Scientific Research Publishing

American Journal of Plant Sciences, 2018, 9, 311-324
http://www.scirp.org/journal/ajps
ISSN Online: 2158-2750
ISSN Print: 2158-2742

Molecular Variability and Genetic Structure of IYMV in Burkina Faso

Moustapha Koala[1,2*], Drissa Sérémé[1,2], Florence Vignols[3], Eric Lacombe[3], Martine Bantgratz[2,3], Bouma James Neya[1,2], Christophe Brugidou[2,3], Nicolas Barro[4], Oumar Traoré[1,2]

[1]Laboratoire de Virologie et de Biotechnologie Végétale (LVBV), Institut de l'Environnement et de Recherches Agricoles (INERA), Ouagadougou, Burkina Faso
[2]Laboratoire Mixte Internationale (LMI) Patho-Bios, Institut de l'Environnement et de Recherches Agricoles (INERA), Ouagadougou, Burkina Faso
[3]UMR IPME (UM, IRD, CIRAD), Institut de Recherche pour le Développement (IRD), Montpellier Cedex, France
[4]Université Ouaga I Professeur Joseph Ki-Zerbo, Ouagadougou, Burkina Faso
Email: *koalamoustapha@gmail.com

How to cite this paper: Koala, M., Sérémé, D., Vignols, F., Lacombe, E., Bantgratz, M., Neya, B.J., Brugidou, C., Barro, N. and Traoré, O. (2018) Molecular Variability and Genetic Structure of IYMV in Burkina Faso. *American Journal of Plant Sciences*, **9**, 311-324.
https://doi.org/10.4236/ajps.2018.92025

Received: December 19, 2017
Accepted: January 28, 2018
Published: January 31, 2018

Abstract

Imperata yellow mottle virus (IYMV, Sobemovirus) was first described in 2008 in the south-western region of Burkina Faso (West Africa). The genetic diversity of IYMV was not documented up to day. In this study, the variability of CP of IYMV was evaluated through the molecular characterization of 38 isolates collected in the western part of Burkina Faso. Comparison of sequences of these new isolates and one IYMV sequence available in GenBank revealed that the average nucleotide diversity was low. The ratio of non- synonymous over synonymous nucleotide substitutions per site was low, indicating a CP diversification under strong purifying selection. Despite of the low nucleotide diversity, phylogenetic analyses revealed segregation of IYMV isolates into six major clades. There was no correlation of phylogenetic grouping of isolates based on geographical location. This is the first study of the genetic diversity of IYMV.

Keywords

Imperata yellow mottle virus (IYMV), Coat Protein (CP), Genetic Variability, Phylogenetic Analysis

1. Introduction

The perennial grass *Imperata cylindrica* (L.) P. Beauv., a common and persistent weed in many food crops such as like cassava, maize, sorghum and rice is consi-

dered as traditional and important medicinal plant in several African country such as Uganda, Ghana and Cameroun, where the roots of *I. cylindrica* have been described as snakebite treatment in Uganda [1]. At Ghana, the properties of management of hypertension have been identified to *I. cylindrica* leaf extract while in Cameroon the properties of management of typhoid fever have been identified to *I. cylindrica* [2] [3].

Recently, only one viral disease has been reported that affects *I. cylindrica*, and was defined as the *Imperata yellow mottle virus* (IYMV) because of the typical mottled yellowing appearing at the *I. cylindrica* leaf surface. IYMV was first characterized in *I. cylindrica* in West Africa in 2008 [4] and classified as a new member of the sobemovirus genus. Like all sobemoviruses, IYMV is readily transmitted mechanically. Up to now, natural infection with IYMV has been observed and demonstrated conclusively in *Zea mays* and *I. cylindrica* [4]. Experimentally, the virus has a crop host range including two cereals (*Sorghum bicolor*, *Pennisetum glaucum*) [5] and three wild grasses (*Rottboellia exaltata Setaria verticillata, Brachiaria xantholeuca* [4] [5]. Contrary to other sobemoviruses, it remains unknown whether insects such as beetles or even the *I. cylindrica* seeds themselves can serve as vector for IYMV infection.

IYMV is a positive single stranded RNA virus with the particle of 32 nm in diameter. Its genome is 4.447 nucleotide long and comprises five ORFs [6]. ORF1 (45 - 686 nt), which is located at the 5' end of the genome, encodes a P1-like protein. P1 is involved in the cell-to-cell and systemic movement of the virus [7]. ORF2, has two overlapping ORFs, encodes the putative central polyproteins. ORF2a (713 - 2509 nt) encodes a serine protease and a viral genome-linked protein (VPg), and ORF2b (2176 - 3768 nt), encodes a RNA dependant RNA polymerase (RdRp). ORF4 (3560 - 4381 nt) is translated from the subgenomic RNA at the 3' end of the genome and encodes the coat protein. Recently, the presence of a fifth ORF (ORFx), conserved was reported in all sobemovirus [6]. Such putative fifth ORF is also present in the IYMV genome, and overlaps the 5' end of the ORF2a in the +2 reading frame (**Figure 1**).

Until now, only one complete IYMV genomic sequence from western region of Burkina Faso (West Africa) had been published [4]. The molecular diversity is therefore not documented and several important factors of epidemiology of IYMV are still poorly understood, such as alternative hosts in fields. Nevertheless, the knowledge of IYMV genetic diversity is essential for a better description of its aetiology, pathogenicity, and ecology develop appropriate strategies to counteract the IYMV spread and disease. The most common molecular markers for investigation of genetic diversity of the genus sobemovirus and other plants virus is a coat protein [8] [9]. In addition, on the basis of coat protein genes sequences, various viruses have been grouped [10]. The aim of this study was therefore to investigate the genetic variability based on molecular analyses of CP gene sequences originated from IYMV isolates obtained from different locations of Burkina Faso.

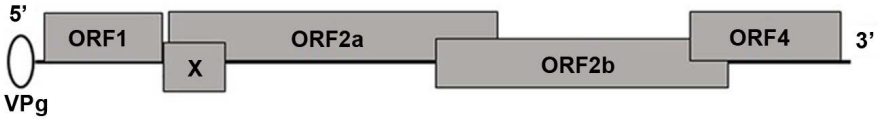

Figure 1. A schematic representation of the genome organization of IYMV; see text for details.

2. Materials and Methods

2.1. Survey and Sample Collection

Imperata cylindrica leaves showing viral symptoms of *Imperata yellow mottle virus* infection were collected in 10 different locations belonging to the high bassins region (Bama, Banzon, N'Dorola, Koloko, Tondogosso, Karangasso Sambla) and cascades regions (Banfora, Karfiguela, Lomouroudougou and Niangoloko) of Burkina Faso (West Africa), as indicated in Figure 2. One virus sample isolated from an individual *I. cylindrica* plant was considered as one isolate. Infected *I. cylindrica* plants were either used for extraction of total RNA or stored at 80˚C for future use.

2.2. RNA Extraction, RT-PCR Amplifications and Sequencing

Total RNA was extracted from frozen infected *Imperata cylindrica* leaves using the RNeasy Plant Mini Kit (Qiagen), according to manufacturer's instructions. Slight modifications were made on the protocol to optimize the quality and quantity of the total RNA. The quality of RNA extraction was compared by measurement of RNA concentration.

Reverse-transcription (RT) was performed using the primers IYMV-R4438-4454 while Polymerase Chain Reaction (PCR) was performed using IYMV-F3483-3502 and IYMV-R4385-4394 described by Koala *et al.,* 2017. All steps and conditions, including, RT and PCR followed the protocol of koala 2017 [5]. All PCR products of the correct size were purified from 1% agarose gels using GENECLEAN turbo Protocols columns before being sent to Genewiz (Essex, UK) for sequencing.

2.3. Recombination and Genetic Diversity Analysis

The sequences contigs obtained in this study were assembled using the Seqman II program in the DNASTAR 10.0 (DNAStar Inc., Madison, USA). The 38 sequences were then compared and analyzed with the available GenBank accession NC-011536 sequence (Table 1). Multiple nucleotide sequence alignments were performed by using CLUSTAL W with default parameter [11].

Alignments were also adjusted manually to guarantee correct reading frames. Noncoding sequences were removed before alignment.

As frequent recombination can provide a false positive signal for positive selection in codon specific analytical methods this paragraph is necessary. So, you need to identify and remove recombinant sequences before implemented selection pressure acting on CP genes. Interestingly, this analysis could provide important results which can improve the paper quality.

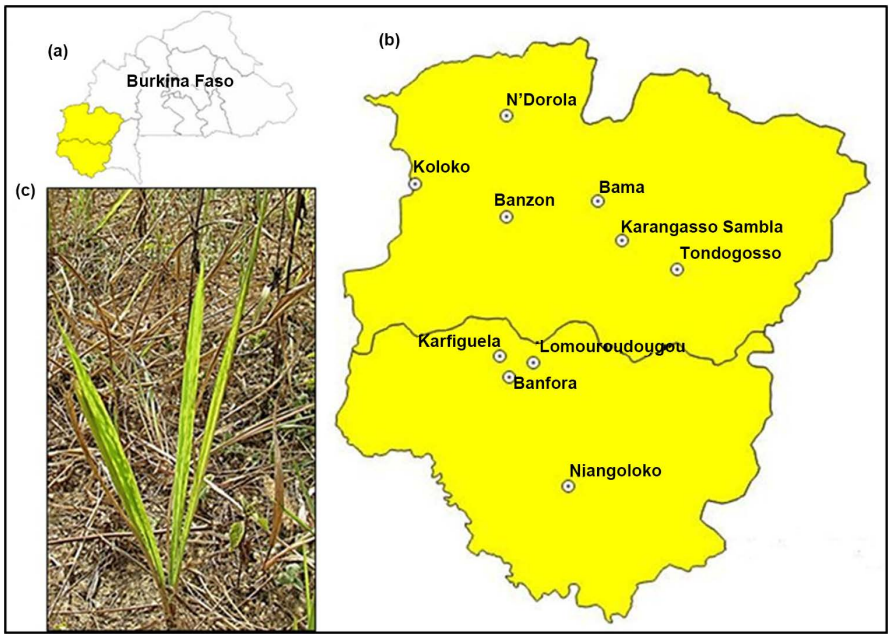

Figure 2. Geographical location of IYMV sample collection sites and symptoms on infected *Imperata cylindrica* in its natural habitat. (a) A Map of Burkina Faso showing the south-Western (in yellow) where sampling was done; (b) Precise rural provinces within the two South-Western regions where IYMV was detected and collected; (c) Typical mottle yellowing of *Imperata cylindrica* leaves guiding plant harvests.

Thus, possible recombination events were analysis using the models RDP, GENECONV, Bootscan, MaxChi Chimaera, SiScan and 3Seq implemented in the software package Recombination Detection Program (RDP, version 4.85) [12]-[18]. The default detection thresholds were used. Only events supported by three kinds of methods were retained.

Pairwise genetic distances among nucleotide and amino acid sequences were calculated using the Kimura's two parameters [19] and using the Jones Taylor Thornton (JTT) model implemented in MEGA v.6.0 [20]. To evaluate variation in selection pressure, during CP evolution, the direction and degree of selective constraints operating in a coding region were assessed by the ratio between nucleotide diversities at nonsynonymous and synonymous positions (dNS/dS).

The extent of IYMV variation among these sequences was evaluated using the index π by DnaSp version 5.0. With a sliding window of 100 nt and a step size of 25 nt. The parameter π is the mean number of nucleotide differences per site between two sequences to measure the nucleotide diversity.The value assigned to the nucleotide was that of the window midpoint.

2.4. Construction of Phylogenetic Trees

Phylogenetic relationships between isolates were inferred by maximum-likelihood (ML) methods. The best fitting nucleotide substitution model with the lowest BIC score was determined using MEGA v.6.0 [20]. ML analyses were performed under the T92 + G + I model. Isolate CP-BF1 (GenBank

Table 1. IYMV Isolates identified in different sub regions of South-Western Burkina Faso (BF). The unique IYMV sequence identified prior to this analysis [4] is given as the referent accession NC-011536.

Origin	Year	Isolate	GenBank (Acc. No)	Reference
N'Dorola	2016	IYMV1-BF	MF043148	This study
N'Dorola	2016	IYMV2-BF	MF043149	This study
Lomouroudougou	2016	IYMV3-BF	MF043150	This study
Lomouroudougou	2016	IYMV4-BF	MF043151	This study
Banfora	2016	IYMV5-BF	MF043152	This study
Karfiguela	2016	IYMV6-BF	MF043153	This study
Niangoloko	2016	IYMV7-BF	MF043154	This study
Lomouroudougou	2016	IYMV8-BF	MF043155	This study
Banfora	2016	IYMV9-BF	MF043156	This study
Tondogosso	2016	IYMV10-BF	MF043157	This study
Koloko	2016	IYMV11-BF	MF043158	This study
Karfiguela	2016	IYMV12-BF	MF043159	This study
Lomouroudougou	2016	IYMV13-BF	MF043160	This study
Tondogosso	2016	IYMV14-BF	MF043161	This study
Karfiguela	2016	IYMV15-BF	MF043162	This study
Bama	2016	IYMV16-BF	MF043163	This study
Banfora	2016	IYMV17-BF	MF043164	This study
Koloko	2016	IYMV18-BF	MF043165	This study
Tondogosso	2016	IYMV19-BF	MF043166	This study
Tondogosso	2016	IYMV20-BF	MF043167	This study
Tondogosso	2016	IYMV21-BF	MF043168	This study
Tondogosso	2016	IYMV22-BF	MF043169	This study
Karfiguela	2016	IYMV23-BF	MF043170	This study
Banzon	2016	IYMV24-BF	MF043171	This study
N'Dorola	2016	IYMV25-BF	MF043172	This study
Tondogosso	2016	IYMV26-BF	MF043173	This study
Niangoloko	2016	IYMV27-BF	MF043174	This study
Karangasso Sambla	2016	IYMV28-BF	MF043175	This study
Banfora	2016	IYMV29-BF	MF043176	This study
N'Dorola	2015	IYMV30-BF	MF043177	This study
N'Dorola	2014	IYMV31-BF	MF043178	This study
Banzon	2015	IYMV32-BF	MF043179	This study
Banfora	2013	IYMV33-BF	MF043180	This study
Bama	2014	IYMV34-BF	MF043181	This study
Banfora	2014	IYMV35-BF	MF043182	This study
N'Dorola	2014	IYMV36-BF	MF043183	This study
Bama	2014	IYMV37-BF	MF043184	This study
Banfora	2014	IYMV38-BF	MF043185	This study
Banzon	2008	NC-011536	NC-011536	Sérémé *et al.*, 2008

accession number: AJ279901.1) of *Rice yellow mottle virus* (RYMV) was used as outgroup for phylogenetic analysis. Robustness of phylogenetic relationships was assessed by 1000 bootstrap replications.

3. Results

3.1. *Imperata cylindrica* Harvest Campaigns Identify Up to 38 IYMV Isolates in South Western Burkina Faso

Within the frame of a 3-year harvest campaign in distinct areas of South-Western Burkina Faso (Figure 2), a total of 38 samples of *I. cylindrica* leaves were analyzed for *Imperata yellow mottle virus* detection.

As expected, RT-PCR on total RNAs from infected plant materials resulted in the amplification of DNA fragments of about 1000 bp for all sample listed in Table 1. PCR amplifications representative of different plants are shown (Figure 3).

3.2. Recombination Analysis

In total, four Potential Recombinant Events (PREs) named PRE_iymv34-BF, PRE_NC-011536, PRE_iymv2-BF and PRE_iymv9-BF were detected by at least one of the models (Figure 4). PRE_iymv34-BF have been the result of recombination of the major NC-011536 with an iymv38-BF minor parent. PRE_NC-011536 have been the result of recombination of the major parent iymv29-BF with an iymv2-BF minor parent. PRE_iymv2-BF shows the recombination between iymv29-BF as the major parent and Unknown (iymv25-BF) as the minor parent. These PREs (iymv34-BF, NC-011536, iymv2-BF) were detected by MaxChi methods with average P-value 2760×10^{-5}, 1068×10^{-2} and 4298×10^{-2} respectively. PRE_iymv34-BF also have been the result of recombination between iymv37-BF as the major parent and iymv38BF as the minor parent in the Chimaera and SiScan methods with average P-value 1393×10^{-2}, 1393×10^{-3}, respectively. Finally, PRE_iymv9-BF show the recombination between iymv38-BF as the major parent and Unknown (NC-011536) as the minor parent. This recombination event was detected by Chimaera method with average P-value > 1.0.

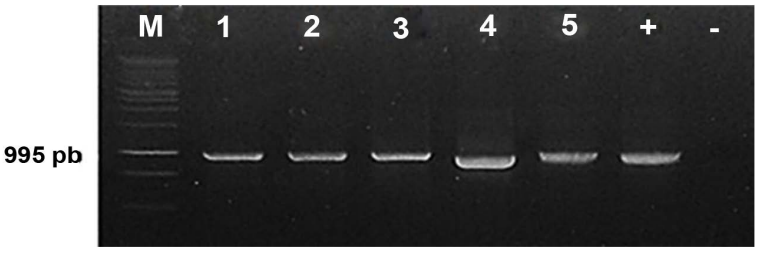

Figure 3. RT-PCR mediated molecular diagnostic for IYMV occurrence in *I. cylindrica* in Burkina Faso. Lane M: 1kb DNA size standard. Lanes 1 to 5: IYMV infected leaves from five individual plants. Lane+: PCR product from the referent accession sample NC-011536 of the *Imperata yellow mottle virus* (IYMV). Lane-: RT-PCR control performed without plant RNAs.

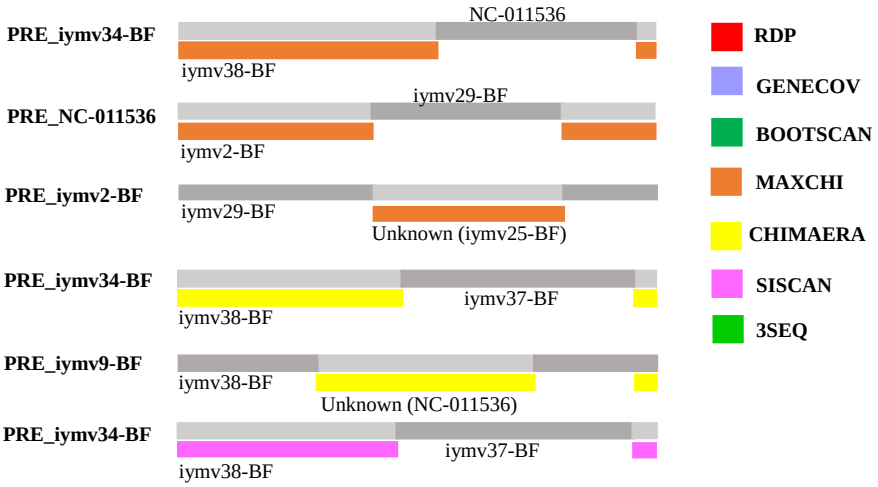

Figure 4. Description of potential recombination events. PRE_iymv34-BF, PRE_NC-011536, PRE_iymv2-BF and PRE_iymv9-BF are the Potential Recombinant Events (PREs). See text for details of major and minor parents.

These four potential recombinants were detected by one or two methods of RDP program with a low degree of confidence. In addition, one of the parental isolates was often unknown. Based on the criteria of recombination selection, these Potential Recombination events were not accepted. No evidence for potential recombination events was found among the other isolates using RPD4.

3.3. Sequence Analysis

The average of genetic diversity among the 39 listed in **Table 1** was 4.6% for nt, with the peak (7.6%) of nucleotide substitutions per site between sequences present at the 5' half N-terminal protein coding region (**Figure 5**). The average number of nucleotide substitutions per synonymous sites was high (π_s = 0.164), yet 18 times higher than the number of nonsynonymous diversity (π_a = 0.009), *i.e.* a ω ratio (π_a/π_s) of 0.07. The maximum of the nonsynonymous and synonymous diversity between two any sequences was 2.1% and 25%, respectively. As $\omega < 1$, this suggests that the CP sequences are under high purifying selective constraints. The p-value of the Z test was highly significant ($P < 0.001$) and confirmed that, diversification in the CP gene of the BF isolates was found under a strong purifying selection. Using Fisher's codon based exact test included in MEGA v.6.0 there was no evidence for positive selection (data not shown) [20].

Total number of nucleotide sites of the 39 IYMV sequences was 822 nt encoding 273 amino acids. The 273 aa residues were dominated by hydrophobic amino acids.

Analyses of the polymorphic sites among sequences of the Burkina Faso isolates revealed 136 variable sites for nucleotide and 24 for amino acid sequences. Indeed, 13% and 10% of amino acids changes resulted of mutations at 1st and 2nd nt positions of codons, respectively. We also noted that conserved amino acid sequence of CP of IYMV exhibit several common features of sobemoviruses. The N-terminal region is rich in basic amino acids and contains an arginine

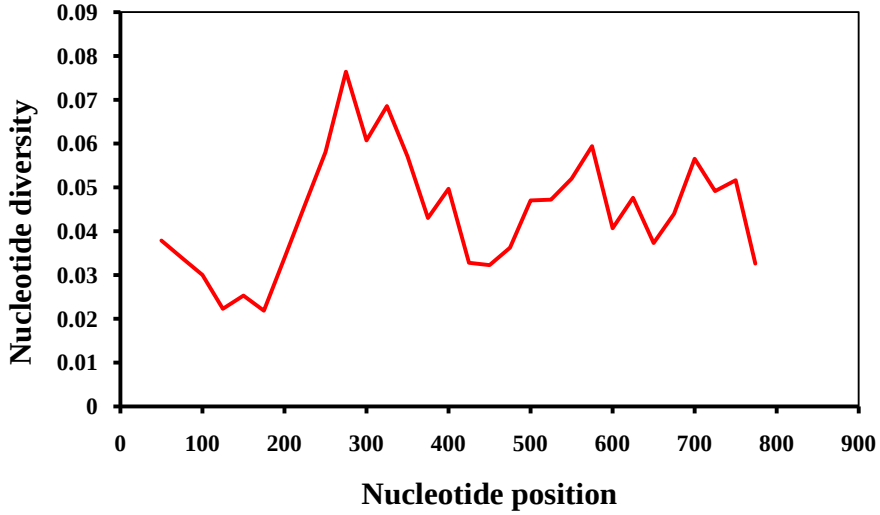

Figure 5. Distribution of IYMV genetic variation estimated by nucleotide diversity (π). The sliding window was 100 sites wide with slide set at 25 site intervals.

rich region predicted to encode a nuclear localization signal and essential for encapsidation. According the two common features of bipartite signal, the two first basic amino acids an arginine and lysine was detected in majority of isolates, in part, and the consensus bipartite targeting motif RKSKKMT13QAAAVKNQQL23APSRR was detected at position 721.

In Addition, basic amino acids (arginine, lysine, proline, and glutamine) located in N-terminal region (16) and responsible for coat protein contacts with the RNA were observed in clade 1. Its amino acids were also observed in clade 2 to 6 except proline which were replaced by threonine and lysine at position 13 and 23 respectively. Amino acid predicted to be involved in Ca^{2+} binding (two residues of aspartic acid [D139, D142], one of valine [V197] and one of asparagine [N252]) were conserved in all isolates [4].

3.4. Phylogenetic Analysis

A total 39 CP gene sequence were analyzed. The phylogenetic relationships among the sequences were constructed using maximum-likehood methods (**Figure 6**). The 39 CP nt sequences revealed segregation of the isolates under study into six clades.

Clade I was composed of (5) isolates from Tondogosso, (1) from Bama, (1) from Banfora and (1) from Banzon. Clade II included of (1) isolate from Karangasso Sambla, (1) isolate from Koloko, (1) from N'Dorola and (1) from Tondogosso. Clade III included (3) isolates from Banfora, (1) isolate from N'Dorola, (1) isolates from Koloko. Clade IV included (2) isolates from Bama, (2) isolates from karfiguela, (2) from Lomouroudougou and (1) isolate from Niangoloko. Clade V included (2) isolates from N'Dorola, (2) isolates from Banzon, (2) isolates from Lomouroudougou, (1) isolate from Banfora and (1) isolate from Karfiguela. Clade VI included (3) isolates from Banfora, (1) isolate from Banzon, (1)

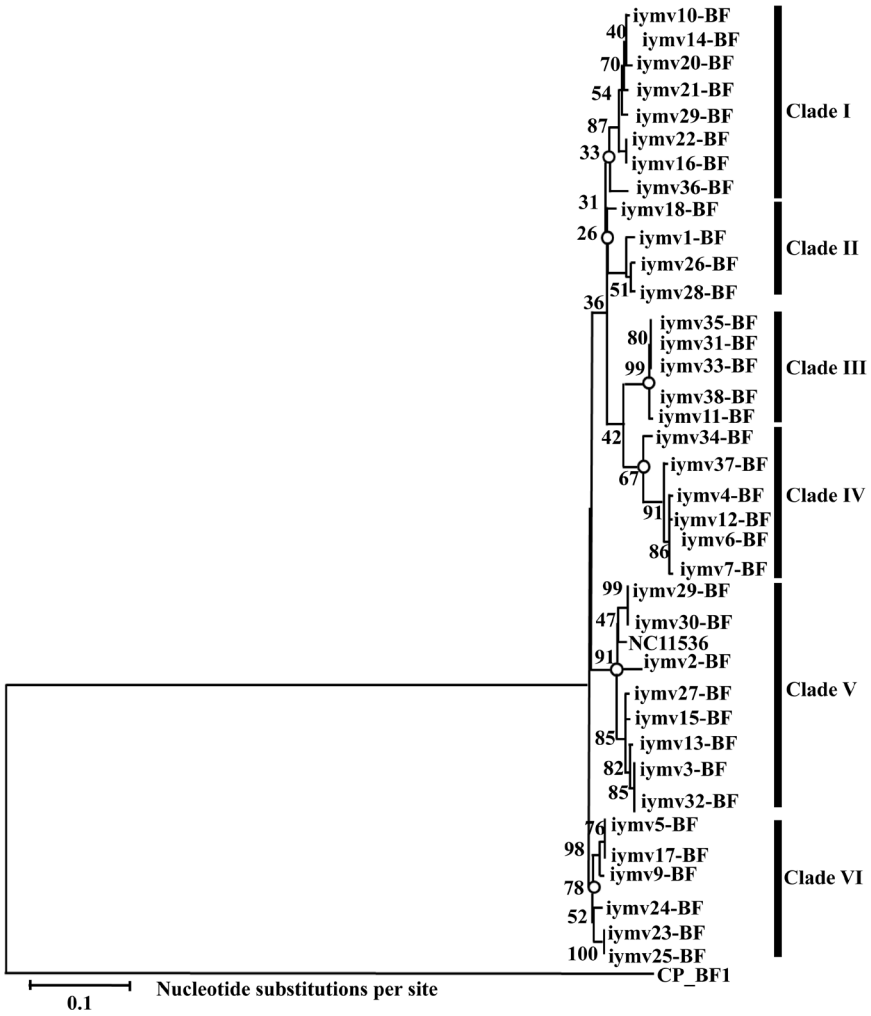

Figure 6. Phylogenetic analysis of 39 IYMV isolates from Western region of Burkina Faso based on the CP gene. The isolate CP_BF1 of RYMV from Burkina Faso was used as an out Clade. Number below branches are bootstrap percentages Scale bar indicates a genetic distance of 0.1.

isolate from N'Dorola and (1) isolate from Karfiguela.However, the Clade I and Clade II are only poorly supported (33% and 26% bootstrap values respectively, (**Figure 6**).

4. Discussion

To assess the genetic diversity of IYMV, we compared CP sequences of 39 isolates from different areas in Burkina Faso. Analysis of CP nucleotide sequences revealed that the global genetic diversity (4.6%) was low according the low CP sequence diversity (3% - 10%) of other RNA viruses [21] [22]. In spite of low genetic diversity of IYMV populations, phylogenetic analysis showed that the isolates of Burkina Faso diverged into six Clades (Clade I to Clade VI). However, the clade I and clade II are poorly supported (33% and 26% bootstrap values respectively, **Figure 6**). Analysis of distribution of IYMV isolates according to the

geographic origin indicates that isolates collected in the same locality belong to two different clades, whereas isolates from distant areas clustered in the same clade. Similarly, at amino acids level, two isolates collected from distant areas belong in the same clade. These results suggest lack of correlation between genetic diversity and geographic distribution of IYMV isolates. Similar results were also reported for *Tobacco mild green mosaic virus* (TMGMV) and *Citrus tristeza virus* (CTV) that infected perennial crops such as *Nicotiama glauna* and citrus species respectively [8] [23] [24]. The fact that *Imperata cylindrica* is a perennial grass, a nonfood and designated as a noxious weed in agricultural and nonagricultural fields the West Africa prevent the exchange of *Imperata cylindrica* propagation material. The spread via rhizomes is main mechanism of spread of *Imperata cylindrica* although some research indicated a spread by seed dispersal [25]. Therefore, *Imperata cylindrica* cannot spread to very long distance. The heterogeneous sequences between isolates of IYMV could be explained by the great potential of genetic variation in *Imperata cylindrica* reported recently [26]. Indeed, perennial grass survive a long time in nature, adapting to different environmental conditions and consequently sometimes involves the development of new ecotypes. This is also true for viruses that infect perennial plants to maintain themselves and adapt to new environmental conditions [26]. Indeed, during the adaptation to the new conditions the multiplication of viruses is accompanied by various mutations due to the lack of repair process associated with their RNA dependent RNA polymerase [21]. In addition, it have been reported that the purifying selection often results in amino acid changes with functional or structural modifications such as genome protection, cell-to-cell movement, transmission between plants, interactions with the host and/or vector, etc. [22].

IYMV CPs sequences are under high purifying selective constraints, the structuration of phylogenetic clades revealed in our analysis that the structure in six clades were associated with amino acid changes, particularly the R-domain region of coat protein (1 - 66) (**Table 2**). These amino acid substitutions are consequent as shown by the strong changes of amino acid physicochemical properties: (P) 13 (T), (P) 28 (L) (P is hydrophilic, T and L are hydrophobic but T is a polar); (G) 65 (A) (G and A are hydrophobic but G is polar uncharged); (A) 66 (S) (A is and S are polar and hydrophobic); (N) 268 (D) (N and D are hydrophilic but N is polar).

Among these amino acid substitutions in the R-domain region, we noticed particularly two amino acid changes with threonine instead proline and leucine instead proline at the position 13 and position 18 respectively. It is well established that the exceptional conformational rigidity of proline affects the secondary structure of protein suggesting strong change for the N-terminus coat protein properties [27].

Interestingly, we noticed that changes occurring particularly in the bipartite sequence (RKSKKMTQAAAVKNQQLAPSRR) of the IYMV CP which allowed to distinguish clade I from other others.

Table 2. Multiple alignment of CP amino acid sequences of IYMV Burkina Faso isolates from different localities. Consensus sequence obtained with CLUSTAL W algorithm is shown above the alignment as a consensus/majority. The amino acids identical to the consensus are indicated by points within the alignment.

	3	7	11	13	17	18	22	23	30	31	39	40	49	65	66	99	164	187	190	227	254	257	261	268	
Cons	**K**	**R**	**K**	**T**	**A**	**V**	**Q**	**L**	**G**	**R**	**P**	**S**	**V**	**G**	**S**	**I**	**C**	**M**	**R**	**A**	**A**	**S**	**V**	**N**	**Clade**
iymv10-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C I
iymv14-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C I
iymv20-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C I
iymv21-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C I
iymv19-BF	.	.	.	.	.	.	.	.	.	.	.	N	.	.	A	.	.	.	.	.	.	.	.	.	C I
iymv22-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C I
iymv16 BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C I
iymv36 BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C I
iymv18 BF	.	.	.	.	.	.	.	.	.	K	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C II
iymv1-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	C II
iymv26-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	K	.	.	.	.	D	C II
iymv28-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	C II
iymv35-BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	V	.	.	.	C III
iymv31-BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	V	.	.	.	C III
iymv33-BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	V	.	.	.	C III
iymv38-BF	.	.	.	.	V	.	P	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C III
iymv11 BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	C III
iymv34-BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	.	.	I	.	.	.	.	.	.	C IV
iymv37-BF	.	.	.	.	V	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	L	A	.	C IV
iymv4-BF	.	K	.	.	V	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	C IV
iymv12-BF	.	K	.	.	V	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	C IV
iymv6-BF	.	K	.	.	V	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	C IV
iymv7-BF	.	K	.	.	V	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	C IV
iymv8-BF	.	K	.	.	V	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	C IV
iymv29-BF	R	K	.	P	.	.	.	P	.	K	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv30-BF	R	K	.	P	.	.	.	P	.	K	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
NC 011536	R	K	R	P	.	.	.	P	.	.	.	.	M	.	A	.	.	.	.	S	.	.	.	.	C V
iymv2-BF	R	K	.	P	.	.	.	P	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv27-BF	R	K	R	P	.	.	.	P	E	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv15 BF	R	K	R	P	.	.	.	P	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv13-BF	R	.	R	P	.	.	.	P	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv3-BF	R	.	R	P	.	.	.	P	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv32-BF	R	.	R	P	.	.	.	P	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C V
iymv5-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	K	.	.	.	.	D	C VI
iymv17-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	K	.	.	.	.	D	C VI
iymv9-BF	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C VI
iymv24-BF	.	.	.	.	.	.	.	.	.	.	L	.	.	.	A	.	W	.	.	.	.	.	.	.	C VI
iymv23-BF	.	.	.	.	V	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C VI
iymv25-BF	.	.	.	.	V	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C VI

Significance C, C I: Clade I, C II: Clade II, CIII: Clade III, C IV: Clade IV, C V: Clade V, CVI: Clade VI.

Several authors have shown that the bipartite targeting sequence plays an essential role in addressing the CP protein to the nucleus [28]. However, although the basic residues R and K have similar properties we cannot say whether all substitutions at different positions in the bipartite targeting sequence have structural consequences on RNA encapsidation, stability of viral particles or other unknown properties of CP during the biological cycle of IYMV. It remains to be determined if these amino acids substitutions involve biologically distinct strains.

5. Conclusion

This is the first study of the genetic diversity of IYMV in Burkina Faso and we think that will allow contributing to a better understanding of IYMV evolution and epidemiology in Burkina Faso. In addition, the diagnosis using the specific primers will make of useful tool for population structure studies of IYMV in Burkina Faso. Although, its results are a prerequisite for further management of imperata yellow mottle disease, it would be interesting to study the genetic diversity in the neighboring countries such as in Mali and Benin (Data not shown) where the presence of IYMV has been suspected. As the global diversity of IYMV is low, it would be interesting to obtain the complete sequence of other proteins in a number of other viral protein from the different isolates representative of the 6 clades. In this context, will be particularly interesting to sequence the P1 protein as it has been demonstrated for *Rice yellow mottle virus* [29] that P1 displayed the highest diversity in the RYMV genome, and the VPg protein has is the major determinant for resistance breaking in RYMV [30].

Acknowledgements

Financial supports for this study were provided in part by the Mixed International Laboratory LMI Patho-Bios (www.pathobios.com), by the PROVEG program (Proveg: a network on plant protection through Program to support network based research in Africa, PARRAF http://proveg.org/accueil) and by the International Foundation for Science (IFS) through fellowship N˚ C/5358-1 to Moustapha KOALA. M.K was also the recipient of a financial support by the project n˚1102004 funded by the International Agropolis Foundation for his education & training in molecular cloning at IRD Institute, France. We thank Nils Poulicard for helpful discussions.

References

[1] Ntume, R. and Anywar, G. (2015) Ethnopharmacological Survey of Medicinal Plants Used in the Treatment of Snakebites in Central Uganda. *Current Life Siences*, **1**, 6-14.

[2] Tsobou, R., Mapongmetsem, P.M. and Van Damme, P. (2013) Medicinal Plants Used against Typhoid Fever in Bamboutos Division, Western Cameroon. *Ethnobotany Research & Applications*, **11**, 163-174.

[3] Mak-Mensah, E.E., Komlaga, G. and Terlabi, E.O. (2010) Antiypertensive Action of

Ethanolic Extract of *Imperata Cylindrica* Leaves in Animal Models. *Journal of Medicinal Plants Research*, **4**, 1486-1491.

[4] Sérémé, D., Lacombe, S., Konaté, M., Pinel-Galzi, A., Traoré, V.S.E., Hébrard, E., Traoré, O., Brugidou, C., Fargette, D. and Konaté, G. (2008) Biological and Molecular Characterization of a Putative New Sobemovirus Infecting *Imperata Cylindrica* and Maize in Africa. *Archives of Virology*, **153**, 1813-1820. https://doi.org/10.1007/s00705-008-0190-y

[5] Koala, M., Traoré, V.S.E., Sérémé, D., Neya, B.J., Brugidou, C., Barro, N. and Traoré, O. (2017) Imperata Yellow Mottle Virus an Emerging Threat to Maize, Sorghum and Pearl Millet in Burkina Faso. *Agricultural Sciences*, **8**, 397-408. https://doi.org/10.4236/as.2017.85030

[6] Ling, R., Pate, A.E., Carr, J.P. and Firth, A.E. (2013) An Essential Fifth Coding ORF in the Sobemoviruses. *Virology*, **446**, 397-408. https://doi.org/10.1016/j.virol.2013.05.033

[7] Bonneau, C., Brugidou, C., Chen, L., Beachy, R.N. and Fauquet, C. (1998) Expression of the Rice Yellow Mottle Virus P1 Protein *in Vitro* and *in Vivo* and Its Involvement in Virus Spread. *Virology*, **244**, 79-86. https://doi.org/10.1006/viro.1998.9100

[8] Vigne, E., Bergdoll, M., Guyader, S. and Fuchs, M. (2004) Population Structure and Genetic Variability within Isolates of Grapevine Fanleaf Virus from a Naturally Infected Vineyard in France: Evidence for Mixed Infection and Recombination. *Journal of General Virology*, **85**, 2435-2445. https://doi.org/10.1099/vir.0.79904-0

[9] Cuevas, J.M., Delaunay, A., Rupar, M., Jacquot, E. and Elena, S.F. (2012) Molecular Evolution and Phylogeography of Potato Virus Y Based on the CP Gene. *Journal of General Virology*, **93**, 2496-2501. https://doi.org/10.1099/vir.0.044347-0

[10] Neeraj, V., Mahinghara, B.K., Ram, R. and Zaidi, A.A. (2006) Coat Protein Sequence Shows That Cucumber Mosaic Virus Isolate from Geraniums (*Pelargonium* spp.) Belongs to Subgroup II. *Journal of Biosciences*, **31**, 47-54. https://doi.org/10.1007/BF02705234

[11] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research*, **22**, 4673-4680. https://doi.org/10.1093/nar/22.22.4673

[12] Martin, D.P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015) RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evolution*, **1**, vev003.

[13] Martin, D. (2000) RDP: Detection of Recombination amongst Aligned Sequences. *Bioinformatics*, **16**, 562-563. https://doi.org/10.1093/bioinformatics/16.6.562

[14] Padidam, M., Sawyer, S. and Fauquet, C.M. (1999) Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*, **265**, 218-225. https://doi.org/10.1006/viro.1999.0056

[15] Martin, D.P., Posada, D., Candall, K.A. and Williamson, C. (2005) A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. *AIDS Research and Human Retroviruses*, **21**, 98-102.

[16] Smith, J.M. (1992) Analyzing the Mosaic Structure of Genes. *Journal of Molecular Evolution*, **34**, 126-129. https://doi.org/10.1007/BF00182389

[17] Posada, D. and Crandall, K.A. (2001) Evaluation of Methods for Detecting Recombination from DNA Sequences: Computer Simulations. *Proceedings of the National Academy of Sciences*, **98**, 13757-13762. https://doi.org/10.1073/pnas.241370698

[18] Gibbs, M.J., Armstrong, J.S. and Gibbs, A.J. (2000) Sister-Scanning: A Monte Carlo Procedure for Assessing Signals in Recombinant Sequences. *Bioinformatics*, **16**, 573-582. https://doi.org/10.1093/bioinformatics/16.7.573

[19] Kimura, M. (1980) A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *Journal of Molecular Evolution*, **16**, 111-120. https://doi.org/10.1007/BF01731581

[20] Koichiro, T., Glen, S., Daniel, P., Alan, F. and Sudhir, K. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725-2729. https://doi.org/10.1093/molbev/mst197

[21] Garcia-Arenal, F., Fraile, A. and Malpica, J.M. (2001) Variability and Genetic Structure of Plant Virus Populations. *Annual Review of Phytopathology*, **39**, 157-186. https://doi.org/10.1146/annurev.phyto.39.1.157

[22] Varanda, C.M.R., Nolasco, G., Clara, M.I. and Félix, M.R. (2013) Genetic Diversity of the Coat Protein of Olive Latent Virus 1 Isolates. *Archives of Virology*, **13**, 1953-1957.

[23] Rubio, L., Ayllon, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P. and Falk, B.W. (2001) Genetic Variation of Citrus Tristeza Virus Isolates from California and Spain: Evidence for Mixed Infections and Recombination. *Journal of virology*, **75**, 8054-8062. https://doi.org/10.1128/JVI.75.17.8054-8062.2001

[24] Fraile, A., Malpica, J.M., Aranda, M.A., Rodriguez-Cerezo, E. and Garcia-Arenal, F. (1996) Genetic Diversity in Tobacco Mild Green Mosaic Tobamovirus Infecting the Wild Plant *Nicotiana glauca*. *Virology*, **223**, 148-155. https://doi.org/10.1006/viro.1996.0463

[25] Yager, L.Y., Miller, D.L. and Jones, J. (2011) Woody Shrubs as a Barrier to Invasion by Cogongrass (*Imperata cylindrica*). *Invasive Plant Science and Management*, **4**, 207-211. https://doi.org/10.1614/IPSM-D-10-00052.1

[26] Chiang, Y.C., Tsai, C.C., Hsu, T.W. and Chou, C.H. (2012) Characterization of 21 Microsatellite Markers from Cogongrass, *Imperata cylindrica* (Poaceae), a Weed Species Distributed Worldwide. *American Journal of Botany*, **99**, e428-e430. https://doi.org/10.3732/ajb.1200152

[27] Sõmera, M., Sarmiento, C. and Truve, E. (2015) Overview on Sobemoviruses and a Proposal for the Creation of the Family Sobemoviridae. *Viruses*, **7**, 3076-3115. https://doi.org/10.3390/v7062761

[28] Dingwall, C. and Laskey, R.A. (1991) Nuclear Targeting Sequences—A Consensus? *Trends Biochemical Sciences*, **16**, 478-481. https://doi.org/10.1016/0968-0004(91)90184-W

[29] Sérémé, D., Séverine, L., Konaté, M., Bangratz, M., Pinel-Galzi, A., Fargette, D., Traoré, A.S., Konaté, G. and Brugidou, C. (2014) Sites under Positive Selection Modulate the RNA Silencing Suppressor Activity of Rice Yellow Mottle Virus Movement Protein P1. *Journal of General Virology*, **95**, 213-218. https://doi.org/10.1099/vir.0.057026-0

[30] Traoré, O., Pinel-galzi, A., Issaka, S., Poulicard, N., Aribi, J., Aké, S., Ghesquière, A., Séré, Y., Konaté, G., Hébrard, E. and Fargette, D. (2010) The Adaptation of Rice Yellow Mottle Virus to the eIF (iso) 4G-Mediated Rice Resistance. *Virology*, **408**, 103-108. https://doi.org/10.1016/j.virol.2010.09.007