

RESEARCH ARTICLE

Open Access

# Organization and molecular evolution of a disease-resistance gene cluster in coffee trees

Alessandra F Ribas<sup>1</sup>, Alberto Cenci<sup>1</sup>, Marie-Christine Combes<sup>1</sup>, Hervé Etienne<sup>2</sup> and Philippe Lashermes<sup>1\*</sup>

## Abstract

**Background:** Most disease-resistance (R) genes in plants encode NBS-LRR proteins and belong to one of the largest and most variable gene families among plant genomes. However, the specific evolutionary routes of NBS-LRR encoding genes remain elusive. Recently in coffee tree (*Coffea arabica*), a region spanning the  $S_H3$  locus that confers resistance to coffee leaf rust, one of the most serious coffee diseases, was identified and characterized. Using comparative sequence analysis, the purpose of the present study was to gain insight into the genomic organization and evolution of the  $S_H3$  locus.

**Results:** Sequence analysis of the  $S_H3$  region in three coffee genomes,  $E^a$  and  $C^a$  subgenomes from the allotetraploid *C. arabica* and  $C^c$  genome from the diploid *C. canephora*, revealed the presence of 5, 3 and 4 R genes in  $E^a$ ,  $C^a$ , and  $C^c$  genomes, respectively. All these R-gene sequences appeared to be members of a CC-NBS-LRR (CNL) gene family that was only found at the  $S_H3$  locus in *C. arabica*. Furthermore, while homologs were found in several dicot species, comparative genomic analysis failed to find any CNL R-gene in the orthologous regions of other eudicot species. The orthology relationship among the  $S_H3$ -CNL copies in the three analyzed genomes was determined and the duplication/deletion events that shaped the  $S_H3$  locus were traced back. Gene conversion events were detected between paralogs in all three genomes and also between the two sub-genomes of *C. arabica*. Significant positive selection was detected in the solvent-exposed residues of the  $S_H3$ -CNL copies.

**Conclusion:** The ancestral  $S_H3$ -CNL copy was inserted in the  $S_H3$  locus after the divergence between Solanales and Rubiales lineages. Moreover, the origin of most of the  $S_H3$ -CNL copies predates the divergence between *Coffea* species. The  $S_H3$ -CNL family appeared to evolve following the birth-and-death model, since duplications and deletions were inferred in the evolution of the  $S_H3$  locus. Gene conversion between paralog members, inter-subgenome sequence exchanges and positive selection appear to be the major forces acting on the evolution of  $S_H3$ -CNL in coffee trees.

## Background

In their natural environment, plants encounter a vast array of pathogenic microorganisms such as viruses, bacteria, oomycetes, fungi and nematodes. To defend themselves against infection by these pathogens, plants employ a network of intertwined mechanisms. One such line of defense is based on dominant disease resistance (R) genes that mediate resistance to pathogens possessing corresponding avirulence (Avr) genes [1]. The largest class of known R genes includes those that encode the nucleotide binding site (NBS) and the leucine-rich repeat (LRR) domains. The

deduced NBS-LRR proteins can be subdivided in classes based on their amino-terminal features [2,3]. The most frequent classes possess a TIR domain with similarity to either the intracellular signaling domains of *Drosophila* Toll and the mammalian Interleukin-1 Receptor or a CC domain (coiled-coil) in the N-terminal and are named TNL (TIR-NBS-LRR) and CNL (CC-NBS-LRR), respectively [1,3,4]. Each domain of NBS-LRR protein is predicted to have a specific function. The NBS domain is suggested to have NTP-hydrolyzing activity (ATPase or GTPase, etc), regulating signal transduction through conformational changes [4,5]. The LRR domain contains tandemly arrayed repeats in the carboxy-terminal region of R-genes and its predicted biochemical function is to mediate protein-protein interaction. It was hypothesized and

\* Correspondence: philippe.lashermes@ird.fr

<sup>1</sup>IRD - Institut de Recherche pour le Développement, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, France  
Full list of author information is available at the end of the article

experimentally confirmed that the LRR domain is involved in the specific recognition of pathogen effectors [6-9]. Both TIR and CC domains are assumed to be involved in protein-protein interactions and signal transduction [10,11].

Genes encoding NBS-LRR protein represent one of the largest and most variable gene families found in plants, with most plant genomes containing several hundred family members. NBS-LRR genes are unevenly distributed in plant genomes and are mainly organized in multi-gene clusters [2,12-16]. Furthermore, results of nucleotide polymorphism analyses demonstrated extremely high levels of inter and intraspecific variation of NBS-LRR genes, which presumably evolved rapidly in response to changes in pathogen populations [16-18]. The clustered distribution of R-genes is assumed to provide a reservoir of genetic variation from which new pathogen specificity can evolve via gene duplication, unequal crossing-over, ectopic recombination or diversifying selection [19]. However, the specific evolutionary routes of NBS-LRR encoding genes remain elusive. Several comparative sequence analyses of R-gene clusters have been performed across haplotypes or related genomes in different plant species including *Arabidopsis* [20,21], wild potato [22], tomato [23-25], *Brassicaceae* [26], wheat [27], rice [28] soybean [29] and common bean [30]. Available data suggest that different R genes can follow strikingly different evolutionary trajectories. Kuang et al. [31,32] divided NBS-LRR-genes into two evolutionary categories: Type I includes genes whose evolution is accelerated by frequent sequence exchange among paralogs. Consequently, their sequences have chimeric structure and a clear allelic/orthologous relationship between different genotypes cannot be easily established. Type II includes slowly evolving genes whose sequence mainly evolves through the accumulation of amino acid substitution. Orthology relationships are highly conserved among accessions [33].

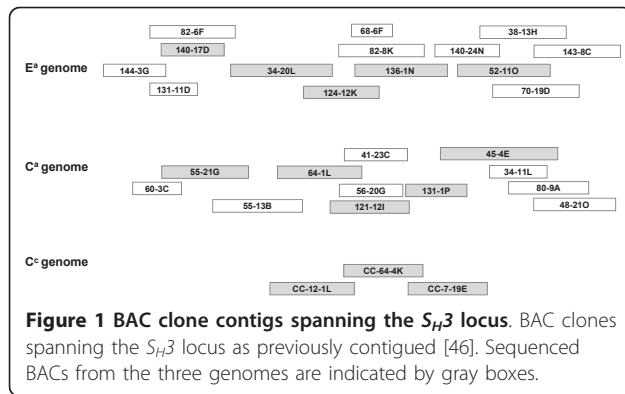
The evolutionary rate of each domain of individual NBS-LRR-encoding genes has been shown to be heterogeneous [31]. The NBS domain appears to be subject to purifying selection, whereas the LRR region tends to be highly variable [34]. Nucleotide polymorphisms found in the LRR region of R genes have been shown to be responsible for pathogen specificity [8]. In particular, codons encoding putative solvent-exposed residues in the LRR domain are hypervariable among different R proteins and show significantly elevated ratios of non-synonymous to synonymous substitutions, suggesting that the LRR domain is subject to positive selection for amino acid diversification [19,35-39].

Coffee is one of the world's most important agricultural commodities and is the main livelihood of more than 80 million people worldwide. Although the *Coffea*

subgenus *Coffea* includes more than 95 species [40], commercial coffee production relies mainly on two related species: *Coffea arabica* L. and *C. canephora* Pierre, which account for 65% and 35% of world coffee production, respectively (International Coffee Organization, <http://www.ico.org>). *C. arabica* L. ( $2n = 4x = 44$ ), the only polyploid species in the *Coffea* genus, is an allotetraploid containing two diploid subgenomes, C<sup>a</sup> and E<sup>a</sup>, which originated from two different diploid species ( $2n = 2x = 22$ ), *C. canephora* and *C. eugenioides*, respectively [41]. While diversification in the *Coffea* subgenus *Coffea* probably occurred in the second half of the Middle Pleistocene (450,000-100,000 years BP), it is most likely that the allopolyploid speciation of *C. arabica* took place in relatively recent times i.e. from historical times to 50,000 years ago [41,42].

Among the diseases affecting cultivated coffee, coffee leaf rust, caused by the obligate parasitic fungus *Hemileia vastatrix* Berk. & Br. (Uredinales), is one of the most serious diseases and greatly limits Arabica coffee production in almost all coffee growing countries around the world. Therefore, the development of coffee varieties resistant to coffee leaf rust has been a breeding objective of the highest priority in many countries [43]. A number of resistance genes to coffee leaf rust have been identified in the cultivated or wild *Coffea* gene pool. In particular, one resistance gene (i.e. *S<sub>H3</sub>* resistance factor) has been successfully introgressed from *C. liberica* into agronomically important Arabica cultivars. In the last few years, the genetic and physical maps of the *S<sub>H3</sub>* locus were completed [44-46]. Furthermore, using fluorescence *in situ* hybridization in *C. arabica*, the *S<sub>H3</sub>* locus was located in a distal position on a chromosome belonging to the homeologous group 1 [47]. Recently, a region of 800 kb spanning the *S<sub>H3</sub>* locus was sequenced and annotated [48]. Tandem arrays of CNL R genes were identified suggesting that the *S<sub>H3</sub>* locus corresponds to a complex multi-gene cluster.

The purpose of the present study was to gain insight into the genomic organization and evolution of the *S<sub>H3</sub>* R gene cluster in coffee. The agronomic importance of this locus as well as the recent origin and the perennial characteristic of coffee species make these objectives especially appealing. Sequences of the *S<sub>H3</sub>* region in three different genomes, the C<sup>a</sup> and E<sup>a</sup> subgenomes of *C. arabica* and the C<sup>c</sup> genome of *C. canephora* were analyzed to investigate the genomic organization and evolution of the *S<sub>H3</sub>* locus. In addition, we performed comparative analyses of the identified NBS-LRR encoding sequences to identify the forces that drive evolution in the *S<sub>H3</sub>* R gene cluster. Our results highlight the importance of intra and inter subgenomic gene conversion as an important evolutionary mechanism for the evolution of disease resistance genes.

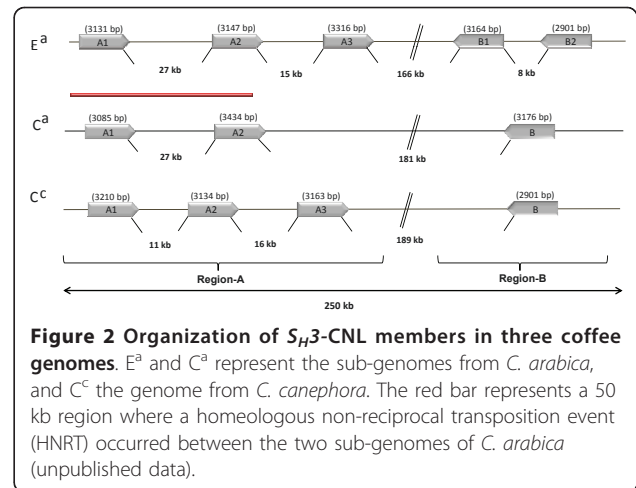


## Results

### Organization of the $S_{H3}$ R gene cluster

The sequences of a total of 13 BAC clones spanning the  $S_{H3}$  locus (Figure 1) in three coffee genomes (i.e.  $E^a$  and  $C^a$  sub-genome from *C. arabica* and  $C^c$  from *C. canephora*), were examined for the presence of R-genes using the previously determined annotation [48]. Depending on the genome concerned ( $E^a$ ,  $C^a$ ,  $C^c$ ), a total of 5, 3 and 4 R-genes were identified, respectively. These sequences shared more than 95% of identity. According to Chen et al. [49], a R gene family is defined when 60% identity is shared by members. The R-gene sequences detected in  $S_{H3}$  were therefore further considered as members of a unique R-gene coffee family. BLAST analysis of the non-redundant database (All non-redundant GenBank CDS translations + RefSeq Proteins + PDB + SwissProt + PIR + PRF) revealed high similarity with several R-genes belonging to the CNL class. While the highest identity was observed with the hypothetical CNL R protein in *Vitis vinifera* (36% identity) in *Ricinus communis* and *Glycine max* (35%), the most similar functionally characterized protein was the RPP8 gene from *Arabidopsis thaliana* that confers resistance to *Peronospora parasitica* [36] and shares 32% of identity and more than 50% of similarity. Among the species belonging to the subclass Asteridae, which includes *Coffea*, several proteins were found in *Solanum* spp. which show 30% of identity and 50% of similarity with the R-gene family found at the  $S_{H3}$  locus of *Coffea*.

The CNL-like sequences identified in the  $S_{H3}$  R-gene cluster were distributed in two regions separated by more than 160 kb (Figure 2). In the first region (hereafter called region A) two or three copies were found in the same orientation. The other region (B) contained 1 or 2 copies repeated in tandem. Region A and B had opposing coding orientations. Orthology relationship among R-genes was established by comparing flanking sequences and each member was identified by the group letter followed by a number. A homeologous non-reciprocal transposition event (HNRT) occurred between the two genomes of *C. arabica* and involved a region of around



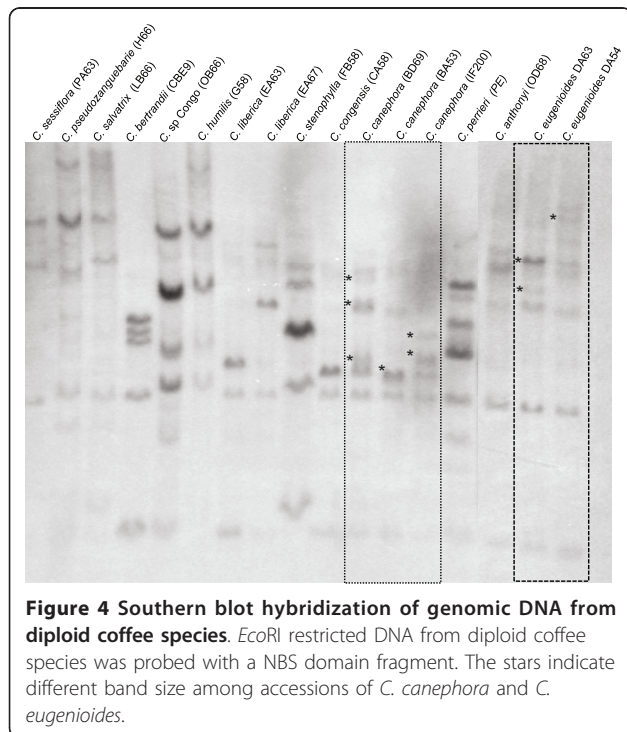
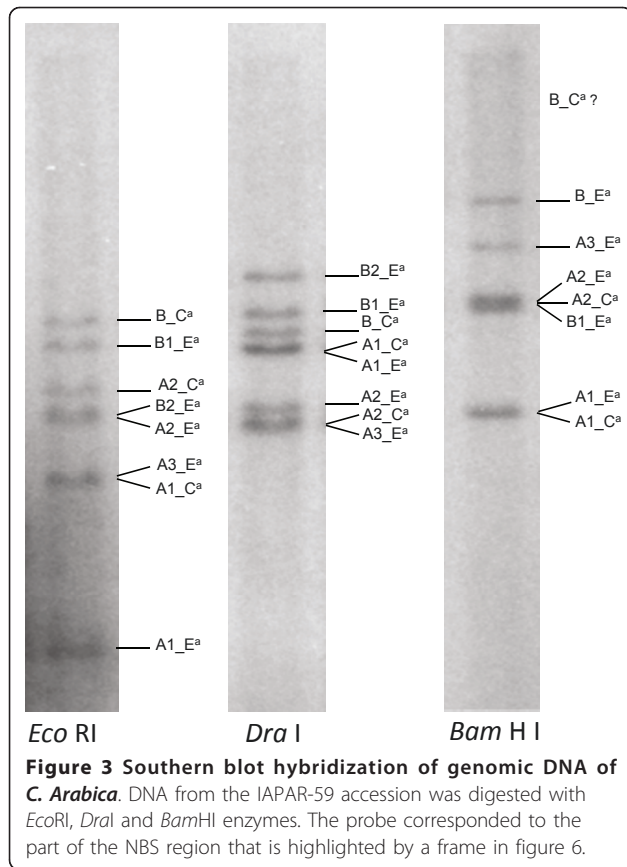
50 kb that includes the A1 and A2 members (unpublished data).

To test for the presence of a possible additional copy of the  $S_{H3}$ -CNL in the Arabica coffee genome, Southern blot analysis was performed using a specific probe corresponding to a conserved part of the NBS region (Figure 3). Whatever the restriction enzyme used, only a limited number of hybridization bands was detected. Based on the restriction profiles predicted from sequence analysis of *C. arabica* cv. IAPAR-59 BAC, it was possible to assign all the bands to one of the eight members (five in the  $E^a$  genome and three in the  $C^a$  genome) present at the  $S_{H3}$  locus. No additional band was detected, suggesting that this family is only present at the  $S_{H3}$  locus in *C. arabica* cv. IAPAR-59. In fact, even if it is possible that additional hybridization fragments have size out of the detectable range, this should have happened for all the three restriction enzymes and can be considered as a very improbable event.

The presence and number of members of this gene family in a panel of diploid coffee species were investigated by Southern blot analysis (Figure 4). While the  $S_{H3}$ -CNL family was always present, the number of members ranged from three to eight depending on the species. Intra-specific variability was also observed for different accessions of *C. canephora* and *C. eugenioides*.

### Origin and evolution of the $S_{H3}$ R gene cluster

To investigate the origin of the  $S_{H3}$ -CNL genes present at locus  $S_{H3}$  we performed comparative analysis of the available sequences of three *Coffea* genomes and among the  $S_{H3}$ -CNL copies including their flanking regions. Since members of the  $S_{H3}$ -CNL family were found to be collinear in the comparisons of the three *Coffea* genomes (Figure 5A), we concluded that the observed organization of this locus predates the divergence between *C. eugenioides* and *C. canephora* lineages. The most parsimonious



scenario for the evolution of this locus is illustrated in figure 5B. Two tandem duplications and several deletions shaped region A, whereas a distant duplication/insertion event gave birth to the  $S_{H3}$ -CNL member(s) in region B.

Locus  $S_{H3}$  was compared with the putative orthologous region in the tomato genome (*Solanum lycopersicum*) which is, to date, the closest species to *Coffea* for which whole genome sequence is available (<http://solgenomics.net>). Micro-synteny was found between the coffee  $S_{H3}$  locus and two tomato genomic regions which shared 53.2 and 23.4% of the *Coffea* genes, respectively (data not shown), but no CNL genes were found in these regions of the tomato genome.

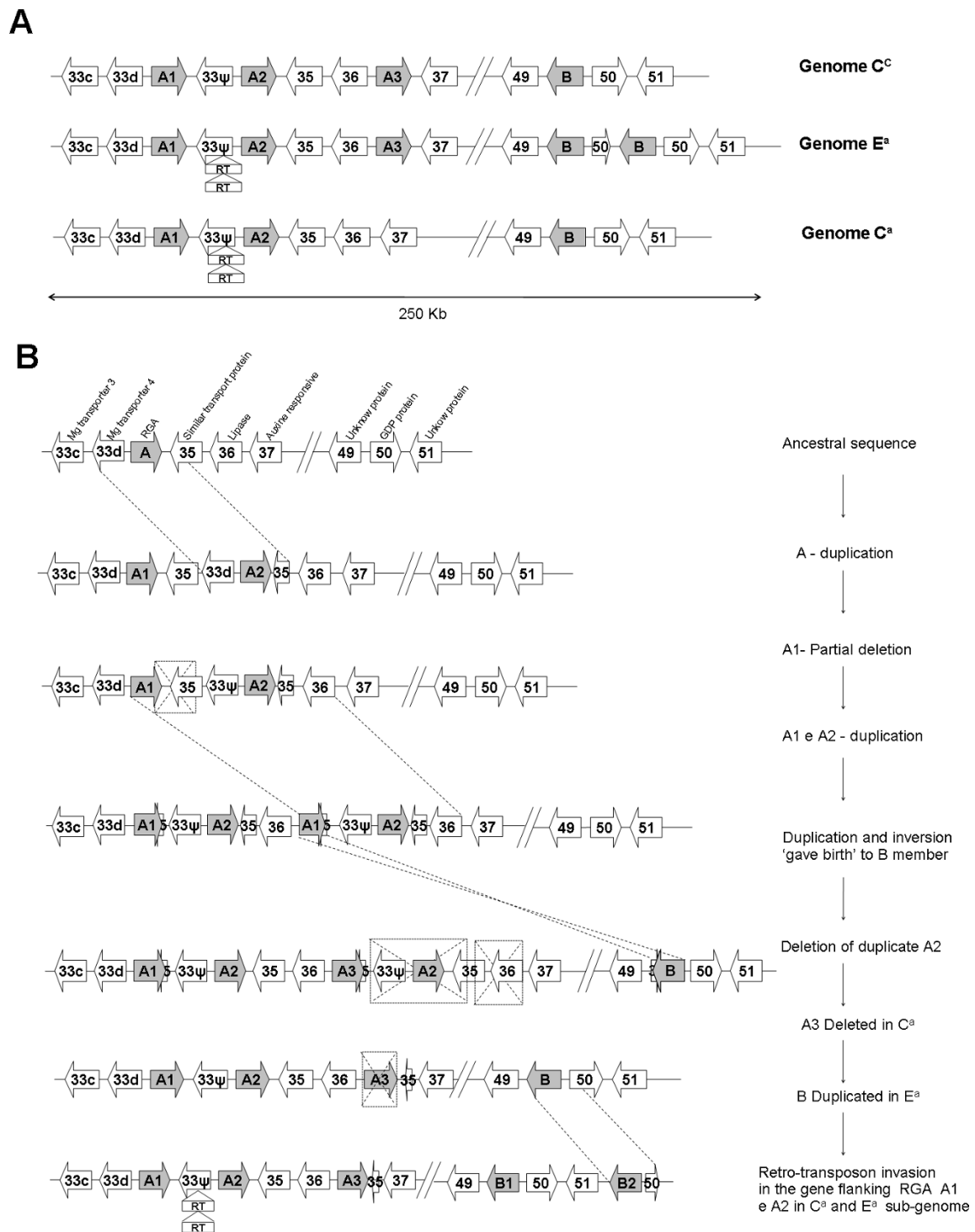
### Sequence characterization of the $S_{H3}$ -CNL family

The coding sequence of all  $S_{H3}$ -CNL members is composed of two exons separated by an intron ranging from 157 to 272 nucleotides in length. The first exon spanned 1042 nt while the second exon extended from 1703 to 2003 nt (Table 1). The protein sequence extended from 915 to 1015 aa (Table 1). The protein sequence alignment of the identified 12  $S_{H3}$ -CNL members (eight from *C. arabica* and four from *C. canephora*) is shown in figure 6.  $S_{H3}$ -CNL\_A2\_C<sup>a</sup> was chosen as query to annotate protein domains. BLASTp analysis against the Pfam database predicted a NBS domain between positions 173 and 465 aa, while analysis of the Conserved Domain Database predicted the beginning of the LRR region at position 625 aa of the query protein. COILS analysis revealed a coiled-coil region located between position 17 and 56 aa, confirming that this family belongs to the CC sub-family of NBS-LRR genes (or non-TIR sub-family). The LRR region of all genes consists of 12 repeats ranging from 23 to 31 aa. These repeats are sufficiently different to ensure an unambiguous alignment of amino-acid sequences. A 8 bp deletions modified the reading frame of B2\_E<sup>a</sup> and induced an early stop codon after the 10th LRR; similarly, an 1 bp insertion in the A2\_E<sup>a</sup> made this member a pseudogene. Both INDEL modifying the reading frame were disregarded in figure 6 and in the following analyses.

### Cloning of $S_{H3}$ -CNL\_A2 members from diploid species of coffee

To study interspecific diversity, the SH3-CNL\_A2 member was selected at random for further analysis. The  $S_{H3}$ -CNL\_A2 member was cloned from six coffee species (*C. anthonyi*, *C. sp. Congo*, *C. canephora*, *C. eugenioides*, *C. liberica*, *C. pseudozanguebarie*). The cloned fragments were around 4 kb in size. Their sequences were determined and compared with those from C<sup>a</sup>, E<sup>a</sup> and C<sup>c</sup> genomes.





**Figure 5 Evolution of the  $S_H3$  locus in coffee species.** **A.** Current organization of the  $S_H3$  locus in *Coffea canephora* (C<sup>c</sup>) and *C. arabica* (sub-genome E<sup>a</sup> and sub-genome C<sup>a</sup>). **B** - A model of the evolution of locus  $S_H3$  in coffee plants involving genome expansion and retraction by gene duplication and deletions. Gray arrows indicate members of the  $S_H3$  family. Open arrows indicate other non-R genes flanking R genes in the locus as numbered in [48]. Short arrows indicate truncated versions of corresponding genes.

**Table 1 Exon, intron size (bp) and protein size (aa) of the *S<sub>H</sub>3*-CNL members identified in the three genomes analyzed**

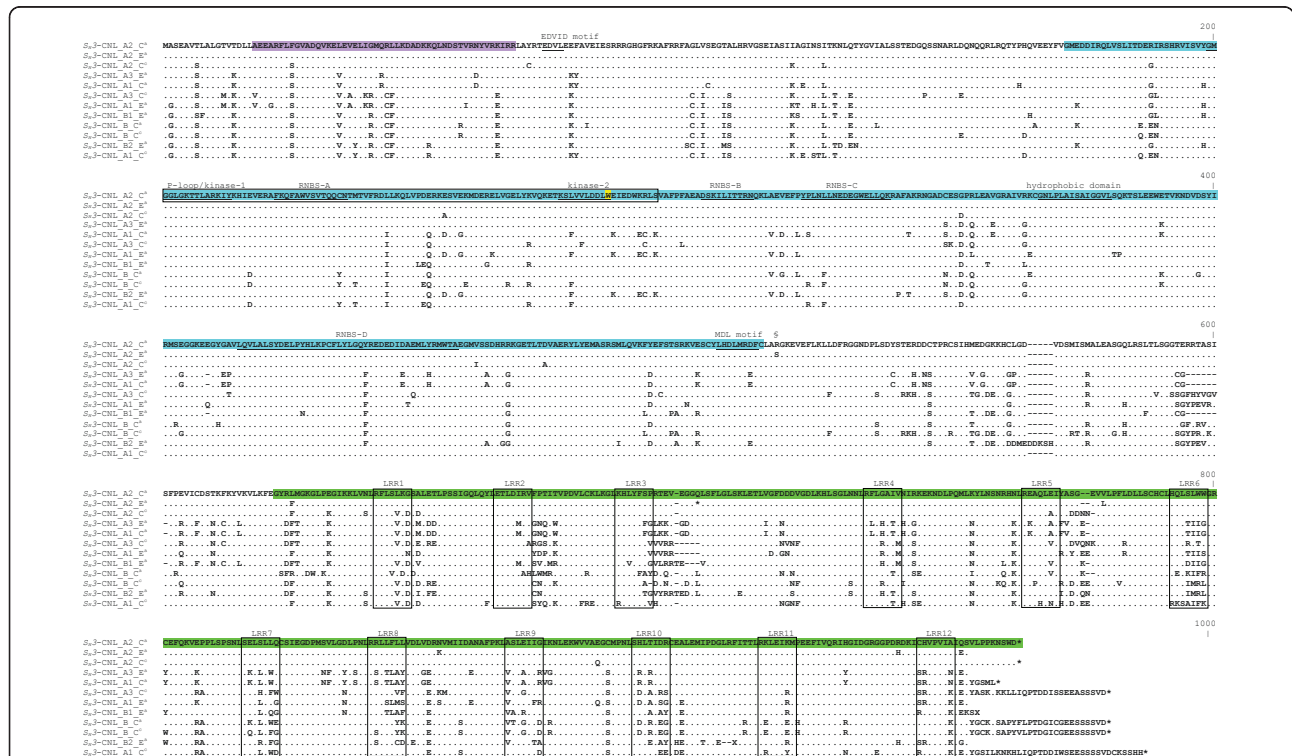
Copy	genome	Exon 1	Intron	Exon 2	Protein (aa)
A <sub>1</sub>	E <sup>a</sup>	1042	272	1787	943
A <sub>1</sub>	C <sup>a</sup>	1042	255	1787	943
A <sub>1</sub>	C <sup>c</sup>	1042	270	1898	980
A <sub>2</sub>	E <sup>a</sup>	1042	270	1908	955
A <sub>2</sub>	C <sup>a</sup>	1042	270	1820	954
A <sub>2</sub>	C <sup>c</sup>	1042	269	1823	954
A <sub>3</sub>	E <sup>a</sup>	1042	268	2003	1015
A <sub>3</sub>	C <sup>c</sup>	1042	256	1865	969
B <sub>1</sub>	E <sup>a</sup>	1042	258	1865	969
B <sub>2</sub>	E <sup>a</sup>	1042	157	1703	915
B	C <sup>a</sup>	1042	260	1895	970
B	C <sup>c</sup>	1042	260	1874	972

**Sequence diversity analysis of the *S<sub>H</sub>3*-CNL family**

Using the RDP3 software [50] and regardless of the method used for the analysis, significant traces of gene conversion were detected among the member of the *S<sub>H</sub>3*-CNL family, both in *C. arabica* and *C. canephora*. As an example, the conversions detected with the RDP

method were reported in Table 2. Among the nine different gene conversions detected, two events involved inter subgenomic exchanges.

The DNA sp program (v.5) was used to estimate polymorphism among the four *S<sub>H</sub>3*-CNL members in the genome of *C. canephora* species (C<sup>c</sup>). The highest level



**Figure 6 Alignment of the predicted amino acid sequences from *S<sub>H</sub>3*-CNL members.** The coiled-coil, NBS and LRR domains are highlighted in lilac, blue and green, respectively. The motif EDVID [79] as well as the motifs P-loop/kinase 1, RNBS-A, kinase II, RNBS-B, RNBS-C, hydrophobic domain in NBS domain are underlined. The first sequence is shown in full, while for other proteins only amino acids that differ from the first one are indicated. A 8 bp deletions in B2\_E<sup>a</sup> and an 1 bp insertion in A2\_E<sup>a</sup> modifying the reading frame were disregarded. The xxLxLxx motif in the LRR domain is boxed, where L is any aliphatic amino acid and x is any amino acid. Gaps introduced at alignment are indicated by dashes, while asterisks indicate the presence of stop codons. NBS probe used in Southern hybridization is highlighted by a frame. The tryptophan residue (W), specific to the non-TIR-NBS-LRR class of plant disease R gene, located at the end of the kinase 2 motif [20], is highlighted in yellow.

**Table 2 Gene conversions detected among  $S_{H3}$ -CNL members with the RDP method [50]**

Sub-genome analyzed	$S_{H3}$ -CNL members	pValue	Begin	End	Length (Nc)
$C^c$	A1_ $C^c$ × A2_ $C^c$	$2.72 \times 10^{-14}$	992	2249	1258
$C^c$	A1_ $C^c$ × B_ $C^c$	$6.57 \times 10^{-3}$	1	977	977
$C^a$	A2_ $C^a$ × A1_ $C^a$	$2.13 \times 10^{-3}$	2850	3049	200
$C^a$	A2_ $C^a$ × A1_ $C^a$	$4.15 \times 10^{-3}$	1	579	579
$E^a$	A1_ $E^a$ × A2_ $E^a$	$9.46 \times 10^{-8}$	1	1335	1335
$E^a$	A3_ $E^a$ × B1_ $E^a$	$7.63 \times 10^{-4}$	218	1584	1367
$E^a$	A2_ $E^a$ × A1_ $E^a$	$7.62 \times 10^{-3}$	2916	3054	139
$C^a + E^a$	B2_ $E^a$ × A1_ $C^a$	$8.99 \times 10^{-14}$	429	1232	804
$C^a + E^a$	A1_ $E^a$ × A2_ $C^a$	$7.82 \times 10^{-8}$	105	1335	1231

p Value (Bonferroni-corrected Karlin-Altschul); Begin, first nucleotide of the potential converted region; End, last nucleotide of the potential converted region; Length, length of the converted region (Nc).

of DNA polymorphism was detected in the LRR domain ( $\pi = 0.17, 0.20$  and  $0.15$ ) while the most conserved regions were in the NBS domain, especially in the P-loop, Kinase 2 and hydrophobic domains (Figure 7).

To check the type of selection that acted on genes in the  $S_{H3}$ -CNL family, the ratio between non-synonymous (Ka) and synonymous substitutions (Ks) was estimated using DnaSP v.5. The Ka/Ks substitution rate was calculated for each pair between ortholog and/or paralog members in *C. arabica* and *C. canephora* species. We also calculated the Ka/Ks between each pair of A2 members cloned from diploid coffee species together with A2 members from sequenced genomes (*C. arabica* and *C. canephora*). The analysis was performed on the complete coding sequence as well as on different domains (CC, NBS, LRR). Analysis also focused on codons encoding the solvent-exposed amino acids of the

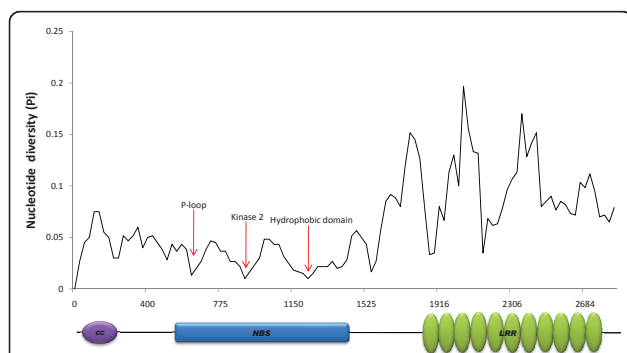
$\beta$ -strand/ $\beta$ -turn motifs (x residues in xxLxLxx motifs) in the LRR domains.

Strong evidence for positive selection (Ka/Ks > 1) was observed only for residues in xxLxLxx motifs. Among the 66 pair combinations between 12 BAC derived R-genes, 19, of which 16 involving B members, appeared under positive selection (Figure 8). When the same region was analyzed among orthologous A2 members, no significant Ka/Ks > 1 was found.

## Discussion

### Organization and evolution of the $S_{H3}$ -CNL family

Comparative analyses of R-gene clusters across different haplotypes or species demonstrated that the evolution of resistant genes is a dynamic process mostly involving duplication, deletion, sequence exchange, point mutation, diversified selection, recombination, gene conversion and retroelement insertion [21,23,24,28-30,32,33,51,52]. The cluster arrangement of R-genes represents an important reservoir of diversity and a source of genetic variation allowing the generation of novel resistance specificities via gene conversion, gene duplication, unequal crossing-over, ectopic recombination or diversifying selection [12,19]. To explore the organization and to characterize the mechanisms involved in the evolution of the  $S_{H3}$  locus, where a putative R-gene cluster was identified [48] a ~550 kb sequence was analyzed in three coffee genomes,  $E^a$  and  $C^a$  from *C. arabica* and  $C^c$  from *C. canephora*. Sequence analysis revealed the presence of a variable number of NBS-LRR genes belonging to the CC subclass at the  $S_{H3}$  locus. All these genes belong to the same family (hereafter called  $S_{H3}$ -CNL family). Sequence analysis of regions flanking the  $S_{H3}$ -CNL genes helped determine the orthology relationship among the copies in different genomes. At the same time, several traces of ancient duplications made it possible to trace back the duplication/deletion events which, consistently with the birth and death evolution model, shaped the  $S_{H3}$  locus from the most recent common ancestor of all  $S_{H3}$ -CNL



**Figure 7 Nucleotide diversity among  $S_{H3}$ -CNL members from *C. canephora*.** Nucleotide diversity ( $\pi$ ) is the average number of nucleotide differences per site between two sequences calculated by DnaSP v.5. Nucleotide diversity was calculated using the sliding window method where a window (segment of DNA) is moved along the sequences step by step. The parameter is calculated in each window, and the value is assigned to the nucleotide at the midpoint of the window. Both the default values were used: window length of 100 sites, and step size of 25 sites (midpoint). The alignment gaps were not counted in the window length (or slide).

Ka/Ks	P												
	A1_E <sup>a</sup>	A1_C <sup>a</sup>	A1_C <sup>c</sup>	A2_E <sup>a</sup>	A2_C <sup>a</sup>	A2_C <sup>c</sup>	A3_E <sup>a</sup>	A3_C <sup>c</sup>	B1_E <sup>a</sup>	B_C <sup>a</sup>	B_C <sup>c</sup>	B2_E <sup>a</sup>	
A1_E <sup>a</sup>		0.0577	0.0982	0.0838	0.0838	0.0405	0.0577	0.0606	0.0696	0.0097	0.0275	0.0383	
A1_C <sup>c</sup>	3.09		0.0448	0.2066	0.2066	0.1503	N/A	0.1033	0.0139	0.0157	0.0080	0.0362	
A1_C <sup>c</sup>	1.09	3.02*		0.2430	0.2430	0.1963	0.0448	0.0517	0.1212	0.1800	0.0987	0.0613	
A2_E <sup>a</sup>	2.65	1.89	1.86		N/A	0.6055	0.2066	0.2294	0.0988	0.0950	0.2670	0.0503	
A2_C <sup>a</sup>	2.65	1.89	1.86	N/A		0.6055	0.2066	0.2294	0.0988	0.0950	0.2670	0.0503	
A2_C <sup>c</sup>	3.52*	2.16	2.03	0.53	0.53		0.1503	0.1392	0.0563	0.0530	0.1487	0.0214	
A3_E <sup>a</sup>	3.09	N/A	3.02*	1.89	1.89	2.16		0.1033	0.0139	0.0157	0.0080	0.0362	
A3_C <sup>c</sup>	1.30	2.46	0.73	1.95	1.95	2.46	2.46		0.3088	0.0610	0.1039	0.1541	
B1_E <sup>a</sup>	3.82	6.95*	2.25	2.49	2.49	3.14	6.95*	1.80		0.0203	0.0156	0.0475	
B_C <sup>a</sup>	5.75**	5.28*	1.38	2.57	2.57	3.30	5.28*	3.19	4.55*		N/A	0.0076	
B_C <sup>c</sup>	4.38*	7.16**	2.45	1.82	1.82	2.31	7.16**	2.75	5.66*	N/A		0.0652	
B2_E <sup>a</sup>	4.03*	3.65*	0.78	3.45	3.45	5.26*	3.65*	2.38	4.09*	7.49**	4.55		

**Figure 8** Ka/Ks ratio in the solvent-exposed residue of *S<sub>H</sub>3*-CNL members. The Ka/Ks ratio was calculated in the solvent-exposed residue of the LRR domain by pairwise comparison of *S<sub>H</sub>3*-CNL members. Values below the diagonal are the Ka/Ks ratio and values above are the probability, significance level for Ka > Ks indicated by \* P < 0.05, \*\* P < 0.01. N/A = not applicable.

copies. Since the structure of the *S<sub>H</sub>3* locus was well conserved in all three *Coffea* genomes analyzed, one can conclude that the origin of most of the *S<sub>H</sub>3*-CNL copies predates the divergence between *Coffea* species.

Homologs of *S<sub>H</sub>3*-CNL genes were found in several dicot species including *Solanum* spp., but comparative genomics failed to find a CNL R-gene in the orthologous regions of three Rosid species [48] and *Solanum lycopersicum*. Consequently, it can be suggested that the ancestral *S<sub>H</sub>3*-CNL copy was inserted in the *S<sub>H</sub>3* locus after the divergence between *Solanum* and *Coffea* lineages.

In a similar study by David et al. [30] the authors compared the B4 locus of *Phaseolus vulgaris* (that has 26 CNL genes) with three sequenced legume genomes, *Medicago trunculata* (*Mt*), *Lotus japonicus* (*Lj*) and *Glycine max* (*Gm*). Their analysis revealed that conserved microsynteny existed among legumes species, except for the CNL sequences, which appeared to be completely absent in the corresponding regions of *Mt* and *Lj* and only a truncated CNL was found in syntenic regions of *Gm* [30]. They suggested that CNL were inserted in the ancestor of the B4 locus after *Mt*, *Lj*, and *Pv* diverged but before the divergence of *Pv* and *Gm* through an ectopic recombination event between non-homologous chromosomes. Phylogenetic analysis including those legume CNL sequences and all known *Pv* B4-CNL sequences supported this hypothesis [30].

Structural instability induced by repetitive mobile elements is one of the mechanisms that could lead to diversification into R gene families. The presence of very similar sequences increases the possibilities of mispairing during recombination, giving rise to unequal crossovers and interlocus gene conversions [19,53-55]. However, the edges of the duplications involved in the birth of new *S<sub>H</sub>3*-CNL copies were not related to mobile elements and mobile elements identified in the region did not appear

to play a role in the structural evolution of the *S<sub>H</sub>3* locus of *Coffea* species.

Gene conversion (i.e. the substitution of a portion of a gene sequence by the homologous sequence of another related gene) is more frequent among members of highly similar, tightly clustered families [56]. Gene conversion is a common phenomenon and it has been detected between paralogs in many R-gene clusters [9,21,35-37, 57-62]. Sequence exchanges between different sub-genomes have previously been detected in a R1 resistance-gene cluster of one CNL subfamily in allohexaploid, *Solanum demissum* [22]. In that study, seven of the 17 sequence exchanges among R1 homologs occurred between different genomes. Two explanations were proposed: first, sequence exchanges among different haplotypes could be generated through gene conversions or alternatively, they might have occurred through recombination before speciation and have been conserved in *S. demissum* [22]. At the *S<sub>H</sub>3* locus, gene conversion events were detected between paralogs in all three coffee genomes analyzed and also between members of the two sub-genomes of *C. arabica*.

Conversion events were detected between *S<sub>H</sub>3*-CNL members independently of their orientation (i.e. between members in region A and B). The inverted orientation of the loci might allow rare interlocus gene conversion or unequal exchange while minimizing the risk of gross chromosomal rearrangement [23]. The gene orientation in a cluster of NBS-LRR has also been studied in rice [12] and *A. thaliana* [56]. These studies demonstrated that conversion can occur between genes in the same or in opposite orientation, however conversion is more frequent in gene families arranged as direct repeats because they have higher similarity than those found in opposite orientation.

In addition, gene conversion was suggested to be more frequent in perennial than in annual plants. Yang et al.[18], compared the gene conversion events among



NBS-encoding genes in two perennial and two annual plants. A total of 823 and 468 gene conversion events involving 299 and 187 NBS-encoding genes were detected in grapevine and poplar, respectively, while only 143 and 81 gene conversion events were detected in *Arabidopsis* and rice, respectively [18]. Since the long-generation time of woody species slows down the accumulation of evolutionary change, the authors suggested that an excess of recent duplications and a higher conversion rate in grapevine and poplar could generate novel resistance profiles to compensate for life history traits. According to Kuang et al. [31,32]  $S_{H3}$ -CNL members should be classified as evolutionary type I (fast evolving genes) since several conversion events were detected between members.

#### Effect of selection on molecular evolution of the $S_{H3}$ -CNL family

Natural selection influences the molecular evolution of sequences by increasing or reducing the fixation probability of a given mutation which, respectively, increases or reduces the fitness of the individuals carrying it.

The effect of natural selection on a gene sequence can be investigated by analyzing nucleotide substitutions that occurred between two variants of this gene. Since synonymous substitutions (i.e. nucleotide substitutions that do not change the amino acid sequence) are supposed to not modify the phenotype, their accumulation is considered not to be influenced by natural selection. Conversely, non-synonymous substitutions (nucleotide substitutions that modify the coded amino acid) could increase, reduce, or not influence the fitness of the individuals carrying it; consequently, their accumulation could be influenced by natural selection. The ratio of non-synonymous ( $K_a$ ) to synonymous ( $K_s$ ) substitution rates could be used to infer the effect of natural selection of a given gene or a part of it. When  $K_a$  and  $K_s$  have similar values ( $K_a/K_s \approx 1$ ), one could infer a neutral effect of selection; when  $K_a$  is significantly lower than  $K_s$  ( $0 < K_a/K_s < 1$ ), it could be deduced that the selection purges the gene sequence of most non-synonymous substitutions (purifying selection); finally, when  $K_a$  is significantly higher than  $K_s$  ( $K_a/K_s > 1$ ), the selection is assumed to favor fixation of new variants (positive or diversifying selection) [63,64].

In many NBS-LRR genes, analysis of corresponding proteins revealed high non-synonymous:synonymous substitution ratios in the leucine-rich (LRR) domain, mainly concentrated on the putative solvent-exposed residues, indicating that the LRR domain is subject to positive selection for amino acid diversification, [19,35,36,38,39,59,60,65-68]. These results are consistent with the observation that nucleotide polymorphisms

found in the leucine-rich (LRR) region of  $R$  genes are often responsible for pathogen specificity [66].

In the  $S_{H3}$ -CNL family, significant positive selection was only detected when the  $K_a/K_s$  analysis was focused on solvent-exposed residues (i.e. the x residues in xxLxLxx motif from LRR domain) most frequently among paralog members. Conversely, when larger regions were considered, the effect of natural selection was diluted and not detectable.

In the co-evolutionary arms race between hosts and their pathogens, genes involved in their interaction are expected to evolve under positive selection. The positive selection detected in the solvent-exposed residue of the  $S_{H3}$ -CNL members could indicate involvement in recognition of pathogen attack.

#### Conclusions

The  $S_{H3}$ -CNL family appears to have evolved following the birth-and-death model, since duplications and deletions were inferred in the evolution of the  $S_{H3}$  locus. Gene conversion between paralog members from the same or different sub-genomes, and positive selection appear to be the major forces influencing the evolution of  $S_{H3}$ -CNL in coffee trees.

#### Materials and methods

##### Plant material and DNA extraction

The cv. IAPAR 59 of *Coffea arabica* and six *Coffea* species were analyzed in this study: *C. canephora* (IF200), *C. anthonyi* (OD68), *C. sp.* Congo (OB66), *C. eugenioides* (DA54), *C. liberica* (EA67), *C. pseudozanguebarie* (H66). Genomic DNA was isolated from leaves of greenhouse grown plants located at IRD (*Institut de Recherche pour le Développement*) Montpellier, France. Leaves were frozen in liquid nitrogen and DNA was extracted using a CTAB procedure [69] with modified extraction buffer (3% CTAB, 1.4 mM NaCl, 100 mM Tris HCl, 20 mM EDTA, pH 0.8).

##### BAC sequences

Several Bacterial Artificial Chromosome (BAC) clones spanning the  $S_{H3}$  locus were isolated from a *C. arabica* (IAPAR59) [70] and a *C. canephora* (HD-200-94) (unpublished data) libraries. Based on fingerprint data and overlapping sequence analysis, BAC sequences were assembled in contigs specific to the *C. arabica* and *C. canephora* genomes (hereafter called  $C^c$  for *C. canephora* genome;  $E^a$  and  $C^a$  for “eugenioides” and “canephora” sub-genomes of *C. arabica*) (Lashermes et al. 2010). Gene annotation of the BACs was already available [48]. Sequences of the thirteen selected BACs were deposited to GeneBank [accession numbers, Genebank:GU123894 to GU123899 and HQ696507 to HQ696513].

### Primer design and cloning procedure

Orthologous specific primers to amplify A2 members of the  $S_H3$ -CNL family from wild *Coffea* species were designed based on sequence alignments of A2 members in *C. canephora* and *C. arabica*: A2\_Left: 5'-CCTTGATAA-GAAACATGAATGAAATACACGA-3' and A2\_right 5'-AAGGATAAATGAGAAGAACTACTGAGCCTG-3'. DNA amplification was performed with Expand™ 20Kb<sup>plus</sup> PCR System (Roche Applied Science, Mannheim Germany). PCR were performed as follows: one cycle of 1 min at 95°C, 10 cycles of 10 sec at 94°C, 45 sec at 50°C, 5 min at 68°C followed by 20 cycles of 10 sec at 94°C, 45 sec at 50°C, 7 min at 68°C plus 10 sec per cycle, and final extension of 7 min at 68°C. A10 µl aliquot from each PCR amplification was analyzed by electrophoresis in a 1.2% agarose gel. The amplicons were gel-stained using Crystal violet and the DNA bands were purified using a S. N.A. P™ purification column (Invitrogen Carlsbad, CA). The PCR products were cloned into the pCR®-XL-TOPO® kit from Invitrogen and chemically competent cells (Invitrogen Carlsbad, CA) according to the manufacturer's protocol. Eight colonies were randomly selected for screening. Colony PCR of eight random samples was used to select clones containing the complete amplicon. For this purpose, two new primers were designed to amplify the extremities of the genes in combination with the primers used to amplify the whole gene: 5'-CGACAGTGG-GAACGAAACCC-3' combined with A2\_Left and 5'-TGGAGGACCGGATCATGAACA-3' combined with de A2\_RIGHT. The colony PCR was performed as follows: 10 min at 94°C, followed by 30 cycles of 30 sec at 94°C, 30 sec at 55°C, 4 min at 72°C and final extension of 10 min at 72°C. The colonies shown to contain the complete insert were transferred to 5 ml LB broth with 50 µg/ml kanamycin and incubated at 37°C overnight. Plasmid DNA was isolated using Promega Wizard® Plus Minipreps DNA purification System (Promega Corporation, Madison, WI, USA) according to manufacturer's instructions. Two independent PCR and sequencing were carried out to ensure quality.

### Sequencing and analysis of cloned $S_H3$ -CNL members

Plasmid DNA was sequenced at Genome Express (Grenoble, France) using M13-universal- forward and reverse primers and five other internal primers were designed using the Primer3 program (Whitehead Institute, USA) to allow whole gene sequencing. INT1-L: 5-TCCATCGTCCAA-GATACAGC-3, INT2-L: 5-TTTGTTGGGATGGAAG ATGA-3, INT3-L: 5-GCTGGGAGTTGCTTCAAAG-3, INT4-L: 5-TCGAATGTGGACAGCAGAAG-3, INT5-L: 5-GCCTTGGAGACACTTCCATC-3. The cloned sequence contigs were assembled using the Staden package [71]. The complete sequences of each clone were aligned using Bioedit v.7.0 [72].

### Southern blot analysis

Southern blot analysis was performed as follows: 20 µg of genomic DNA was extracted as described above and digested with a restriction enzyme (only *EcoRI* for the panel of diploid species *EcoRI*, *DraI* and *BamHI* for the *Coffea arabica* cv. IAPAR-59) and separated by agarose gel electrophoresis. The digested DNA was transferred to Hybond-N+ nylon membranes for Southern hybridization as described in Noir et al. [70].  $S_H3$ -CNL family specific probe was obtained by PCR amplification using primers designed on the NBS domain (left primer: 5'-CGGTCTCGGTAAGACCACTC-3' and right primer 5'-CCTCTGCAAATGGAAATGCT-3'). The amplified 516 bp fragment was labeled with [32P]-dATP according to the manufacturer's recommendations (Megaprime DNA Labelling Systems kit, Amersham) and used as probe in the hybridization experiment as described in Sambrook et al. [73].

### Motif predictions

SMART protein motif analyses (<http://smart.embl-heidelberg.de>) and Pfam database (<http://pfam.sanger.ac.uk/search/sequence>) were used to detect motifs in the  $S_H3$ -CNL genes. COILS with a threshold of 0.9 was used to specifically detect CC domains [74].

### Gene Conversion

In order to check the possibility of conversion events among the the  $S_H3$ -CNL members, alignments of sequences from *C. arabica* and *C. canephora* species were analyzed with the RDP3 software [50] using the default settings (but linear instead of circular sequences were selected, in general settings). The program uses simultaneously different recombination detection methods, including RDP and GeneConv [75], to both detect and characterize the recombination events that are evident within a sequence alignment without any prior user indication of a non-recombinant set of reference sequences [50]. Pairwise P values are assigned based on the comparison of each fragment with the maximum fragment length that is expected from the sequence pair by chance.

### Sequence evolution

Protein sequences were manually aligned with the BioEdit program. The amino acid sequence alignments were used to guide the alignments of nucleotides using MEGA version 4.1 [76]. Nucleotide diversity ( $\pi$ ) was calculated by DnaSP v.5 [77] where each paralog was considered as an independent allele of population.

The  $Ka/Ks$  ratio was estimated by DnaSP v5.1 based on Nei and Gojobori's equation [78] for full-length CDS for specific domains: (CC, NBS, LRR), for the xxLxLxx motifs in the LRR domain, and for the solvent-exposed residues (i.e. only the x residues in the xxLxLxx motif).

P values were calculated and the significance level was compared at 0.05 and 0.01%.

### Microsynteny Analysis

The search for putative gene orthologs of the genes at the *S<sub>H3</sub>* locus was performed by TBLASTN analysis on the tomato genome sequence (*Solanum lycopersicum*) available in Solanaceae Genome Network (<http://www.sgn.cornell.edu>).

### Acknowledgements

This study was financially supported by a grant to A. Ribas of CAPES foundation from the Ministry of Education of Brazil and by CIRAD and IRD funds for doctoral support.

### Author details

<sup>1</sup>IRD - Institut de Recherche pour le Développement, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, France.  
<sup>2</sup>CIRAD-BIOS - Centre de Coopération Internationale en Recherche Agronomique pour le Développement - Département des Systèmes Biologiques, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, France.

### Authors' contributions

AFR carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. AC carried out the genomic analyses and helped to draft the manuscript. MC participated in the sequence alignment. HE participated in the design of the study. PL designed the study, and participated in its coordination. All authors read and approved the final manuscript.

Received: 14 December 2010 Accepted: 16 May 2011

Published: 16 May 2011

### References

- Jones JDG, Dangl JL: The plant immune system. *Nature* 2006, **444**(7117):323-329.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND: Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* 1999, **20**:317-332.
- Dangl JL, Jones JDG: Plant pathogens and integrated defence responses to infection. *Nature* 2001, **411**(6839):826-833.
- Martin GB, Bogdanove AJ, Sessa G: Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology* 2003, **54**(1):23-61.
- Leipe DD, Koonin EV, Aravind L: STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. 2004, **343**:1-28.
- Jones D, Jones J: The role of leucine-rich repeat proteins in plant defenses. *Adv Bot Res* 1997, **24**:89-167.
- Hammond-Kosack KE, Jones JDG: Plant Disease Resistance Genes. *Annu Rev Plant Physiol Plant Mol Biol* 1997, **48**:575-607.
- Ellis J, Dodds P, Pryor T: Structure, function and evolution of plant disease resistance genes. *Current Opinion in Plant Biology* 2000, **278**-284.
- Van der Hoorn RAL, Roth R, De Wit PJGM: Identification of distinct specificity determinants in resistance protein Cf-4 allows construction of a Cf-9 mutant that confers recognition of avirulence protein AVR4. *Plant Cell* 2001, **13**:273-285.
- Jebanathirajah JA, Peri S, Pandey A: Toll and interleukin-1 receptor (TIR) domain-containing proteins in plants: a genomic perspective. *TRENDS in Plant Science* 2002, **7**(9):388-391.
- McDowell JM, Woffenden BJ: Plant disease resistance genes: recent insights and potential applications. *TRENDS in Biotechnology* 2003, **21**(4):178-183.
- Hulbert SH, Webb CA, Smith SM, Sun Q: Resistance gene complexes: Evolution and Utilization. *Annual Review of Phytopathology* 2001, **39**(1):285-312.
- Richly E, Kurth J, Leister D: Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol Biol Evol* 2002, **19**:76-84.
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D: Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* 2004, **271**(4):402-415.
- Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu HY, Roe B, Young ND, Cannon SB: Identification and characterization of nucleotide-binding site-Leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiology* 2008, **146**(1):5-21.
- Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q: Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics* 2008, **280**(3):187-198.
- Meyers BC, Kaushik S, Nandety RS: Evolving disease resistance genes. *Current Opinion in Plant Biology* 2005, **8**(2):129-134.
- Yang S, Zhang X, Yue J, Tian D, Chen J: Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics* 2008, **280**(3):187-198.
- Michelmore RW, Meyers BC: Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* 1998, **8**:1113-1130.
- Meyers BC, Kozik A, Griego A, Kuang HH, Michelmore RW: Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 2003, **15**(4):809-834.
- Noel L: Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus in *Arabidopsis*. *Plant Cell* 1999, **11**:2099-2111.
- Kuang H, Wei F, Marano M, Wirtz U, Wang X, Liu J, Shum W, Zaborosky J, Tallon L, Rensink W, et al: The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. *Plant J* 2005, **44**(1):37-51.
- Parniske M, Jones JD: Recombination between diverged clusters of the tomato Cf-9 plant disease resistance gene family. *Proc Natl Acad Sci USA* 1999, **96**(10):5850-5855.
- Kruijt M: Rearrangements in the Cf-9 Disease resistance gene cluster of wild tomato have resulted in three genes that mediate Avr9 responsiveness. *Genetics* 2004, **168**(3):1655-1663.
- Seah S, Telleen AC, Williamson VM: Introgressed and endogenous *Mi-1* gene clusters in tomato differ by complex rearrangements in flanking sequences and show sequence exchange and diversifying selection among homologues. *Theoretical and Applied Genetics* 2007, **114**(7):1289-1302.
- Xiao SY, Emerson B, Ratanasut K, Patrick E, O'Neill C, Bancroft I, Turner JG: Origin and maintenance of a broad-spectrum disease resistance locus in *Arabidopsis*. *Mol Biol Evol* 2004, **21**(9):1661-1672.
- Wicker T, Yahiaoui N, Keller B: Contrasting rates of evolution in Pm3 loci from three wheat species and rice. *Genetics* 2007, **177**(2):1207-1216.
- Dai L, Wu J, Li X, Wang X, Liu X, Jantasuriyarat C, Kudrna D, Yu Y, Wing RA, Han B, et al: Genomic structure and evolution of the Pi2/9 locus in wild rice species. *Theoretical and Applied Genetics* 2010.
- Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NWG, Couloux A, Dalwani A, Denny R, et al: Differential Accumulation of Retroelements and Diversification of NB-LRR Disease Resistance Genes in Duplicated Regions following Polyploidy in the Ancestor of Soybean. *Plant Physiology* 2008, **148**(4):1740-1759.
- David P, Chen NWG, Pedrosa-Harand A, Thareau V, Sevignac M, Cannon SB, Deboucq D, Langin T, Geffroy V: A Nomadic Subtelomeric Disease Resistance Gene Cluster in Common Bean. *Plant Physiology* 2009, **151**(3):1048-1065.
- Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW: Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 2004, **16**(11):2870-2894.
- Kuang HH, Wei FS, Marano MR, Wirtz U, Wang XX, Liu J, Shum WP, Zaborosky J, Tallon LJ, Rensink W, et al: The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues



- and shows substantial structural variation among haplotypes of *Solanum demissum*. *Plant Journal* 2005, **44**(1):37-51.
33. Kuang H, Caldwell KS, Meyers BC, Michelmore RW: Frequent sequence exchanges between homologs of RPP8 in *Arabidopsis* are not necessarily associated with genomic proximity. *Plant Journal* 2008, **54**(1):69-80.
  34. McHale L, Tan X, Koehl P, Michelmore RW: Plant NBS-LRR proteins: adaptable guards. *Genome Biol* 2006, **7**(4):212.
  35. Parniske M: Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. *Cell* 1997, **91**:821-832.
  36. McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL: Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. *Plant Cell* 1998, **10**(11):1861-1874.
  37. Ellis JG, Lawrence GJ, Luck JE, Dodds PN: Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* 1999, **11**:495-506.
  38. Shen KA, Chin DB, Arroyo-Garcia R, Ochoa OE, Lavelle DO, Wroblewski T, Meyers BC, Michelmore RW: Dm3 is one member of a large constitutively-expressed family of NBS-LRR encoding genes. *Mol Plant Microbe Interact* 2002, **15**:251-256.
  39. Yahiaoui N, Brunner S, Keller B: Rapid generation of new powdery mildew resistance genes after wheat domestication. *Plant Journal* 2006, **47**(1):85-98.
  40. Davis A, Govaerts R, Bridson D, Stoffelen P: An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 2006, **152**:465-512.
  41. Lashermes P, Combes M, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A: Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics* 1999, **261**:259-266.
  42. Anthony F, Diniz LEC, Combes M-C, Lashermes P: Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Syst Evol* 2010, **285**:51-64.
  43. van der Vossen HAM: State-of-art of developing durable resistance to biotrophic pathogens in crop plants, such as coffee leaf rust. In *Durable resistance to coffee leaf rust*. Edited by: Zambolim L, Zambolim EM, Várzea VMP. Viçosa: Universidade Federal de Viçosa; 2005:540.
  44. Prakash NS, Marques DV, Varzea VMP, Silva MC, Combes MC, Lashermes P: Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. *Theoretical and Applied Genetics* 2004, **109**(6):1311-1317.
  45. Mahe L, Le Pierres D, Combes M, Lashermes P: Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. *Genome* 2007, **50**:316-324.
  46. Lashermes P, Combes MC, Ribas A, Cenci A, Mahe L, Etienne H: Genetic and physical mapping of the S<sub>4</sub>3 region that confers resistance to leaf rust in coffee tree (*Coffea arabica* L.). *Tree Genetics & Genomes* 2010, **6**:973-980.
  47. Herrera JC, D'Hont A, Lashermes P: Use of fluorescence in situ hybridization as a tool for introgression analysis and chromosome identification in coffee (*C. arabica* L.). *Genome* 2007, **50**:619-626.
  48. Cenci A, Combes M-C, Lashermes P: Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. *Molecular Genetics and Genomics* 2010, **283**(5):493-501.
  49. Chen Q, Han Z, Jiang H, Tian D, Yang S: Strong Positive Selection Drives Rapid Diversification of R-Genes in *Arabidopsis* Relatives. *Journal of Molecular Evolution* 2010, **70**(2):137-148.
  50. Heath L, van der Walt E, Varsani A, Martin DP: Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* 2006, **80**:11827-11832.
  51. Xiao S: Origin and Maintenance of a Broad-Spectrum Disease Resistance Locus in *Arabidopsis*. *Mol Biol Evol* 2004, **21**(9):1661-1672.
  52. Seah S, Telleen AC, Williamson VM: Introgressed and endogenous Mi-1 gene clusters in tomato differ by complex rearrangements in flanking sequences and show sequence exchange and diversifying selection among homologues. *Theoretical and Applied Genetics* 2007, **114**(7):1289-1302.
  53. McDowell J, Simon S: Recent insights into R gene evolution. *Molecular Plant Pathology* 2006, **437**-448.
  54. Meyers B, Kozik A, Griego A, Kuang H, Michelmore R: Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 2003, **809**-834.
  55. Richter TE, Ronald PC: The evolution of disease resistance genes. *Plant Molecular Biology* 2000, **42**(1):195-204.
  56. Mondragon-Palomino M, Gaut B: Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol Biol Evol* 2005, **2444**-2456.
  57. Caicedo A: Diversity and molecular evolution of the RPS2 resistance gene in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 1999, **96**:302-306.
  58. Cooley MB, Pathirana S, Wu H-J, Kachroo P, Klessig DF: Members of the *Arabidopsis* HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* 2000, **12**:663-676.
  59. Dodds P, Lawrence G, Ellis J: Contrasting modes of evolution acting on the complex N locus for rust resistance in flax. *Plant Journal* 2001, **439**-453.
  60. Meyers BC: The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 1998, **10**:1817-1832.
  61. Sun Q, Collins NC, Ayliffe M, Smith SM, Drake J, Pryor T, Hulbert SH: Recombination between paralogues at the rp1 rust resistance locus in maize. *Genetics* 2001, **423**-438.
  62. Luo S, Peng J, Kungpeng L, Wang M, Kuang H: Contrasting evolutionary patterns of the R<sub>1</sub> resistance gene family in different species of Poaceae. *Mol Biol Evol* 2010, **28**:313-325.
  63. Bergelson J: Evolutionary Dynamics of Plant R-Genes. *Science* 2001, **292**(5525):2281-2285.
  64. Hurst L: The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics* 2002, **18**:486-489.
  65. Ellis J, Lawrence G, Luck J, Dodds P: Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* 1999, **11**(3):495-506.
  66. Ellis J, Dodds P, Pryor T: The generation of plant disease resistance gene specificities. *Trends Plant Sci* 2000, **5**(9):373-379.
  67. Botella MA: Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell* 1998, **10**:1847-1860.
  68. Geoffroy V, Macadré C, David P, Pedrosa-Harand A, Sévignac M, Dauga C, Langin T: Molecular analysis of a large subtelomeric nucleotide-binding-site-leucine-rich-repeat family in two representative genotypes of the major gene pools of *Phaseolus vulgaris*. *Genetics* 2009, **181**(2):405-419.
  69. Doyle JJ, Doyle JLA: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 1987, **19**:11-15.
  70. Noir S, Patheyron S, Combes MC, Lashermes P, Chalhouh B: Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). *Theoretical and Applied Genetics* 2004, **109**(1):225-230.
  71. Staden R: The Staden Sequence Analysis Package. *Mol Biotechnol* 1996, **5**:233-241.
  72. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98. *Nucl Acids Symp Ser* 1999, **41**:95-98.
  73. Sambrook J, Fritsch E, Maniatis T: Molecular cloning: a laboratory manual. Nework: Cold Spring Harbor Laboratory; 1989.
  74. Lupas A, Van Dyke M, Stock J: Predicting coiled coil from protein sequences. *Science* 1991, **252**:1162-1164.
  75. Sawyer S: Statistical test for detecting gene conversion. *Mol Biol and Evolution* 1989, **6**:526-553.
  76. Tamura K, Dudley J, Nei M, Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007, **24**:1596-1599.
  77. Librado P, Rozas J: DnaSP v.5 A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, **25**:1451-1452.
  78. Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press; 1987.
  79. Rairdan GJ, Collier SM, Sacco MA, Baldwin TT, Boettlich T, Moffett P: The coiled-coil and nucleotide binding domains of the potato Rx disease resistance protein function in pathogen recognition and signaling. *Plant Cell* 2008, **20**(3):739-751.

doi:10.1186/1471-2164-12-240

Cite this article as: Ribas et al.: Organization and molecular evolution of a disease-resistance gene cluster in coffee trees. *BMC Genomics* 2011 **12**:240.