congo/Cameroon region. This particular plant was certainly introduced into RCI for improvement purposes quite a long time ago and has certainly undergone several crosses and backcrosses leading to an introgressed form bearing the inserted locus on both homologous chromosomes (+/+) (Accession 319, see Supporting Material 2 and 4). The diversity group E thus appears to be the only one among the *C. canephora* groups to be characterized by the absence of *Alex-1* at the *g3* locus.

TABLE 1

(a) Insertion polymorphism of the *Alex-1* MITE at the *g3* locus among a representative set of *Coffea* species and a close relative *Psilanthus ebracteolatus*.

Species analyzed	No. of individuals	+/+	-/-	+/-	Origin
<i>C. arabica</i>	3	3	0	0	E. Africa, Ethiopia
<i>C. eugenoides</i>	9	9	0	0	E. Africa Kenya
<i>C. pseudozanguebariae</i>	11	0	11	0	E. Africa Kenya
<i>C. racemosa</i>	11	0	11	0	E. Africa Tanzania
<i>C. liberica</i> var. <i>liberica</i>	10	0	10	0	W. Africa RCI
<i>C. stenophylla</i>	10	0	10	0	W. Africa RCI
<i>C. humilis</i>	10	0	8	2	W. Africa RCI
<i>C. canephora</i>	71	53	8	10	W. and C. Africa
<i>C. congensis</i>	5	5	0	0	C. Africa RCA
<i>C. liberica</i> var. <i>dewevrei</i>	10	4	3	3	C. Africa RCA
<i>C. liberica</i> var. <i>koto</i>	3	3	0	0	C. Africa Cameroon
<i>C. brevipes</i>	10	10	0	0	C. Africa Cameroon
<i>C. heterocalyx</i>	1	0	1	0	C. Africa Cameroon
<i>C. anthonyi</i>	7	0	7	0	C. Africa Cameroon
<i>C. sp</i> N'Koumbala	10	8	0	2	C. Africa Cameroon
<i>C. sp</i> Mayombé	3	3	0	0	C. Africa Congo R.
<i>C. kapakata</i>	2	2	0	0	C. Africa Angola
<i>C. myrtifolia</i>	3	0	3	0	Mauritius
<i>C. resinosa</i>	1	0	1	0	Madagascar
<i>C. tsirananae</i>	1	0	1	0	Madagascar
<i>C. lancifolia</i>	1	0	1	0	Madagascar
<i>C. perrieri</i>	1	0	1	0	Madagascar
<i>C. sakarahae</i>	1	0	1	0	Madagascar
<i>C. millotii</i>	1	0	1	0	Madagascar
<i>C. dolichophylla</i>	1	0	1	0	Madagascar
<i>C. heimii</i>	1	0	1	0	Madagascar
<i>C. bertrandii</i>	1	0	1	0	Madagascar
<i>C. humblotiana</i>	1	0	1	0	Comoros
<i>P. ebracteolatus</i>	1	0	1	0	W. Africa RCI
Total	200	100	83	17	

E. Africa: East Africa; W. Africa: West Africa; C. Africa: Central Africa.

RCI: République de Côte d'Ivoire (Ivory Coast); RCA: République Centre Africaine (Central African Republic).

(b) Insertion polymorphism of the *Alex-1* MITE at the *g3* locus among a representative set of the diversity groups of the *Coffea canephora* species as defined by Gomez et al. [19].

Diversity group	No. of individuals	+/+	-/-	+/-	Origin
D	29	28	0	1* (DEA)	Guinea/RCI
A	2	1	0	1* (AE)	Cameroon/Congo
B	3	3	0	0	RCA
C	9	8	0	1* (CE)	Cameroon/Congo/RCA
E	16	1	8	7* (DE)	Congo/Cameroon/RCA
O	12	12	0	0	Uganda
Total	71	53	8	10	

+/+ &amp; -/-: Homozygote for presence and absence, respectively; +/-: Heterozygote. \*: intraspecific hybrids.

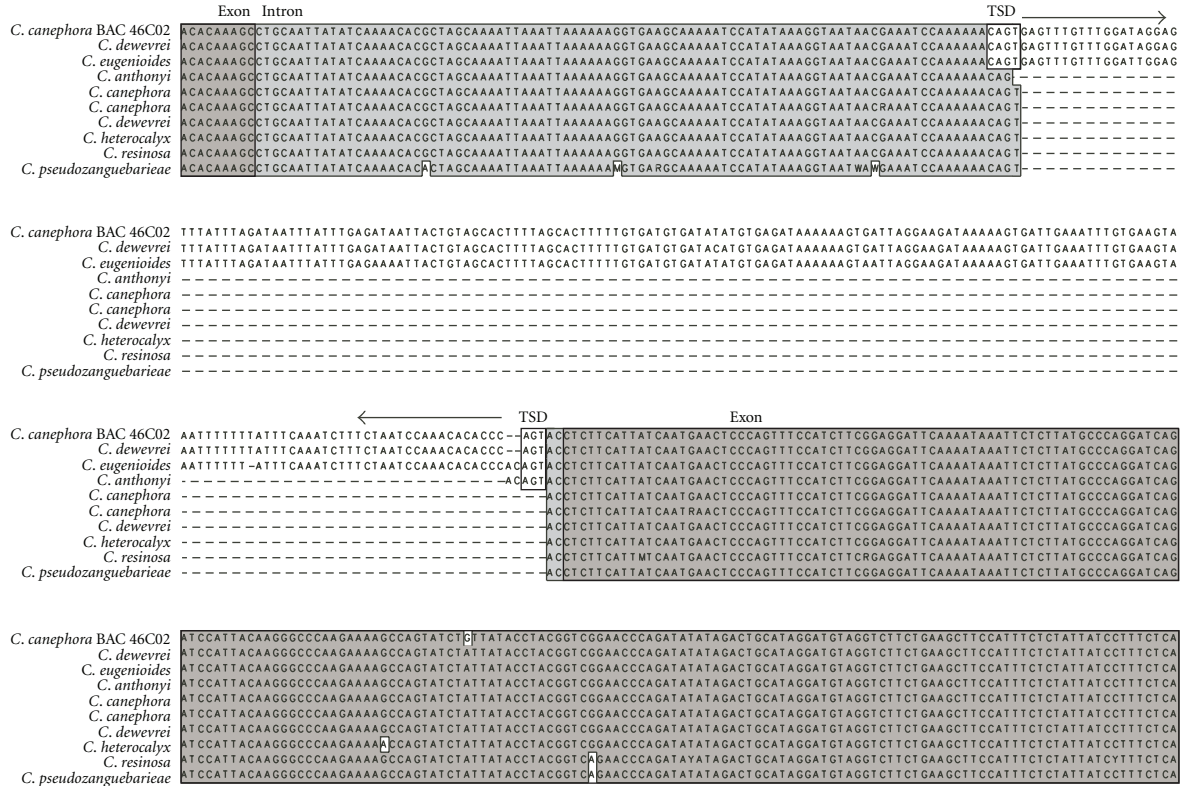


FIGURE 2: Sequence alignment of *Alex-1* insertion site in different *Coffea* species. *C. anthonyi* is the only negative sequence displaying the presence of a Target Site Duplication (TSD) (empty boxes in the figure). Black arrows indicate the presence of the Terminal Inverted Repeats (TIRs).

Analysis of the sequences in seven negative genotypes in the E diversity group and in negative genotypes in other diversity groups revealed no remnant TSD sequences, suggesting that the absence of *Alex-1* at the *g3* locus is more likely due to a lack of insertion than to postdifferentiation excision (Figure 3).

The closest relative to the genus *Coffea*, a plant from the *Psilanthus* genus, also lacked the insertion of the *Alex-1* MITE at the *g3* locus.

#### 4. Discussion

In this paper we describe the first MITE characterized in the genome of a *Coffea* species that has no homologs in other published sequences, it appears thus as specific to the *Coffea* genus. We studied the insertion polymorphism of this MITE at the *g3* locus. The distribution pattern of *Alex-1* at this locus in the *Coffea* and *Psilanthus* species strongly suggests that insertion occurred early in relation with the evolution of these genera, although certainly after the divergence between *Coffea* and *Psilanthus* as *Alex-1* was not found in *Psilanthus*. However, not enough *Psilanthus* species or genotypes were analyzed to confirm this hypothesis.

Because the distribution of the MITE *Alex-1* does not corroborate previous phylogenetic studies [13] and was found in *Coffea* species independently of their eco-geographical distribution (Figure 4), its insertion at the *g3* locus

most probably occurred before the spread of the genus in Africa but probably after the colonization of Madagascar and the other islands in the Indian Ocean by one or several ancestral *Coffea* species if we consider the hypothesis that the genus originated in the African continent.

Its insertion certainly occurred before the formation of the *C. arabica* species, which is the only allotetraploid in the *Coffea* genus, originating from Southern Ethiopia and most probably resulting from a cross between *C. eugenioides* and *C. canephora* [23] both being only or mostly homozygous +/-.

When the insertion of a transposable element occurs, it is always in heterozygous form. The probability that the same TE is inserted at the same locus, at exactly the same spot on both homologous chromosomes, is almost nil. If the insertion does not modify a gene function leading to an advantage or disadvantage in terms of selection, its maintenance in the genome responds to a neutral model and may be conserved or eliminated in the following generations. In the present case, as no link was found between the presence or absence of *Alex-1* at the *g3* locus and the habitat type of the species, the neutral situation probably applies. It is still not clear why species then became preferentially fixed for the presence or absence of the TE, if this was not merely random.

Four species displayed the presence of heterozygous genotypes and only two (*C. canephora* and *C. liberica* var *dewevrei*) showed the three possible patterns (Table 1(a)),



FIGURE 3: Sequence alignment of *C. canephora* accessions displaying the absence of *Alex-1* compared with BAC 46C02 sequence. The absence of a TSD indicates that the MITE was not excised from this site and that insertion never occurred. The accessions sequences were 739, 504, 738, 665, 663, 730, 321, 604, 345, 651, and 725 for wild accessions belonging to A, B, D, and E genetic diversity groups, and to the cultivated BD55 accession from Cameroon. Empty boxes: TSD. Black arrows: TIRs.

homozygous (+/+ and -/-) and heterozygous (+/-). It is very likely that because of the size of the sample, all possible situations have not yet been identified in all the species. It is also possible that some fixation and/or divergence events are actually still underway. The most intriguing example is *C. liberica* var *dewevrei*, which displays the three genotypes (homozygous and heterozygous), while *C. liberica* var *liberica* is fixed for the absence of *Alex-1* (homozygotes -/-) and *C. liberica* var *koto* is fixed for the presence of *Alex-1* (homozygotes +/+). *C. liberica* var *dewevrei* may still be in the fixation process but this could take quite a long time, as the three genotypes are encountered with equal frequency.

In the cases of *C. sp* N’Koumbala and *C. humilis*, no -/- or +/+ homozygotes were identified (Table 1(a)), which does not mean that these types of homozygotes do not exist but simply that they were not present in the sample we analyzed. The presence of heterozygotes (+/-) can result from an allelic equilibrium with a low frequency of positive alleles in *C. sp* N’Koumbala and of a negative allele in *C. humilis*, but in such a situation, the reciprocal homozygote would also be expected to be present, and this was not the case in our sample. Another possible explanation for this low allelic frequency is that interspecific crosses, even if very rare, may happen throughout the Cameroon/Congo region and in RCI, which are hot spots of diversification and secondary centers of speciation for *Coffea* species [24].

The *C. canephora* group E insertion pattern suggests possible interspecific hybridization and gene flow. Indeed, it is the only group in this species that lacked *Alex-1* at the *g3* locus. The absence of TSD in the sequenced amplicons (Figure 3) indicates that *Alex-1* has never been inserted at that locus, and consequently, that its absence is not the result of transposition to another site. All the other genotypes, whatever diversity group they belong to, are +/+ homozygotes, it is thus highly likely that the common ancestor of *C. canephora* underwent the insertion of *Alex-1* at the *g3* locus and then evolved towards the fixation of the insertion (+/+ pattern). Group E, and certainly other unidentified genomic sequences, could then result from an introgression following a cross with a neighboring (sympatric) species and backcrosses to recover *C. canephora* properties.

Interestingly, under this hypothesis of introgression, *C. sp* N’Koumbala is a possible candidate to be the provider of the absence of insertion if an allelic equilibrium remains in this species. Indeed, this taxon grows in the same region as plants from diversity group E, but additional comparative sequencing of the *g3* locus, including flanking regions, is necessary to confirm or reject the hypothesis.

It is also possible that this particular group derives from a sister plant to the plant that integrated the MITE, the latter

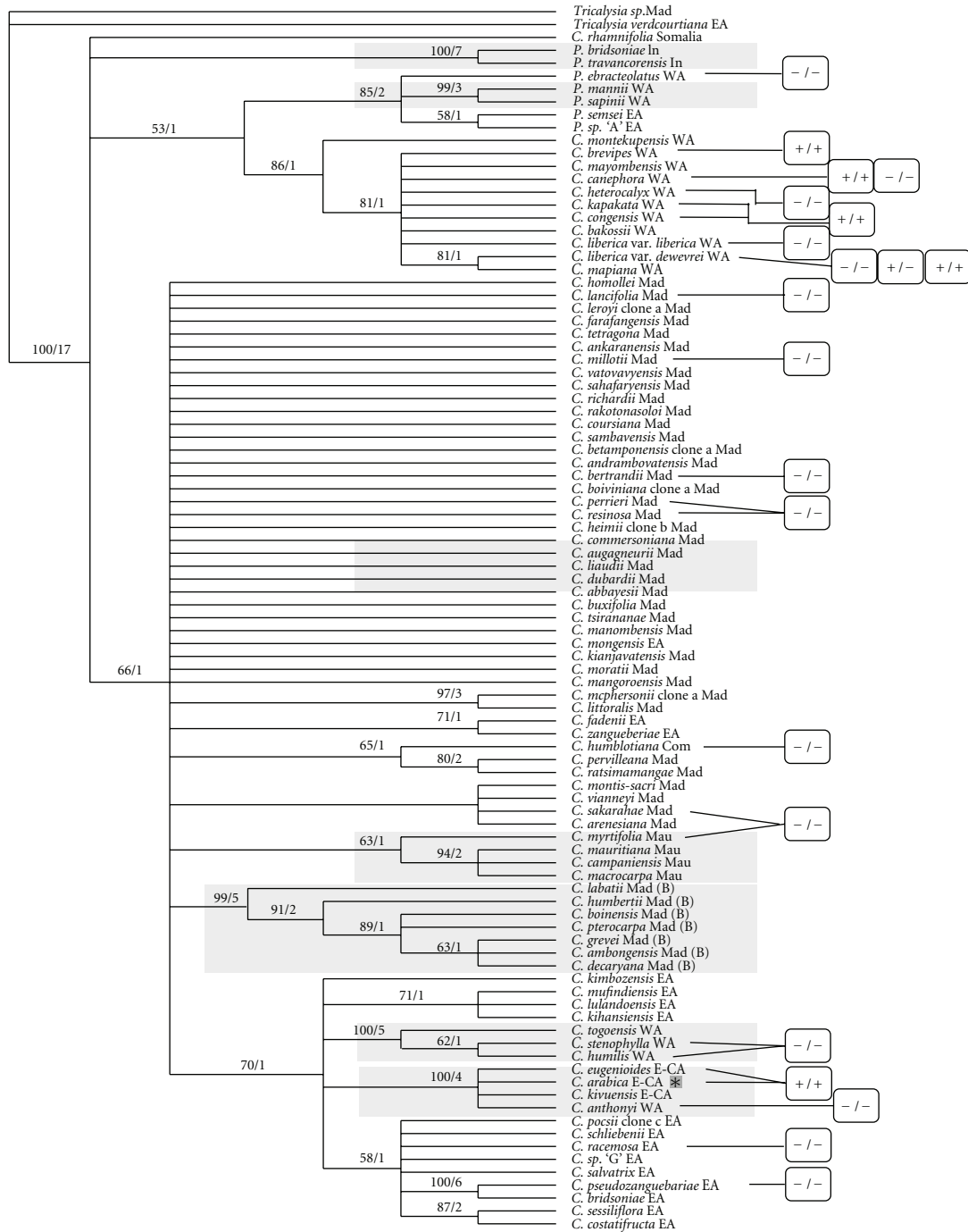


FIGURE 4: Insertion polymorphism of the MITE *Alex-1* superimposed to a phylogenetic tree established by Maurin et al. [13]. Except for the species from the Indian Ocean islands, which are homozygous  $-/-$ , all possible allelic associations are found on the African continent despite the geographical region, West or East. \*: *C. arabica* is the sole tetraploid *Coffea* species. WA: West and Central Africa, Mad: Madagascar, Com: Comoros, Mau: Mauritius, EA: East Africa. All the species analyzed in the present study are not represented on the tree and reciprocally.

resulted in the full *C. canephora* lineage except for group E. In this case, group E may derive from a population that lived in sympatry with *C. canephora*, from which it has never completely genetically separated due to cross hybridization. If this is the case, the genome region that contains the *Alex-1* MITE was preserved from recombination, which should have led to  $+/-$  and  $+/+$  genotypes. However, genotypes that are

found in artificial intraspecific hybrids make this hypothesis unlikely.

## 5. Conclusion

The insertion pattern of the *Alex-1* MITE at the *g3* locus in *Coffea* species indicates an original path of speciesdifferen-

tiation including gene flows between ancestral forms that happened before the present. Recent collecting missions very occasionally identified natural interspecific hybrids or sympatric populations of *Coffea* species. However, it is known that such events can happen in the wild (*C. arabica* being the best example), or in displaced populations in functioning or abandoned coffee plantations [25]. Changing environmental conditions and habitat modification could certainly have led to cohabitation of two or more species in limited areas where their specific phenology was disturbed, thus allowing cross pollination. Subsequent environmental changes could have led to the expansion of favorable habitats, resulting in the isolation of the newly formed species.

MITEs thus appear to be a powerful tool to analyze these speciation events and to trace the phylogenetic relationships between species and if the number of specific insertion sites is sufficient to enable the establishment of an event chronology [8].

## Acknowledgments

The authors thanks go to P. Musoli from NARO/COREC (Uganda) and T. Leroy from CIRAD (France) for kindly providing DNA from Ugandan *C. canephora*.

## References

- [1] S.-J. Kwon, K. L. Ju, S. W. Hong, Y. J. Park, K. L. McNally, and N. S. Kim, "Genetic diversity and phylogenetic relationship in AA *Oryza* species as revealed by Rim2/Hipa CACTA transposon display," *Genes and Genetic Systems*, vol. 81, no. 2, pp. 93–101, 2006.
- [2] K. Takagi, H. Nagano, Y. Kishima, and Y. Sano, "MITE-transposon display efficiently detects polymorphisms among the *Oryza* AA-genome species," *Breeding Science*, vol. 53, no. 2, pp. 125–132, 2003.
- [3] S. J. Wheelan, L. Z. Scheifele, F. Martinez-Murillo, R. A. Irizarry, and J. D. Boeke, "Transposon insertion site profiling chip (TIP-chip)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17632–17637, 2006.
- [4] Z. J. Xu and W. Ramakrishna, "Retrotransposon insertion polymorphisms in six rice genes and their evolutionary history," *Gene*, vol. 412, no. 1-2, pp. 50–58, 2008.
- [5] M. Calonje et al., "Non-coding nuclear DNA markers in phylogenetic reconstruction," *Plant Systematics and Evolution*, vol. 282, no. 3-4, pp. 257–280, 2009.
- [6] J. M. Casacuberta and N. Santiago, "Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes," *Gene*, vol. 311, no. 1-2, pp. 1–11, 2003.
- [7] D. Grzebelus et al., "Population dynamics of miniature inverted-repeat transposable elements (MITEs) in *Medicago truncatula*," *Gene*, vol. 448, no. 2, pp. 214–220, 2009.
- [8] R. J. Mason-Gamer, "Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective," *BMC Evolutionary Biology*, vol. 7, article 92, 2007.
- [9] C. Feschotte, N. Jiang, and S. R. Wessler, "Plant transposable elements: where genetics meets genomics," *Nature Reviews Genetics*, vol. 3, no. 5, pp. 329–341, 2002.
- [10] A. M. Casa, S. E. Mitchell, O. S. Smith, J. C. Register, S. R. Wessler, and S. Kresovich, "Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines," *Theoretical and Applied Genetics*, vol. 104, no. 1, pp. 104–110, 2002.
- [11] S. R. Wessler, T. E. Bureau, and S. E. White, "LTR-retrotransposons and MITEs: important players in the evolution of plant genomes," *Current Opinion in Genetics and Development*, vol. 5, no. 6, pp. 814–821, 1995.
- [12] A. Benjak, S. Boue, A. Forneck, and J. M. Casacuberta, "Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.)," *Genome Biology and Evolution*, vol. 1, pp. 75–84, 2009.
- [13] O. Maurin, A. P. Davis, M. Chester, E. F. Mvungi, Y. Jauferally-Fakim, and M. F. Fay, "Towards a phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences," *Annals of Botany*, vol. 100, no. 7, pp. 1565–1583, 2007.
- [14] R. Waugh, K. McLean, A. J. Flavell et al., "Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP)," *Molecular and General Genetics*, vol. 253, no. 6, pp. 687–694, 1997.
- [15] R. Kalendar, T. Grob, M. Regina, A. Suoniemi, and A. Schulman, "IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques," *Theoretical and Applied Genetics*, vol. 98, no. 5, pp. 704–711, 1999.
- [16] A. J. Flavell, M. R. Knox, S. R. Pearce, and T. H. N. Ellis, "Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis," *Plant Journal*, vol. 16, no. 5, pp. 643–650, 1998.
- [17] R. Guyot et al., "Microcollinearity in an ethylene receptor coding gene region of the *Coffea canephora* genome is extensively conserved with *Vitis vinifera* and other distant dicotyledonous sequenced genomes," *BMC Plant Biology*, vol. 9, no. 1, article 22, 2009.
- [18] P. Cubry et al., "Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding," *Genome*, vol. 51, no. 1, pp. 50–63, 2008.
- [19] C. Gomez, S. Dussert, P. Hamon, S. Hamon, A. de Kochko, and V. Poncet, "Current genetic differentiation of *Coffea canephora* Pierre ex a. Froehn in the guineo-Congolian african zone: cumulative impact of ancient climatic changes and recent human activities," *BMC Evolutionary Biology*, vol. 9, no. 1, article 167, 2009.
- [20] P. Musoli, P. Cubry, P. Aluka et al., "Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda," *Genome*, vol. 52, no. 7, pp. 634–646, 2009.
- [21] P. Hamon et al., "Two novel Ty1-copia retrotransposons isolated from coffee trees can effectively reveal evolutionary relationships in the *Coffea* genus (Rubiaceae)," *Molecular Genetics and Genomics*, vol. 285, no. 6, pp. 447–460, 2011.
- [22] E. L. L. Sonhammer and R. Durbin, "A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis," *Gene*, vol. 167, no. 1-2, pp. GC1–GC10, 1995.
- [23] P. Hamon, S. Siljak-Yakovlev, S. Srisuwan et al., "Physical mapping of rDNA and heterochromatin in chromosomes of 16 *Coffea* species: a revised view of species differentiation," *Chromosome Research*, vol. 17, no. 3, pp. 291–304, 2009.
- [24] A. P. Davis, R. Govaerts, D. M. Bridson, and P. Stoffelen, "An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae)," *Botanical Journal of the Linnean Society*, vol. 152, no. 4, pp. 465–512, 2006.



- [25] C. Gomez et al., "Favourable habitats for *Coffea* inter-specific hybridization in central New Caledonia: combined genetic and spatial analyses," *Journal of Applied Ecology*, vol. 47, no. 1, pp. 85–95, 2010.

## **Supporting material legends:**

### **Supporting material 1:**

**The *g3* locus at the 46C02 sequenced BAC clone in *C. canephora*.** The *Alex-1* MITE was found nested in the 12th intron of the *g3* gene. Primers (G3F and G3R) used to amplify the insertion site of *Alex-1* were designed in the 11th and 13th exons of the gene.

### **Supporting material 2:**

**Analyzed *C. canephora* accessions.** Representativeness of the diversity groups of the *C. canephora* species and insertion polymorphism of the *Alex* MITE at the *g3* locus among these accessions. Sequenced genotypes (Figure 3): 321, 345, 651, 663, 665, 504, 604, 725, 730, 738, 739 (wild). Genetic groups are those defined by Gomez & al. 2009. +/+ & -/-: Homozygote; +/-: Heterozygote.

### **Supporting material 3:**

**Polymorphisms associated with the *Alex-1* MITE at the *g3* locus from 59 different *C. canephora* accessions.** Accession origins are representative of the genetic diversity of *C. canephora* detailed in Supporting material 2. PCR amplifications of Ugandan accessions are not shown.

### **Supporting material 4:**

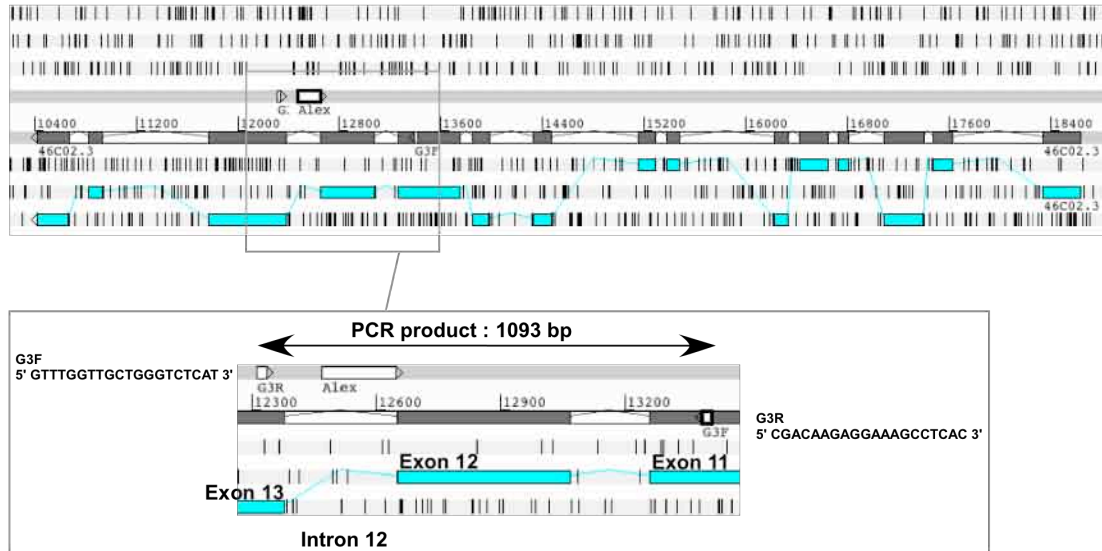
**Geographical distribution of homozygous and heterozygous genotypes for the absence of *Alex-1* among the 71 samples representative of the *C. canephora* diversity groups.** Red stars indicate sequenced individuals (Figure 3). +/+ & -/-: Homozygote; +/-: Heterozygote

RCI = République de Côte d'Ivoire (Ivory Coast)

RCA = République Centre Africaine (Central African Republic)

From Gomez et al. (2009)

## Supporting material 1.

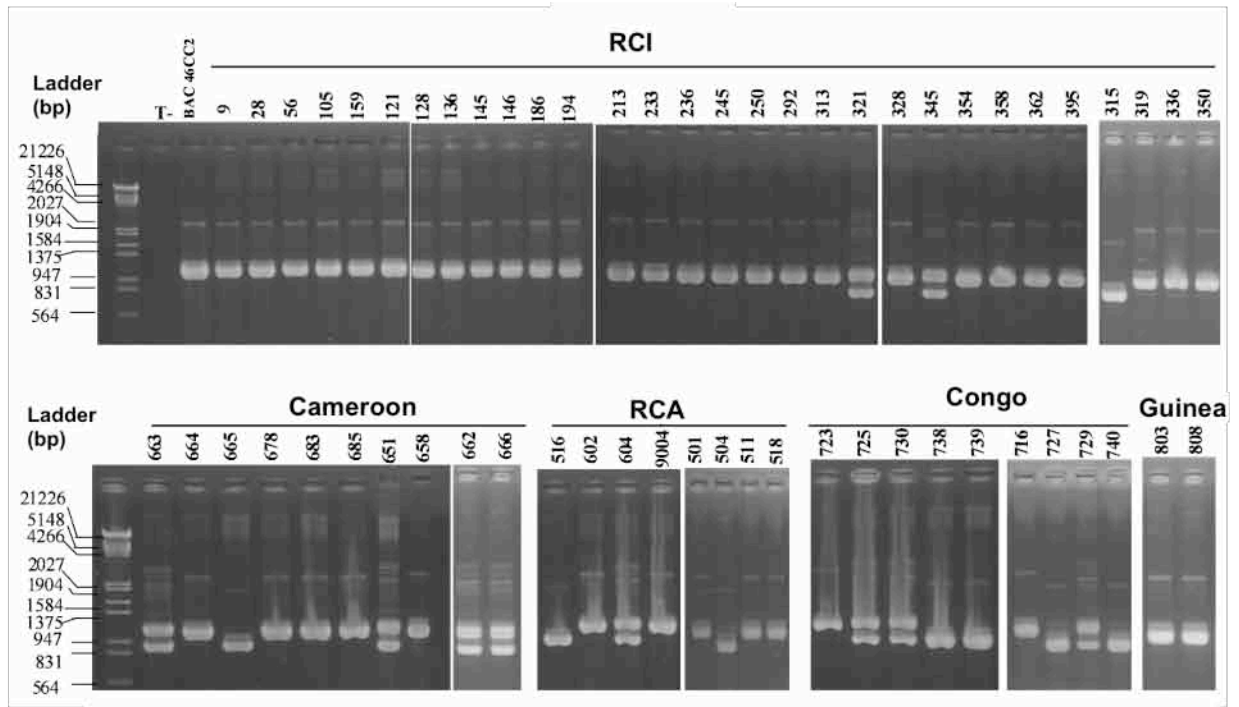


## Supporting material 2.

Wild <i>Coffea canephora</i> genotypes	Coll. #	Diversity Groups	Introgressed plants	Polymorphisms at the <i>g3</i> locus
RCI	9	D		++
	28	D		++
	56	D		++
	105	D		++
	159	D		++
	121	D		++
	128	D		++
	136	D		++
	145	D		++
	146	D		++
	186	D	DA	++
	194	D	DE	++
	213	D		++
	233	D		++
	236	D		++
	245	D		++
	250	D		++
	292	D		++
	313	D		++
	315	E		--
	321	E		+/-
	319	E		++
	328	D		++
	336	D		++
	345	D	DEA	+/-
	350	D	DE	++
	354	D		++
	358	D	DE	++
	362	D		++
	395	D		++
Guinea	803	D		++
	808	D		++
Cameroon	651	E		+/-
	658	C	DA	++
	663	A	AE	+/-
	662	E		+/-
	664	C		++
	665	E		--
	666	E		+/-
	678	C		++
	683	C		++
	685	C		++
RCA	501	B		++
	504	E		--
	511	B		++

	516	E		--
	518	B		++
	602	C		++
	604	C	CE	+/-
	9004	C		++
Republic of Congo	716	C		++
	723	A		++
	727	E		--
	725	E		+/-
	729	E		+/-
	730	E		+/-
	738	E		--
	740	E		--
	739	E		--

### Supporting information 3.



## Supporting information 4.

