Available online at www.sciencedirect.com
SCIENCE DIRECT®
Infection, Genetics and Evolution 3 (2003) 189–206
Infection, Genetics and Evolution
www.elsevier.com/locate/meegid

Origin, world-wide dispersion, bio-geographical diversification, radiation and recombination: an evolutionary history of *Yam mild mosaic virus* (YMMV)

Mustapha Bousalem [a,*], Sylvie Dallot [a,1], Shinichi Fuji [b], Kieko T. Natsuaki [c]

[a] *IRD, 911 Avenue Agropolis, 34032 Montpellier, France*
[b] *Faculty of Bioresource Science, Akita Prefectural University, Shimoshinjo, Akita 010-0146, Japan*
[c] *Department of International Agricultural Development, Tokyo University of Agriculture, Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan*

Received 23 January 2003; received in revised form 6 May 2003; accepted 7 May 2003

Abstract

We developed an evolutionary epidemiological approach to understand the regional and world-wide dispersion of *Yam mild mosaic virus* (YMMV) by retracing its evolutionary history. Analyses of the distribution and the prevalence of YMMV in the Caribbean islands of Guadeloupe and Martinique, and in French Guyana revealed that YMMV has a wide repartition and different prevalence on *Dioscorea alata* L. (Asian and Oceanic origin), on *D. cayenensis* Lam.–*D. rotundata* Poir. (African origin) and on *D. trifida* L. (Amazon and the Caribbean origin) in this region. Considering the data on the current dispersion of the virus and the evolution and the history of the yams, the phylogenetic analysis of the 3′ terminal part of the YMMV genome gave a consistent support of the Asian-Pacific origin of YMMV from *D. alata* species. The YMMV phylogenetic tree is star-like, suggesting an early split of the genetic lineages. An important part of the clades is constituted by a single lineage arisen by recombination. The largest emerging monophyletic group illustrates well YMMV geographical dispersion. This evolutionary pattern contrasts with the one revealed by the African distinct lineages and by the second significant monophyletic group, for which a host adaptation to *D. trifida* is suggested. The analysis of the pattern of nucleotide substitutions in the CP gene revealed that purifying selection dominates the evolution of the CP of potyviruses and strongly operates on the YMMV. Switching events, radiation, host and geographical adaptation and recombination events are proposed as major traits of YMMV evolutionary history.
© 2003 Elsevier B.V. All rights reserved.

Keywords: *Yam mild mosaic virus*; *Dioscorea*; Potyviruses; Evolution; Phylogeny; Diversification; Adaptation; Recombination; Radiation; Purifying selection

1. Introduction

The evolution, dispersion and epidemiological properties of viruses are believed to be determined through a combination of constraints imposed by the host(s), the vectors, the environmental conditions and human activity. Facing emerging and re-emerging viruses associated with increased human migrations, movement of plant material and climatic changes, it appears crucial to understand the evolutionary mechanisms which drive virus selection and diversification. Indeed, comprehension of these mechanisms would help us to manage diseases and the risk of introduction of new viruses.

Many studies have been undertaken with respect to the evolution of viruses including plant viruses (Garcia-Arenal et al., 2001) and have been reviewed for different groups of viruses (Roossinck, 1997; Simon and Bujarski, 1994; Gibbs, 1999). However, very few studies have developed an integrated approach to virus evolution that includes aspects of epidemiology, evolutionary biology and viral population genetics at the same time. Thus, the evolutionary perception may be blurred by the heterogeneity of the data and the diversity of the analytical methods used.

We attempted to develop an integrative approach to understand potyvirus evolution using yam potyviruses as a model (Bousalem et al., 2000a,b). Potyviruses (genus *Potyvirus*, family *Potyviridae*) comprise the largest and economically most important genus of plant RNA viruses (Shukla et al., 1994). Yam potyviruses constitute informative

* Corresponding author. Present address: INRA-URPV, Domaine Duclos, Prise d'Eau, 97170, Petit-Bourg, Guadeloupe, France.
Tel.: +590-590-25-59-15.
E-mail address: bousalem@antilles.inra.fr (M. Bousalem).
[1] Present address: USDA/ARS/USHRL, 2001 South Rock Road, Fort Pierce, 34945 FL, USA.

1567-1348/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S1567-1348(03)00085-6

natural models for comparative molecular evolution because yams (genus *Dioscorea*) are characterized by a very large diversity in the number of species, their geographical origin and in their world-wide distribution (Orkwor, 1998). Moreover, a better understanding of the diversification and evolution of yam potyviruses is needed, especially with the dramatic increase in their incidence in the Caribbean. Vegetative propagation through tubers, insect transmission and uncontrolled introductions of infected germplasm have resulted in the spreading of potyviruses over different cropping areas world-wide (Thouvenel and Fauquet, 1979; Porth et al., 1987; Goudou-Urbino et al., 1996; Odu et al., 1999; Bousalem and Dallot, 2000; Dallot et al., 2001).

In a first approach recently published (Bousalem et al., 2000a,b), we used *Yam mosaic virus* as a model. Our results showed that recombination events and host adaptation were some of the evolutionary mechanisms at the origin of the high genetic diversity and the distant phylogenetic relationships observed among natural populations of YMV. In the current study, we extend our understanding of potyvirus evolution by using *Yam mild mosaic virus* (YMMV) as a complementary model to YMV. YMMV is a recently recognised distinct potyvirus infecting *Dioscorea alata* in the South Pacific (Mumford and Seal, 1997; Fuji et al., 1999). It was originally described as *yam virus I*, but is synonymous to *Dioscorea alata virus* (DaV) and is known to be widespread on *D. alata* in Africa (Odu et al., 1999). We recently identified YMMV in the Caribbean and in South America, and a first molecular characterisation showed a significant molecular divergence between isolates from Colombia, Martinique and Papua New Guinea (Bousalem and Dallot, 2000; Dallot et al., 2001).

The data on the evolution and the history of the yams are essential for a better comprehension of the diversification of the yam potyviruses. Food yams are believed to have originated from tropical areas of three separate continents: Africa (*D. cayenensis–D. rotundata*), Southeast Asia and the South Pacific (*D. alata*), and South America (*D. trifida*). However, human activity has profoundly influenced the distribution of yams as economic plants. *D. alata* has now a world-wide dispersion, *D. cayenensis–D. rotundata* is present in Africa and tropical America whereas *D. trifida* is restricted to its area of origin (Coursey, 1976; Orkwor, 1998; Hahn, 1987; Degras, 1993).

Preliminary available evidence clearly indicates a link between YMMV distribution and the origin, migration and current dispersion of *D. alata* from South Pacific. YMMV may thus be a complementary model to YMV based on the differences in its dispersion pattern and in its geographical and host origins. Indeed, our previous study allowed us to conclude to the hypothesis of an African origin for YMV from the yam complex *D. cayenensis–D. rotundata*, followed by independent transfers to *D. alata* and *D. trifida* during virus evolution (Bousalem et al., 2000b). In this study, we develop an evolutionary epidemiological approach to understand the regional and world-wide dispersion of YMMV, by retracing its evolutionary history. Special attention has been paid to the study of the genetic variation and quantification of the viral populations and to the factors and processes determining such genetic structure.

2. Materials and methods

2.1. Virus sources and plants

A large sampling of yam plants was performed taking into account the major species as well as their geographical centres of origin and diversification. It included (i) *Dioscorea alata* L. (Asian and Oceanic origin), (ii) the specific complex *Dioscorea cayenensis* Lam.–*Dioscorea rotundata* Poir. (African origin) and (iii) *Dioscorea trifida* L. (Amazon and the Caribbean origin) species. Additional sampling was performed on the minor yam species *Dioscorea esculenta* (Asian origin) and the wild yam *Dioscorea dumeturum* (African origin).

Prevalence of YMMV in French Guyana (north-east South America) and in the French Caribbean islands of Guadeloupe and Martinique was evaluated in 377 samples collected from yams showing typical virus symptoms and belonging to *D. cayenensis–D. rotundata* (23 cultivars), *D. alata* (18 cultivars) and *D. trifida* (2 major cultivars and 14 experimental cultivars). Comparisons of YMMV prevalence on the three species (3×2 contingency table) and locations (islands versus continental areas, 2×2 contingency tables) were performed using the Fisher test statistics with exact *P*-values computed by permutation (Freeman and Halton, 1951; Mehta and Patel, 1995). In the islands versus continental areas, the alternative hypothesis was “higher prevalence in the islands”.

The evaluation of the genetic diversity was conducted on a larger sampling including isolates from different cultivated and wild yam species from various parts of the world: Africa (Togo), South America (Colombia, Brazil), Central America, (Costa Rica), Asia (Japan), Oceania (Papua New Guinea, Fiji) (Table 1). Additional sequences of the partial CP core region and the 3′-NCR of 17 isolates were retrieved from Genebank database (Table 1). The isolates were from Africa (Togo, Nigeria, Benin, Ivory coast, Guinea, Ghana and Cameron), Asia (Sri-Lanka) and Oceania (Vanuatu).

2.2. Specific detection of YMMV by continuous immuno-capture-RT-PCR

We developed an IC-RT-PCR procedure as previously performed for YMV diagnosis (Bousalem et al., 2000a). This technique was improved for multiplex PCR detection by using universal potyvirus monoclonal antibodies (AGDIA) at a 1/50 dilution in carbonate buffer for successful YMMV detection. We used the specific YMMV detection primers UTR1R 5′-CACCAGTAGAGTGAACATAG-3′ and CP2F 5′-GGCACACATGCAAATGAAAGC-3′ (Mumford

Table 1
YMMV isolates in relation with their origin (country, host species) and date of sampling sequenced in this study (A) and for which the sequence was retrieved from the Genebank database (B)

Isolate	Collecting country	*Dioscorea* species	Year of sampling	Accession no.
(A) *Isolates sequenced in this study*				
Col 1	Colombia	*D. alata*	1999	*AF548491*
Col 2		*D. alata*	1999	*AF548492*
Mart 1	Martinique	*D. cay.–rot.*[a]	1998	*AF548493*
Mart 2		*D. alata*	1998	*AF548494*
Mart 3		*D. alata*	1998	*AF548495*
Mart 4		*D. alata*	1998	*AF548496*
Mart 5		*D. cay.–rot.*	1998	*AF548497*
Mart 6[b]		*D. alata*	1998	*AF548528*
Br2	Brazil	*D. alata*	2000	*AF548498*
CR 1	Costa Rica	*D. trifida*	1999	*AF548499*
Guad 1	Guadeloupe	*D. alata*	1998	*AF548500*
Guad 2		*D. alata*	1998	*AF548501*
Guad 3		*D. alata*	1998	*AF548502*
Guad 4		*D. trifida*	1998	*AF548503*
Guad 5		*D. trifida*	1998	*AF548504*
Guad 6		*D. esculenta*	1998	*AF548505*
Guy 1	French Guyana	*D. trifida*	1998	*AF548506*
Guy 2		*D. trifida*	1998	*AF548507*
Guy 3		*D. trifida*	1998	*AF548508*
Guy 4		*D. trifida*	1998	*AF548509*
Guy 5		*D. trifida*	1998	*AF548510*
Guy 6		*D. trifida*	1998	*AF548511*
Guy 7		*D. trifida*	1998	*AF548512*
Guy 8		*D. trifida*	1998	*AF548513*
Togo 1	Togo	*D. alata*	1999	*AF548514*
Togo 40[b]		*D. dumetorum*	1999	*AF548527*
Fiji1	Fiji	*D. alata*	2000	*AF548515*
Fiji2		*D. alata*	2000	*AF548516*
Fiji3		*D. alata*	2000	*AF548517*
Fiji4		*D. alata*	2000	*AF548518*
JPN 1	Japan	*D. alata*	1997	*AF548519*
JPN 2		*D. alata*	1998	*AF548520*
JPN 3		*D. alata*	1998	*AF548521*
JPN 4		*D. alata*	1993	*AF548522*
PNG 1	Papua New Guinea	*D. alata*	1993	*AF548523*
PNG 2		*D. alata*	1993	*AF548524*
PNG 3		*D. alata*	1993	*AF548525*
PNG 4		*D. alata*	1994	*AF548526*
(B) *Previously sequenced isolates*				
DaSLK 1	Sri-Lanka	*D. alata*	–	*AJ305464*
DaSLK 2		*D. alata*	–	*AJ305463*
DeSLK 1		*D. esculenta*	–	*AJ305457*
DaVan 1	Vanuatu	*D. alata*	–	*AJ305458*
DaTog 2	Togo	*D. alata*	–	*AJ305462*
DaTog 3		*D. alata*	–	*AJ305461*
DaCDI 19	Ivory Coast	*D. alata*	–	*AJ305470*
DaCDI 37		*D. alata*	–	*AJ305469*
DrCmn 11	Cameron	*D. cay.–rot.*	–	*AJ305456*
DaCmn 42		*D. alata*	–	*AJ305468*
DaBen 1	Benin	*D. alata*	–	*AJ305471*
DaNig 2	Nigeria	*D. alata*	–	*AJ305466*
DaNig 316		*D. alata*	–	*AJ305465*
DaGui 1	Guinea	*D. alata*	–	*AJ305467*
DrGui 1		*D. cay.–rot.*	–	*AJ305455*
DaUga 1	Ghana	*D. alata*	–	*AJ305460*
DaUga 3		*D. alata*	–	*AJ305459*

[a] *D. cayenensis–D. rotundata*.
[b] Sequence including the partial core of the CP and the 3′-NCR.

and Seal, 1997), that amplify a 249 bp fragment including the partial CP gene and the 3′-NCR. The IC-RT-PCR procedure was conducted as previously described (Bousalem et al., 2000a) with the exception of the PCR amplification scheme (40 cycles of 30 s at 95 °C, 30 s at 55 °C, 1 min at 72 °C followed by a last elongation 10 min at 72 °C.

2.3. Uncoupled IC-RT-PCR for sequencing

The Nib C-ter, complete CP and 3′-NCR were amplified using a set of four primers: the upstream primers specific to Nib C-ter were YMMV-S1 5′-GGAAATAGGCATAAGCGTAAGC-3′ as we previously described (Fuji et al., 1999) and YMMVDP 5′-GACTYGAAGCYATYTGYGC-3′. As downstream primers for both retro-transcription and amplification, we used an oligodT or the 3′-NCR specific primer YMMVR1 5′-TACTATGCAATACTTATAT-3′. The amplified fragments were 1200 and 1116 bp for, respectively, oligodT and YMMVR1 primers.

Sample extraction and immunocapture were performed using the same procedure as the continuous IC-RT-PCR. Reverse-transcription was conducted as previously described for YMV (Bousalem et al., 2000a). For amplification, 3 μl of cDNA products were used in a 25 μl final volume PCR mix per tube, containing 1× manufacturer Taq buffer (Tris–HCl, KCl, $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, pH 8.7), 250 μM of each dNTPs, 1 μM of each primer, 2.5 U of HotStar Taq polymerase (QIAGEN). The amplification scheme was 95 °C for 15 min, followed by 35 cycles at 94 °C for 1 min, 52 °C for 1 min and 72 °C for 1 min and then one final elongation at 72 °C for 5 min.

2.4. DNA sequencing

After electrophoresis (1% agarose in TAE buffer), PCR fragments of interest were excised and purified. Sequencing was performed either directly or after cloning.

Direct sequencing was performed by using the previously described primers as well as internal primers CP2F and CP2Fr 5′-GCTTTCATTTGCATGTGTGCC-3′. For isolates Fiji 1–4, PNG 1–4, JPN 1–4, PCR products were cloned into pGEM-T Easy Vector (Promega, WI, USA) and three clones for each isolate were sequenced.

Sequencing was carried out with Taq terminator sequencing kit (Applied Biosystems) and analysed on an Applied Biosystems 373A or 377 sequencer. Sequences were assembled and analysed by the auto-assembler software (Applied Biosystems, Inc.).

2.5. Phylogenetic analysis of the data

2.5.1. Partition of the data set

A region of 1050 nt has been sequenced for 36 isolates that includes the contiguous C-terminal region of NIb (114 nt), the full CP (798 nt) and the 3′-NCR (144 nt). Within the CP, we distinguished: (i) the N-terminal region (N-ter; 189 nt)

encoding the hyper-variable amino-terminal extension of the CP polypeptide (up to the lysine amino-acid K^{31} in the DVN motif in YMMV isolates, homologous to the KDVN motif of PVY-N: (Shukla et al., 1994); (ii) the conserved core and C-ter regions (core: 705 nt) encoding the core (D^{32}–E^{246}) and C-terminal (R^{247}–I^{266} homologous to R^{248}–M^{267} in PVY-N), (Shukla et al., 1994). Phylogenetic analyses were first carried out on the complete sequences combining all the available information and then on each of the four independent partitions (NIb, N-Ter CP, core CP, 3′-NCR).

2.5.2. Phylogenetic analysis

Sequences were aligned by the Clustal algorithm from the MegAlign software of DNASTAR package (version 1.2) Phylogenetic reconstructions were obtained by: (i) the maximum parsimony (MP) method (PAUP 4.0b8.) (Swofford, 1998) with all molecular characters assessed as independent, unordered, and equally weighted (Fitch, 1971); (ii) the maximum likelihood (ML) method (Felsenstein, 1981) using the quartet puzzling method (PAUP 4.0b8.) (Strimmer and Von Haeseler, 1996) with the Tamura and Nei model (Tamura and Nei, 1993) of sequence evolution. The MP analyses were conducted with heuristic searches using 100 replicates of random taxon addition, tree bisection–reconnection swapping (TBR), saving of all shortest trees (MULPARS on), and ACCTRAN optimization. The ML analyses were conducted with 10,000 puzzling steps. For the CP gene analysis, the two most closely related potyviruses *Potato virus Y* (PVY) (2) and *Pepper mottle virus* (PepMoV) were used as outgroups. For partial gene analysis, PNG 1 isolate was used as outgroup to avoid the misalignment of the variable regions of YMMV with PVY and PepMov. The choice of the isolate PNG 1 was based on a preliminary analysis with unambiguously alignable nucleotides of the CP conserved core region, including PVY and PepMoV as outgroups.

Robustness of the nodes of the phylogenetic trees was assessed by (i) the bootstrap (BP) computed after re-samplings followed by a MP reconstruction (1000 replicates of heuristic search, one random sequence addition per replicate and MAXTREES = 100), (Felsenstein, 1985); (ii) the reliability percentages (RP), i.e. the number of times the group appears after 10,000 ML puzzling steps (PUZZLE 4.0; Strimmer and Von Haeseler, 1996).

2.5.3. Estimation of evolutionary distances

Sequence data were used to estimate the number of nucleotide substitutions per site (d) using the Kimura's two-parameters method (Kimura, 1980). To examine the pattern of nucleotide substitutions in the CP gene, the number of synonymous substitutions per site (d_s) and the number of non-synonymous substitutions per site (d_n) were estimated. Contribution of positive selection to sequence variation has been assessed by the comparison of the ratio d_n/d_s. A ratio of less than 1 indicates that purifying selection is uppermost, while a ratio greater than 1 is taken as evidence of an excess of positive selection.

We used two groups of methods based on different assumptions: (i) the evolutionary pathway methods: the Nei–Gojobori method (Nei and Gojobori, 1986; Ina, 1995), (ii) methods based on Kimura's two-parameter model: the Li–Wu–Luo method (Li et al., 1985), the Pamilo–Bianchi–Li method (Pamelo and Bianchi, 1993; Li, 1993), the Kumar method (Nei and Kumar, 2000). All the analyses were performed with the MEGA2 program (Kumar et al., 2001).

2.6. Recombination analysis

Recombination events were investigated according to a global approach previously performed for YMV (Bousalem et al., 2000b). This approach included a phylogenetic comparative analysis as well as the implementation of three programs designed for the detection of recombination, each one based on a different method (Posada and Crandall, 2001): (i) the phylpro program (bêta version 0.9) (Weiller, 1998) is a distance-based method looking for inversions of distance patterns among the sequences, (ii) Geneconv 1.81, a new version of the VTDIST2 program (Sawyer, 1989) can be categorized as a substitution distribution method and examines pairs of aligned sequences for a significant clustering of substitutions or fit to an expected statistical distribution; (iii) RDP program implements a phylogenetic approach, inferring recombination when phylogenies from different parts of the genome result in discordant topologies (Martin and Rybicki, 2000).

The phylogenetic approach was carried out by performing comparisons of the phylogenies reconstructed from: (i) the full sequence of 1056 nt comprising the contiguous C-terminal region of NIb, the complete CP and the 3′-NCR, (ii) the complete CP, (iii) the 3′-NCR, and (iv) various partitions of the sequences. Among them, the most informative were (i) the C-terminal region of NIb and the contiguous N-terminal part of the CP and (ii) the partial core and C-terminal part of the CP and the contiguous 3′-NCR.

The three recombination programs were carried out with the full sequences of 36 isolates and with the partial core CP and the 3′-NCR of 54 isolates including the additional sequences retrieved from the Genebank database.

For the Phylpro analysis, pairwise distances of all sequences including all variables sites were determined within a split window of 40 columns. Conversion events in the Geneconv program were evaluated by a pairwise comparison of sequence segments excluding monomorphic sites. The likelihood of genetic rearrangements in the data set was determined by 10,000 random permutations of the sequences. The significance was evaluated for both global and pairwise comparisons. The option allowing the detection of "within-group" gene conversions was used. The RDP program analyzed every permutation of three sequences in the aligned set, attempting to identify two "parental" and one recombinant sequence. We performed successive analyses modifying the window size and the maximum associated probability (*P*). For large window sizes (50–100

nucleotides), P was fixed for a range from 0.05 to 0.001, and from 0.001 to 0.000001 for small window sizes (5–20 nucleotides).

3. Results

3.1. Epidemiological observations: high prevalence and wide distribution of YMMV in the Caribbean and French Guyana

Field surveys were performed in all the areas of yam cropping, to be representative of the different agro-ecosystems of the French Caribbean islands and of the French Guyana. Three hundred seventy seven yam samples were collected from the three main species *D. cayenensis–D. rotundata* (23 cultivars), *D. alata* (18 cultivars) and *D. trifida* (2 major cultivars and 14 experimental cultivars. *Yam mild mosaic virus* was detected in 15% of the 377 samples analyzed by IC-RT-PCR. YMMV prevalence was found to differ significantly on the three yam species (Fisher statistic = 59.9 with 2 d.f., $P = 1 \times 10^{-13}$). *D. trifida* was the most infected species (37.6% of infected samples) when compared with the other two species (Fisher statistics = 57.8 with 1 d.f., $P = 1.7 \times 10^{-13}$), whereas YMMV prevalence was significantly higher on *D. alata* than on *D. cayenensis–D. rotundata* (respectively, 11.2 and 3.6% of infected samples, Fisher statistics = 5.1 with 1 d.f., $P = 3.3 \times 10^{-2}$).

Distinct situations in YMMV prevalence on the three yam species were also observed between continental and islands areas, (Fig. 1). In French Guyana, only *D. trifida* was found infected by YMMV in 29.5% of the samples tested. The virus was detected in all the areas where *D. trifida* is grown, in the traditional cropping areas along the pacific coast and the Maroni river as well as in the indigenous ancestral cropping areas in the rain forest. The situation was found nearly identical in Guadeloupe and Martinique with a wide distribution of YMMV, independent of the traditional or modern cropping system. *D. trifida* was frequently infected with an infection rate of 52.6% in Guadeloupe and 46.6% in Martinique, followed by *D. alata* (respectively, 9.5 and 28.6% of positive samples). Both *D. trifida* and *D. alata* were found significantly more infected in the Caribbean islands than in French Guyana ($P = 0.009$ and 0.037, respectively, for *D. trifida* and *D. alata*). YMMV prevalence in *D. cayenensis–D. rotundata* was low (1.96 and 9.1% of infected samples in Guadeloupe and Martinique) and the situation in the islands did not differ significantly from the absence of infections observed in French Guyana ($P = 0.149$). Additionally, YMMV was detected in five of the seven samples collected in Guadeloupe and Martinique on the minor cultivated species *D. esculenta*.

3.2. Phylogenetic structuring of YMMV diversity

3.2.1. Evidence for eight phylogenetic groups from the analysis of the CP gene sequences of 36 isolates

A first phylogenetic analysis was conducted on the 798 nt of the CP gene from 36 isolates reported in Table 1. The corresponding phylogram is reported on Fig. 2. Eight groups (groups I–VIII) were evidenced by the two phylogenetic ML and MP reconstruction methods, using as outgroups the two most closely related potyviruses PVY and PepMoV. Six groups (I–VI) are represented by individual taxon with non-significant inter-relationship supports (BP an RP less than 50%). The polytomie structure of the resulting trees makes difficult to establish a relationship between the different clades. However, two major groups supported by high BP and RP values, the groups VII and VIII, emerged. Among each group, three sub-groups (a, b, c) were identified with significant BP and RP values.

To avoid any misleading conclusion due to the polytomie structure of the tree, we performed a second analysis based on the full sequences of 1056 nt. Congruent results with CP analysis were observed (data not shown), except for isolate Mart 2 (group II) that shifted to group

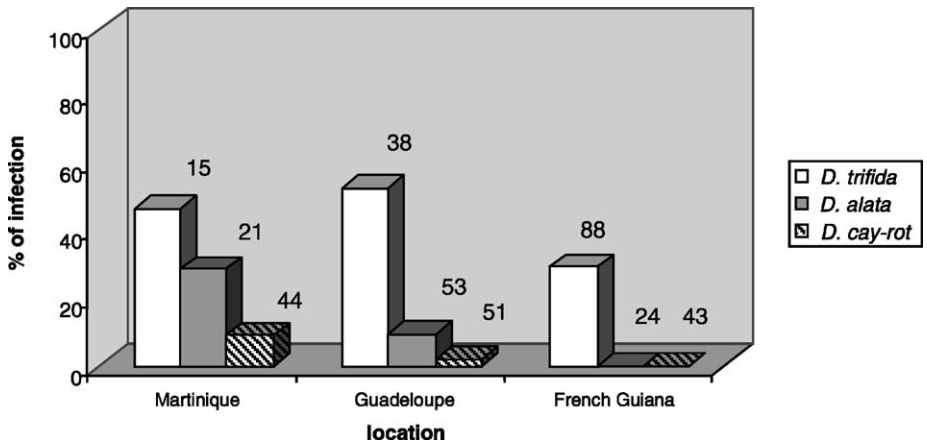

Fig. 1. YMMV prevalence on the three main yam species (*D. trifida*, *D. alata* and *D. cayenensis–D. rotundata*) in the islands of Guadeloupe and Martinique (Caribbean) and in French Guyana (South America). The numbers at the top of the bars indicate the number of samples tested.

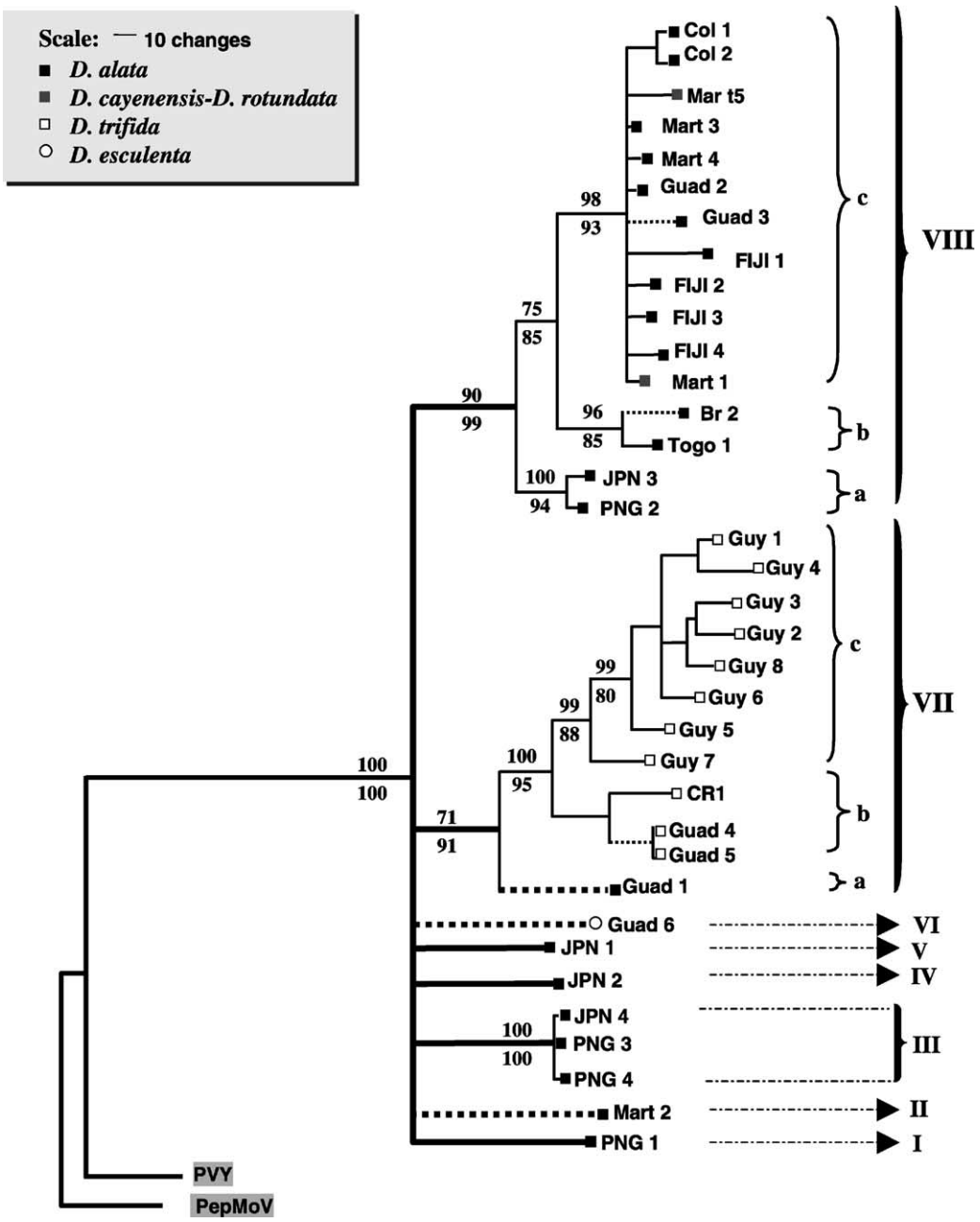

Fig. 2. Phylogenetic analysis of the CP gene of 36 isolates of YMMV: phylograms reconstructed from the CP nucleotide sequences using the maximum parsimony analysis and rooted by the most closely related potyviruses PVY and PepMV as out-groups. The robustness of the nodes was assessed by bootstrap percentages (50% majority-rule consensus trees, numbers above branches). The quartet puzzling maximum likelihood (ML) method analysis was conducted on the same nucleotide matrix and robustness of the nodes was assessed by the reliability percentages (below branches). Branch lengths are proportional to the number of substitutions inferred. The main phylogenetic groups are indicated by thick bars and labelled horizontally. Recombinant isolates are indicated with the dotted bars. The host of origin of the isolates is mapped on the phylogram and the legend is reported on the figure: black, white, and hatched squares correspond to YMMV, respectively, sampled on *Dioscorea alata*, the complex *D. cayenensis–D. rotundata* and *D. trifida*; white circle corresponds to YMMV sampled on *D. esculenta*.

VII (BP = 61). This change of clustering of Mart 2 has been related to a recombination event. Pairwise sequences inter-group comparisons (Table 2) confirmed the divergence of each group and provided supplementary support of the structure of YMMV diversity in eight phylogenetic groups.

3.2.2. Evidence for additional grouping from the analysis of 54 sequences of the partial CP core region and the 3′-NCR

An additional phylogenetic analysis was conducted on the partial CP core and the 3′-NCR sequences of 54 isolates, including 16 isolates retrieved from the Genebank database,

Table 2
Genetic diversity of YMMV

Phylogenetic analysis groups	Assessment of genetic diversity of the CP gene												
	YMMV intra and inter-group comparison[a]								Potyviruses species comparison[b]				
	I	II	III	IV	V	VI	VII	VIII	Species	% Divergence (aa)	d	d_n	d_s	d_n/d_s
I		0.174	0.149	0.148	0.148	0.165	0.196	0.175	YMV 27	11.5 (18.5)	0.130	0.058	0.356	0.162
	–	0.871	0.590	0.595	0.568	0.649	0.842	0.776						
		0.014	0.023	0.017	0.024	0.032	0.042	0.028						
II	2.3		0.156	0.158	0.167	0.182	0.173	0.170	BYMV 24	8.2 (15.8)	0.143	0.050	0.504	0.100
	–	–	0.612	0.626	0.714	0.758	0.674	0.739						
			0.027	0.023	0.031	0.031	0.040	0.029						
III	4.3	3.4	**0.004**	0.150	0.145	0.158	0.175	0.156	BCMV 19	7.3 (12.9)	0.078	0.036	0.200	0.180
	4.5	3.8	**0.013**	0.565	0.549	0.578	0.638	0.632						
			0.001	0.025	0.029	0.034	0.049	0.026						
IV	3.8	3.8	4.5		0.130	0.160	0.175	0.158	PPV 34	6.9 (26.2)	0.087	0.041	0.243	0.168
	–	–	5	–	0.442	0.612	0.655	0.625						
					0.026	0.030	0.041	0.025						
V	4.5	4.9	4.5	5		0.184	0.158	0.154	YMMV 36	6.2 (12.4)	0.140	0.033	0.519	0.063
	5	–	5.3	–	–	0.718	0.536	0.597						
						0.043	0.049	0.032						
VI	6.8	6.8	7	8	8.9		0.179	0.198	PRSV 77	5.4 (10.3)	0.084	0.027	0.261	0.103
	–	–	7.5	–	–	–	0.645	0.873						
							0.051	0.037						
VII	5.6	5.6	7	6.4	7.5	10	**0.096**	0.169	PVY 34	5.2 (10.5)	0.078	0.028	0.228	0.122
	8.7	8.7	10.6	10	11.3	12.4	**0.297**	0.603						
							0.027	0.047						
VIII	3.8	3.4	4	4	5	7	7.5	**0.065**	SCMV 85	5.2 (10.5)	0.073	0.084	0.065	1.292
	7	3.8	7	6	8.3	10.5	13	**0.209**						
								0.012						

[a] YMMV intra and inter-group comparison: above the diagonal: the inter-group values: respectively, d, d_s and d_n. The corresponding intra-group values are indicated (excluding group constituted by one isolate) on the diagonal. Below the diagonal: respectively, minimum and maximum inter-group amino acid divergence.

[b] Potyviruses species comparison: the number of sequences analysed is indicated under the virus name. Mean and maximal (under brackets) amino acid divergence values of the CP are given. The nucleotide diversity (d) is expressed as nucleotide substitution by site, while d_n represents the non-synonymous substitution and d_s the synonymous substitution. The accession numbers of the potyviruses sequences used in this study are available under request to the first author.

our 36 sequences (Table 1) and two other isolates from Martinique and Togo (Mart 6 and Togo 40).

Although the phylogenetic analysis of these sequences did not conform completely to the branching pattern displayed by the complete CP analysis, the phylogenetic grouping of the new sequences provided supplementary information (Fig. 3). The most interesting point in a geographical perspective analysis is the grouping of 12 African isolates into three closely additional groups (IX, X and XI). The 13th African isolate (DaCDI 37) significantly grouped with group VIII (subgroup a). Among the three sequences from Sri-Lanka isolates, one (DeSLK 1) clustered with the group IV (JPN 2 from Japan) with a low RP and a non-significant BP value. The two other sequences grouped significantly with the group VIII, subgroup c. Similarly, the additional isolate from Martinique (Mart 6) clustered with group VIII. An additional clustering, supported by high MP and RP values, was evidenced between isolates DaVan 1 from Vanuatu and Guad 6 from Guadeloupe (group VI).

3.3. Evidence of extensive homologous recombination among YMMV

Conflicting phylogenetic signals in the sequence data were detected by conducting various phylogenetic reconstruction analyses. In addition to the phylogeny of the complete CP (Fig. 2), we show the most significant results obtained with the sequences coding for the C-terminal region of NIb and the contiguous N-terminal part of the CP (Fig. 4A), the core and the C-terminal part of the CP (Fig. 4B) and the 3′-NCR (Fig. 4C). All gave basically the same relative phylogenetic groupings except for the 3′-NCR. The inconsistency of phylogenetic grouping in the 3′-NCR is more related to a low phylogenetic signal and to a difference in the evolutionary histories between the coding and the non-coding region rather than to recombination events. However, recombination events may be evidenced by the analysis of conflicting groupings supported by high PM and RP values.

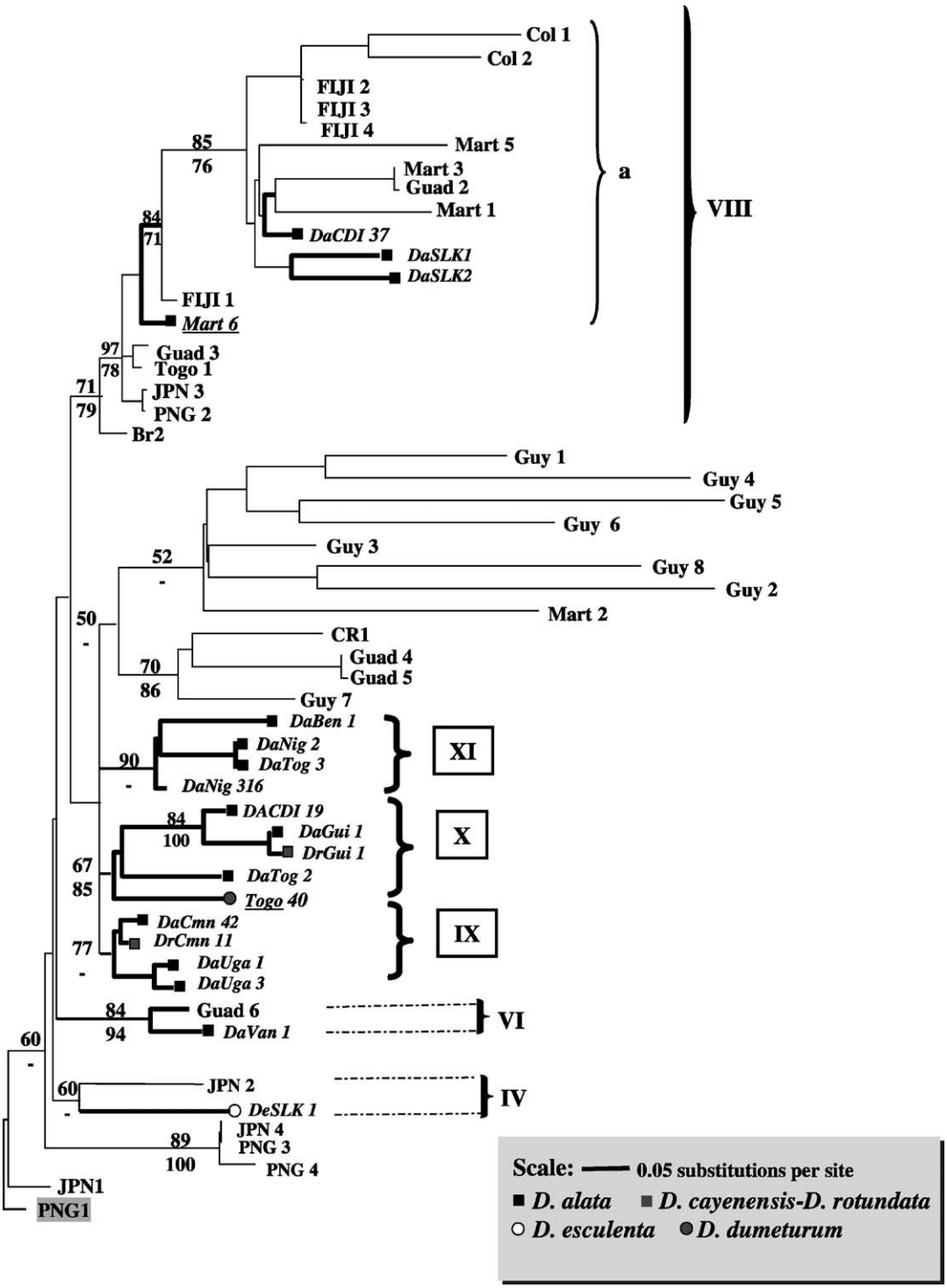

Fig. 3. Phylogenetic analysis of the partial CP core region and the 3′-NCR of 54 YMMV isolates. Phylogram was produced by the quartet puzzling maximum likelihood (ML) approach on nucleotide sequences beginning at the WCIEN motif of the CP up to the 3′-NCR of 54 YMMV isolates, including the 36 previously analysed isolates (in bold), two isolates partially sequenced (in italic and underlined) and 18 sequences (in bold and italic) retrieved from Genebank database (Table 1). The taxon used as outgroup (i.e. PNG 1 isolate) is indicated in grey. Values at the nodes (RP) are indicated above branches. Maximum parsimony analysis was conducted on the same nucleotide matrix and the robustness of the nodes was assessed by BP (below branches). Branch lengths are proportional to the number of substitutions inferred. The new isolates are indicated by thick bars and the corresponding phylogenetic groups are labelled horizontally. Only the modified groups are indicated and the new phylogenetic groups (groups IX–XI) are boxed. The host origin of the new isolates is mapped on the phylogram and the legend is reported on the figure.

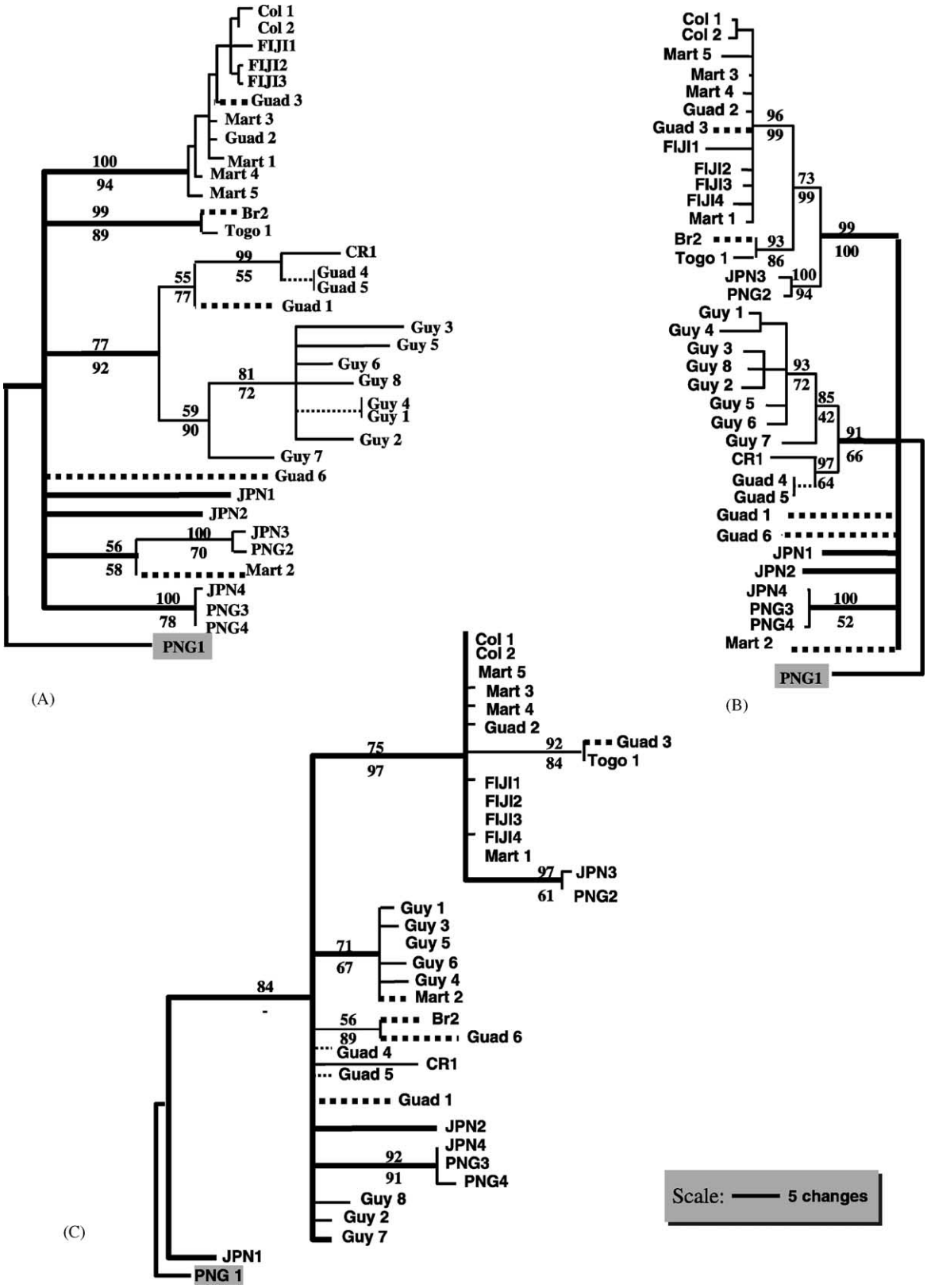

Fig. 4. Assessment of recombination events by comparative phylogenetic analysis of the C-ter of NIb + the N-ter of CP part (A); core + C-Ter of the CP (B) and the 3′-NCR (C) of 36 isolates of YMMV. Phylograms were reconstructed from the nucleotide sequences by maximum parsimony (BP value above branches) and by maximum likelihood (RP value below branches). The taxon used as outgroup (i.e. PNG 1 isolate) is boxed in grey. Branch lengths are proportional to the number of substitutions inferred and all phylograms are drawn to the same scale. Recombinant isolates are indicated with the dotted bars.

Distinct topological positions according to the part of the sequences analyzed were observed for three isolates from Guadeloupe (Guads 1, 3 and 6), one isolate from Martinique (Mart 2) and one isolate from Brazil (Br2). The robustness of these differences was demonstrated by high PM and RP values, (Fig. 4; Table 3).

The analyses performed with the Phylpro, Geneconv and RDP programs provided concordant results for the majority of the sequences tested and supported the hypothesis that the conflicting phylogenetic signals previously observed were due to recombination events (Table 3). Moreover, the three programs detected two other isolates (Guads 4 and 5 from Guadeloupe) as potential recombinants that have not been detected previously by the topology of the phylogenetic trees. These two isolates are closely related (0.1% nt divergence) and were considered as two samples of the same isolate.

Additional analysis was conducted on a second data set, including the 17 sequences of the partial CP core and the 3′-NCR as well as the 36 previous sequences partitioned on the same region. An African isolate (DaCDI 19 from Ivory Coast) was indicated as a potential recombinant (Table 3).

In most of the cases, the closest putative parental sequences were identified and belonged to the groups VII and VIII (Table 3). These sequences are implicated in inter- or intra-group recombination events. The recombination occurred at different locations of the region analyzed, including one or multiple recombination breakpoints. These sites of recombination were identified in the core of the CP (Guad 6), in the C-terminal part of the CP (Br2, Mart 2) and in the 3′-NCR (Guads 3 and 6), (Table 3). The Guad 1 isolate was the only one not detected as a recombinant by both Geneconv and RDP but the analysis of the phylogeny revealed distinct topologies from the different portions of genes chosen confirmed by the phylpro program. Multiple sites were involved (located mostly in the N-ter part and core of the CP), leading to a very complex mosaic pattern with no clear parental sequences always identified.

To avoid any misleading interpretation in the topology of the tree due to the recombinants, we performed an analysis excluding these isolates. No significant changes were obtained (data not shown).

3.4. Structure of the YMMV population in relation with yam species and geographical origin

The different phylogenetic groups and recombinant isolates are presented with their corresponding hosts and geographical origins in Table 4. The sampling may appear skewed but it is explained by obvious epidemiological reasons. Our study focused mainly on the three main yam species *D. alata*, *D. cayenensis–D. rotundata* and *D. trifida*. A comparative analysis of the prevalence of YMMV on these species was possible only in the Caribbean and South America, where *D. trifida* is restricted. The low number of Caribbean and South American isolates from the African species *D. cayenensis–D. rotundata* in our sampling is related to the very low prevalence of YMMV on this species in this region of the world (see Fig. 1). Moreover, the restriction of French Guyana samples to *D. trifida* reflects the epidemiological situation, since neither *D. alata* nor *D. cayenensis–D. rotundata* species were found infected in this region.

Given this sampling situation, different types of relationships are evidenced by the evolutionary pattern of YMMV. The largest emerging monophyletic group is the group VIII. This group illustrates well a lack of phylo-geographical relation, as it contains isolates collected from all the regions sampled. The subgroup “c” contains isolates from Africa, Caribbean, South America, Asia and Oceania whereas the subgroup “b” includes isolates from Brazil and Africa and the subgroup “a” is restricted to Asia and Oceania. This evolutionary pattern contrasts with the one exhibited by the second significant monophyletic group (group VII) for which a clear host adaptation to *D. trifida* is established, with the exception of the subgroup “a” that includes the recombinant Guad 1 as unique member. We cannot conclude that there is a geographical specialization of this group, due to the over-representation of *D. trifida* isolates from French Guyana in our sampling.

There is strong evidence for geographical clustering of isolates from Africa, as supported by the three African groups (IX–XI). The samples cover the most cultivated yam species in Africa (*D. alata* and *D. cayenensis–D. rotundata*) and an indigenous wild yam *D. dumeturum*. No strict regional relation was found among these African isolates. The isolates from West Africa (Nigeria, Benin, Togo, Ivory Coast and Guinea) cluster together (group X and XI) or with the isolates from East Africa (Fig. 3).

The evolutionary pattern of the isolates from Japan and Papua New Guinea is particular. These isolates are positioned into five distinct groups. Four of these groups constitute the most basal groups in the phylogram (Fig. 2). The 5th is the sister group (subgroup a) of the major monophyletic group VIII. A close association between Japan and Papua New Guinea isolates is supported by the group III and the subgroup “a” of the group VIII. Interestingly, the remaining isolates represent distinct individual clades and do indeed represent distinct genotypes.

3.5. Sequence diversity analyses

To quantify the amount of variation in the nucleic acid that results in variation in the encoded CP protein and to understand the impact of selection, patterns of nucleotide substitution were analyzed. The number of synonymous (d_s) and non-synonymous (d_n) substitutions per site was preliminary estimated by different methods as described in Section 2. The obtained values were very close to one another with a slight difference in the estimation of the values at synonymous sites that was overestimated by the Li–Wu–Luo

Table 3
Evidence of recombinant isolates supported by phylogeny and by the programs phylpro, Geneconv and RDP

Isolates	Recombination evidence										Recombination type
	Phylogeny		Phylpro		Geneconv			RDP			
	Parents[a]	Portion gene	Parents	Site(s)[b]	Parents	Pos.[b]	*P*	Parents	Pos.	*P*	
Global analysis with 36 sequences including partial Nib, CP and 3′-NCR											
Guad 3	VIIIc VIIIb	Nib + CP 3′-NCR	VIIIc and VIIIb	1024	VIIIb	780–937	$<10^{-4}$	VIIIc VIIIb	1–781 >781	3×10^{-5}	Intra-group
Br2	VIIIb Ungrouped	Nib + CP Nter core CP Cter + 3′-NCR	VIIIb and VII	785	VIIIb	–	$<10^{-4}$	VIIIb VII	1–786 >786	2.7×10^{-4}	Inter-group Single site
Guad 1	VII Ungrouped	CP Nter CP Cter + 3′-NCR	VIIc, VIIb, VIII	126, 187, 504, 597, 858, 936	–	–	–	–	–	–	Intra- and inter-group Mosaic
Mart 2	Ungrouped VII	Nib + CP CP Cter + 3′-NCR	VIII, IV? VII	241, 399, 498, 806	VII	819–856	4×10^{-2}	Unknown VII	1–792 >792	5×10^{-4}	Inter-group Mosaic
Guad 6	Ungrouped[c]	Full sequence	VIII, VII, II	206, 339, 438, 860	Un-known	>748	3×10^{-4}	–	–	–	Probable mosaic
Guads 4 and 5	–	–[d]	VII and VIII	813	VIII	735–814	7×10^{-3}	Unknown VIII	1–816 >816	1.2×10^{-4}	Intra-group
Supplementary analysis including the partial CP core and the 3′-NCR of 54 isolates											
DaCDI 19	X	Core CP + 3′-NCR	VIIIc and X, XI	>260	X XI	88–155 245–320	2×10^{-3}	–	–	–	Inter-group Single site

[a] Parents correspond to the group of the sequences determined as the potential source of recombination.

[b] Site corresponds to the peak(s) of low phylogenetic correlation in Phylpro. Pos. in Geneconv program corresponds to the significant fragment where recombination events are thought to occur. For RDP program, Pos. indicates the fragments the most related to the one or the other of the parental sequences.

[c] A supplementary analysis including the partial sequences (WCIEN motif up to 3′-NCR) of 54 isolates showed a significant cluster (99% bootstrap) with DaVan isolate (*D. alata* isolate from Vanuatu).

[d] The recombination event was not shown by the different phylogenies reconstructed from the chosen portion genes.

Table 4
Phylogenetic classification of YMMV isolates in relation with their geographic and yam species origin

Phylogenetic and recombination analysis			Origin	
Sequence	Groups	Recombinant	Geographic	Host
C-terminal part of Nib + CP + 3′-NCR[a]	I	–	Papua New Guinea	*D. alata*
	II	Recombinant	Caribbean	*D. alata*
	III	–	Japan, Papua New Guinea	*D. alata*
	IV	–	Japan	*D. alata*
	V	–	Japan	*D. alata*
	VI	Recombinant	Caribbean	*D. esculenta*
	VII (c)	–	South America	*D. trifida*
	VII (b)	–	Central America, Caribbean	*D. trifida*
	VII (a)	1 Recombinant	Caribbean	*D. alata*
	VIII (c)	1 Recombinant	Oceania, Caribbean, South America	*D. alata, D. cay.–rot.*
	VIII (b)	1 Recombinant	South America, Africa	*D. alata*
	VIII (a)	–	Oceania, Asia	*D. alata*
Partial CP + 3′NCR	IX	–	Africa	*D. alata, D. cay.–rot.*
	X	1 Recombinant	Africa	*D. alata, D. cay.–rot. and D. dumeturum*
	XI	–	Africa	*D. alata*

[a] Two distinct phylogenetic analyses were performed: the first one was performed with 37 sequences comprising partial Nib + CP + 3′-NCR. The second one included partial CP and 3′-NCR sequences retrieved from Genebank.

method (Li, 1997) and underestimated by the modified Nei and Gojobori method (Mumford and Seal, 1997). Thus, for the rest of the analysis, the Pamelo and Bianchi (Pamelo and Bianchi, 1993) and Li (Li, 1993) method was used to allow the comparison with the published data.

The mean pairwise value of nucleotide diversity for the whole set of 36 YMMV CP sequences was 0.140. Comparatively with seven other potyviruses (total of 333 sequences), YMMV exhibits, with BYMV, the largest diversity (Table 2). This diversity was about two times larger than the lowest diversity among potyviruses, observed for the SCMV (Table 2). This high nucleotide diversity of YMMV was correlated to a high d_s value and to a low d_n value comparatively to the other potyviruses. Thus, the resulting d_n/d_s was very low. On the contrary, the low diversity of SCMV was related to a lower d_s and to a higher d_n values than observed for other potyviruses, resulting in a d_n/d_s ratio >1. The other potyviruses analyzed had ratios between 0.1 and 0.18 with intermediate d_n and d_s values.

Inter-group YMMV analysis showed diversity values ranging from 0.130 to 0.198 (Table 2). The smallest value was observed for the Japanese groups (V and IV) while the largest implicated the major groups VIII (0.198 with group VI) and VII (0.196 with group I). Interestingly, the average of the inter-group diversity values showed by the *D. trifida* group VII (0.176) was slightly larger than the one showed by the cosmopolitan YMMV group VIII (0.168). Between these minimal and maximal values, the genetic diversity shown by the different groups was of a similar order of magnitude, indicating significant genetic differences between these phylogenetic groups and that correlated well with the phylogenetic analysis.

Inter-group diversity values were consistently larger than those resulting from intra-group comparisons. This might indicate population differentiation, i.e. differences among populations were larger than differences within population. The diversity within the groups VII and VIII was more than two times smaller (0.096 and 0.065, respectively) than the maximum inter-group divergence value. In contrast, the diversity value within the group III (isolates from Papua New Guinea and Japan) was five times smaller (0.004).

The same potyvirus sequence data was used to assess YMMV diversity by comparison of the mean pairwise percentage of divergence and the maximum divergence of CP amino acid sequences. The sequence comparisons of the full length CP of 36 isolates of YMMV revealed a maximum of divergence of 12.4% with an average of 6.2% (Table 2). This value corresponded to an intermediate diversity value comparatively to the high average diversity value of YMV (11.5%), to the maximum divergence value of PPV (26.2%) and the low diversity of SCMV (average 5.2%, maximum divergence 10%). Intra- and inter-YMMV amino acid divergence analysis, showed, as for the nucleotide diversity analysis, a relative equal distribution of the diversity between the different groups, with a maximum for the groups VII and VIII.

A complementary analysis was performed using the 3′-NCR as a second molecular marker. Compared to eight other potyviruses (103 sequences), the YMMV 3′-NCR appeared one of the most variable with an average value of 9.37% and a maximum of divergence of 22.7%, just after the PVY which presented the largest diversity (results not shown).

4. Discussion

4.1. New insights on Yam mild mosaic disease outbreaks in the Caribbean and South America: evidence for the cosmopolitan trait of YMMV

The analyses of the distribution and the prevalence of YMMV provided new insights on the disease outbreaks in yam in this area: YMMV shows a wide repartition in Guadeloupe, Martinique and French Guyana but occurs with different prevalence since *D. trifida* appears significantly more infected than *D. alata* and *D. cayenensis–D. rotundata*. This high prevalence of YMMV on the relative new host *D. trifida* would suggest the selection of a specific YMMV population with a better fitness to invade rapidly this species. The absence of infections on *D. alata* and *D. cayenensis–D. rotundata* in French Guyana, the centre of origin of *D. trifida*, would also support this hypothesis. At this point, the apparent host adaptation of the isolates of group VIII to *D. trifida* does not allow us to conclude to a strict host specificity, that can be shown only by biological experiments. However, in the case of another potyvirus, the *Plum pox virus* (PPV), the anisotropy of distribution of the two main PPV-D and PPV-M phylogenetic groups on apricot and peach species (Bousalem et al., 1994) was not related to a strict pathotype but rather to a myriad of biological factors, each of them bringing a better fitness to invade one or the other *Prunus* species (Dallot et al., 1998; Quiot et al., 1995).

YMMV was already known to infect *D. alata* and *D. cayenensis–D. rotundata* (Bousalem and Dallot, 2000; Odu et al., 1999), *D. japonica* and *D. opposita* (Fuji et al., 1999, 2001), *D. numelarie* and *D. Penthaphila* (Bousalem and Loubet, unpublished results). This study constitutes the first report of YMMV infection on the American yam species *D. trifida*, on the Asiatic-Oceanian species *D. esculenta* and *D. transversa* and on the African yam *D. dumetorum*. In addition, computation of the geographical origin of the YMMV isolates analyzed in this study (Table 1) and those previously reported (Dallot et al., 2001; Fuji et al., 1999, 2001; Mumford and Seal, 1997; Odu et al., 1999) provided the evidence of the world-wide distribution and of the cosmopolitan character of this virus.

4.2. Bio-geographical evidence of the link between the pattern of distribution of YMMV and the geographical origin, dispersion and current distribution of D. alata

The reconstruction of the YMMV history by analysis of its current distribution is inextricably linked to the history of the genus *Dioscorea*. The current pattern of yam species distribution is dominated by the world-wide distribution of *D. alata*, the African and tropical American distribution of *D. cayenensis–D. rotundata* and the restriction of *D. trifida* to tropical America, its area of origin (Coursey, 1976; Orkwor, 1998; Hahn, 1987; Degras, 1993). A global analysis clearly indicates a link between the pattern of YMMV dispersal and the origin, migration and current dispersion of *D. alata* (Fig. 5). East to west movement of yam species is well known and there is no significant corresponding transfer of American species to the Old Word (Davies, 1968; Coursey, 1976; Orkwor, 1998; Hahn, 1987; Degras, 1993). The Asiatic yam *D. alata* probably originated in tropical Burma and Thailand and spread from Southeast Asia to India and the Pacific Ocean more than 2000 years ago (Coursey, 1976). The major centre of diversification and dispersion of *D. alata* appears to be now Papua New Guinea (Martin and Rhodes, 1977; Lebot, 1999; Coursey, 1976). *D. alata* reached the East coast of Africa from Malaysia at about 1500 AD with Portuguese and Spanish seafarers (Coursey, 1976; Hahn, 1987) and was transferred to the Caribbean from West Africa, with the African species *D. cayenensis–D. rotundata*, by the slave trade in the 16th century. Direct introductions of *D. alata* from South-east Asia to the Caribbean had also probably occurred in the 16th century but also in the 20th century.

4.3. Phylogenetic support of the Asian-Pacific origin of YMMV

Our phylogenetic analysis gives a consistent support of the Asian-Pacific origin of YMMV from *D. alata* species. This hypothesis corresponds to the most parsimonious scenario of the evolution of YMMV deduced from phylogenetic analysis and from the current dispersion of the virus. Phylogeny of the complete nucleotide sequences of the YMMV CP shows eight distinct genetic clades that share a common ancient ancestor (recombinant isolates excluded).

The basal positions in the phylogenetic tree are occupied by individual clades (I, III, IV and V) that have Japan, Papua New Guinea or both origins. In addition, two other isolates (JPN 3 and PNG 2) from this same geographical zone occupy the most ancestral node (subgroup a) of the main YMMV group (VIII). This cosmopolitan group includes isolates showing a world-wide origin: Asia, Pacific, Caribbean, Tropical America and Africa and illustrates well the pattern of migration of *D. alata*. In conformity with this parsimonious scenario, the African clades and the *D. trifida* clade probably arose by bio-geographical diversification.

4.4. High accumulation of synonymous mutation and purifying selection act in the evolution of YMMV

Analyses of nucleotide substitution accompanying the YMMV evolution shows that YMMV is the potyvirus that has the most variable CP nucleotide genetic diversity. YMMV CP diversity is due to an accumulation at a higher level of synonymous mutations and the resulting amino acid diversity has been found moderate, especially by comparison with YMV and PPV. For these two viruses, the most divergent values are consistent with the shift of host or/and a geographical diversification. Inter and intra-YMMV analyses are in concordance with these observations. The cosmopolitan group VIII and group VII linked to *D. trifida* are

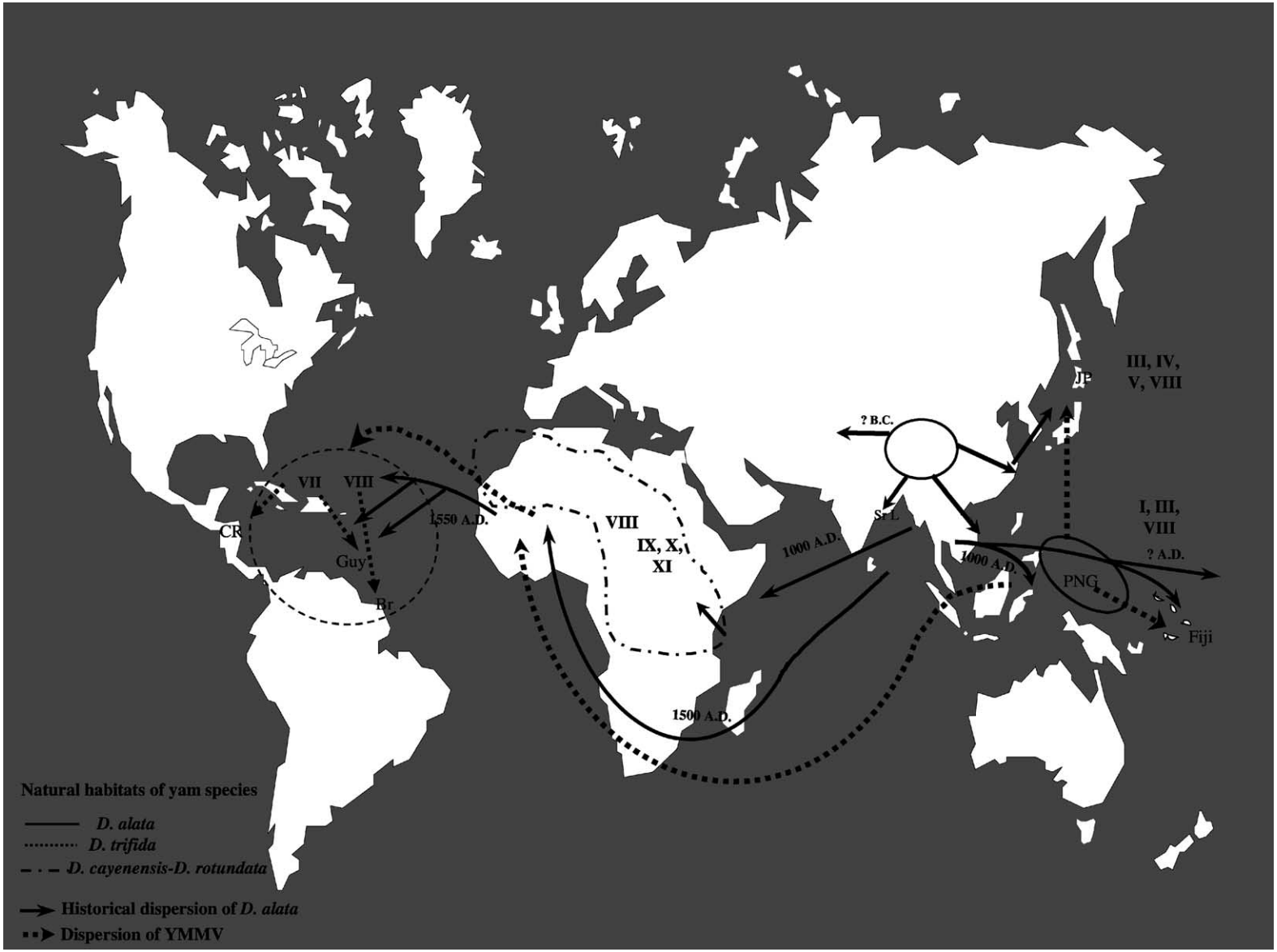

Fig. 5. World-wide dispersion of the phylogenetic groups of YMMV in relation with the natural origin habitats of three main yam species and the dissemination of *D. alata* (modified and completed from Coursey, 1976). Natural habitats of yam species are represented by dotted line for *D. trifida*, by stripe-dotted line for the complex *D. cayenensis–D. rotundata* and by a continuous line for *D. alata*. Plain arrows indicate historical dissemination of *D. alata*. The approximate date of dissemination is indicated above the arrows (AD: "Anno domini"). Introductions of *D. alata* directly from South-east Asia and Oceania to the Caribbean and South America might have also occurred, since the 16th century and in the 20th. Dotted-line arrows indicate the hypothesized world-wide movement of YMMV and roman numbers indicate the YMMV phylogenetic groups evidenced.

the most divergent. However, a large fraction of the total YMMV diversity is also distributed among the different phylogenetic groups, indicating other sources of diversification.

Rapid divergence driven by positive selection has been rarely demonstrated at the molecular level, since it can be confounded with variation at RNA and protein sites that tolerate genetic drift (Fares et al., 2001). The estimation of the d_n/d_s ratio was used to evaluate the hypothesis of an evolutionary response of the YMMV CP to accommodate new host species (as *D. trifida*) or geographical adaptation. We have found that purifying selection dominates the evolution of the CP of potyviruses and strongly operates on the YMMV. However, the overall negative selection can mask punctual positive selection operating on a few positions as evidenced recently with potyviruses (Moury et al., 2002). The d_n/d_s values found for the studied potyviruses are in the same range as those reported for RNA and DNA viruses infecting animals and plants (Li, 1997; Garcia-Arenal et al., 2001). Positive selection is considered to be less frequent than negative selection (Gojobori et al., 1990; Sala and Wain-Hobson, 1999). The most compelling evidence for positive selection is provided by the human and animal viruses as a consequence of the immune response (Nielsen, 1999).

The identification of positive selection acting on the SCMV CP was thus not expected. This feature could be probably related to the host and geographical adaptation that operate for this virus (Alegria et al., 2003). In the other hand, the potyviruses d_s values change little as they could be near to saturation. As suggested by Sala and Wain-Hobson (1999), the reliable interpretation of d_n/d_s ratios is possible only when the degree of nuclei acid divergence is small. Indeed, analysis of nucleotide diversities data for some virus genes (Garcia-Arenal et al., 2001) shows that higher d_n/d_s values are more correlated to low d_s value than to high d_n value.

4.5. Recombination events as a source of YMMV lineage differentiation

The sequence of the six recombinants evidenced in our data set were obtained by direct sequencing, suggesting that these recombinants represent the master population of the viral quasi-species with a selective advantage over parental genomes. Perhaps most intriguingly, the results demonstrate that the majority of the closest putative parental sequences belonged to the majors groups VII and VIII and that these extremely divergent groups are linked by recombination into a single gene pool as a source of a new distinct YMMV lineages (groups II and VI). Comparable results have been reported for three YMV lineages (Bousalem et al., 2000b). In recent years, recombination in natural populations has been reported for several plant viruses (Simon and Bujarski, 1994; Aranda et al., 1997; Revers et al., 1996; Cervera et al., 1993; Glasa et al., 2001, Ohshima et al., 2002) and it has been proposed that recombination can be advantageous for RNA viruses. Our data, indeed, constitute a powerful evidence that recombination is an important evolutionary strategy for potyviruses.

The current geographical distribution of YMMV recombinants in tropical America and particularly in the Caribbean could be explained by the sympatric distribution of the two potentially parental genotype groups VII and VIII. In case of allopatric distribution, such as in French Guyana where the group VII is restricted to *D. trifida*, intra-group recombination could be an alternative strategy of diversification.

4.6. Switching events, radiation, host and geographical adaptation and recombination events: major traits of YMMV evolutionary history

In the absence of positive selection, mechanisms like aphid sampling inducing bottlenecks and founder effects associated to a new bio-geographical environment have more likely contributed, with the recombination events, to the formation of the distinct YMMV lineages.

Phylogenetic analyses revealed the presence of multiple clades with most having a single taxon. The connection between the clades is not resolved and the overall topology of the YMMV phylogenetic tree is star-like, suggesting an early split of the genetic lineages. An important part of the clades is constituted by a single lineage arisen by recombination that could reflect fairly recent events between different YMMV lineages. The creation of individual lineage is also consistent with the hypothesis of radiation by geographical isolation due to the island trait of the area of origin. This is compatible with a star phylogeney produced by this lineage indicating that, under the assumption of a constancy rate, all lineages diverged at the same time (Kimura, 1983). The amino acid and nucleotide substitutions among this lineage are as expected very close and conform to a star phylogeny.

Sampling might have produced an “Iceberg bias” and thus providing a partial picture of the total phylogenetic diversity of the virus. To address this possibility, a complementary study about YMMV in relation with other Asian-Oceanian yam species (*D. numelaria*, *D. penthephila*, *D. bulbifera*, *D. japonica*, *D. hispida*) will be developed.

This explanation is applicable to the origin of the African groups, for which both geographical diversification and host diversification involving the wild African *Dioscorea* species are possible.

The second main group (group VII) is restricted to the Caribbean and tropical America. The restricted geographical area of this later group is related to the geographical centre of diversification of *D. trifida*. We hypothesize that the group VII arose by host selection and adaptation. The strict specificity of this group to *D. trifida* has not been biologically established yet. An alternate hypothesis may be related to the theory of an ecological specialization, which generally assumes that radiation of a parasite into a novel host affects its ability to compete on the original host (Turner and Elena, 2000). The cost associated with host radiation in a

RNA virus model is compatible with this theory (Chao et al., 1977; Novella et al., 1995; Crill et al., 2000; Turner and Elena, 2000). In this view, *D. trifida* adaptation could be a recent property in comparison with a host switches event. A similar situation involving the YMV originating from Africa and *D. trifida* was described (Bousalem et al., 2000b) and the strict correlation of host-lineage was recently confirmed (Bousalem, unpublished results).

It is important to note that the genetic diversity of the *D. trifida* group is high comparatively to the major group VIII. We attributed this observation to the fact that the size of the quasi-species (i.e. level of genetic diversity in a viral population) is not constant but rather varies depending on the host environment, and that the viral quasi-species is selected for a particular level of variation specific to the virus-host combination (Schneider and Roossinck, 2001). This suggests the possibility that different hosts may accelerate or decelerate the rate of viral evolution by permitting or denying high level of diversity in a viral population.

Evidence of host-adaptation is clear with only *D. trifida* for both YMMV and YMV (Bousalem et al., 2000b). *D. trifida* belongs to the most divergent yam *Macrogynodium* section while the African yam *D. rotundata–D. cayenensis* and the Asiatico-Oceanian *D. alata* are close and belong to the same section *Enantiophyllum* (Coursey, 1976).

4.7. An evolutionary epidemiological approach to understanding regional and world-wide dispersion of YMMV: global and regional time scale

The correlation between the migration of the YMMV and the migration of *D. alata* (and occasionally *D. esculenta*) would allow us establishing a time scale of the major evolutionary events: (1) the YMMV ancestor occurred in the Asian-Oceanic area where the major yam *D. alata* is originated, (2) a first diversification of YMMV by rapid radiation in the South Pacific islands where *D. alata* diversity is greater, (3) introduction of *D. alata* in Africa at about 1500 AD: selection, diversification and formation of the YMMV African lineage, (4) migration in the 16th century of the YMMV from Africa to the Caribbean via both *D. alata* and *D. cayenensis–D. rotundata* and transfer onto a novel host, the American species *D. trifida*, inducing the selection and the formation of a new lineage, (5) this lineage and the major and cosmopolitan group (VIII) generated variability through recombination, (6) fast colonization of the Caribbean via aphid transmission, (7) transfer of new lineage in South America via tubers of *D. trifida*, (8) colonization of the French Guyana via both vegetative and aphid transmission, (9) acceleration of the epidemics due to the intensive introduction and exchange of yam cultivars in the 1960s, (10) current prevalence and dissemination of YMMV as presented in the first part of this study.

A scale of a potyvirus dissemination rate after its introduction to an island is provided by the example of Zucchini yellow mosaic virus. The introduction of ZYMV to Martinique island may have occurred in 1992 either through importation of infected plants or seeds, and has been considered to be widespread only 7 years later (Desbiez et al., 1996).

Due to the increase in human migrations and to climatic changes, emerging and re-emerging viruses constitute one of the major challenges for virologists in this new century. The use of a concept such as “integrated genetic epidemiology” for interpreting molecular data is now accepted in the domain of human and animal infectious diseases (Tibayrenc, 1999) but remains limited for plant viruses. Our evolutionary epidemiological approach may lead to novel ways of analyzing and predicting the risk of emergence/re-emergence of new virus/strain in a global disease management approach.

Acknowledgements

The authors wish to thank Pr. Garcia-Arenal, Dr. J.B. Quiot and the two anonymous referees for their critical review of the manuscript and valuable suggestions. The technical assistance of S. Loubet and the contribution of Dr. M. Hilf to improve the English of this manuscript are gratefully acknowledged. We wish to thank: G. Ano, R. Arnolin (INRA Guadeloupe); R. Corranson (SPV Martinique), and G. Tiego (SPV Guyane) for assistance in the collect of samples. This work was supported by the Institut de Recherche pour le Développement (IRD). We are grateful to J.L. Pham for giving us the facility to complete this study within the research unit “DYNADIV”.

References

Alegria, O.M., Royer, M., Bousalem, M., Chatenet, M., Petterschmitt, M., Girard, M., Rot, P., 2003. Genetic diversity in the coat protein coding region of eighty six sugarcane mosaic virus isolates from eight countries, especially from Cameroon and Congo. Arch. Virol. 148, 357–372.

Aranda, M.A., Fraile, A., Dopazo, J., Malpica, J.M., Garcia-Arenal, F., 1997. Contribution of mutation and RNA recombination to the evolution of a plant pathogenic RNA. J. Mol. Evol. 44, 81–88.

Bousalem, M., Dallot, S., 2000. First report and molecular characterization of *Yam mild mosaic virus* on *Dioscorea alata* in the island of Martinique. Plant Dis. 84, 91005.

Bousalem, M., Candresse, T., Quiot-Douine, L., Quiot, J.B., 1994. Comparison of three methods for assessing *Plum plox virus* variability: further evidence for the existence of major groups of isolates. J. Phytopathol. 142, 163–172.

Bousalem, M., Dallot, S., Guyader, S., 2000a. Using phylogenetic data to develop molecular tools for the detection and genotyping of *Yam mosaic virus*. Potential application in molecular epidemiology. J. Virol. Meth. 90, 25–36.

Bousalem, M., Douzery, E.J.P., Fargette, D., 2000b. High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution. J. Gen. Virol. 81, 243–255.

Cervera, M.T., Riechmann, J.L., Martin, M.T., Garcia, J.A., 1993. 3′-Terminal sequence of plum pox virus PS and o6: evidence for RNA recombination within the potyvirus group. J. Gen. Virol. 74, 329–334.

Chao, L., Levin, B.R., Stewart, F.M., 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology 58, 369–378.

Coursey, D.G., 1976. Yams. *Dioscorea* sp. (*Dioscoreaceae*). In: Simmonds, N.W. (Ed.), Evolution of Crop Plants. Simmonds-Longman Group Limited, London, pp. 70–74.

Crill, W.D., Wichman, H.A., Bull, J.J., 2000. Evolutionary reversals during viral adaptation to alternating hosts. Genetics 154, 27–37.

Dallot, S., Labonne, G., Quiot-douine, L., Boeglin, M., Candresse, T., Quiot, J.B., 1998. Peculiar plum pox potyvirus D-populations are epidemic in peach trees. Acta Hortic. 472, 355–365.

Dallot, S., Bousalem, M., Gudzman, M., 2001. Occurrence of potyviruses on Yam (*Dioscorea* spp.) in Colombia and first molecular characterization of *Yam mild mosaic virus*. Plant Dis. 85, 803.

Davies, O., 1968. Origins of African agriculture. 1. The origins of agriculture in West Africa. Curr. Anthropol. 9 (5), 425–479.

Degras, L., 1993. The yam. A tropical root crop. In: Coste, R. (Ed.), London and Basingstoke. Macmillan/CTA, New York.

Desbiez, C., Wipf-Scheibel, C., Granier, F., Robaglia, C., Delaunay, T., Lecoq, H., 1996. Biological and molecular variability of zucchini yellow mosaic virus on the island of Martinique. Plant Dis. 80, 203–207.

Fares, M.A., Moya, A., Escarmis, C., Baranowscki, E., Domingo, E., Barrio, E., 2001. Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. Mol. Biol. Evol. 18, 10–21.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Fitch, W.M., 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Sys. Zool. 20, 406–416.

Freeman, G.H., Halton, J.H., 1951. Note on an exact treatment of contingency, goodness of fit and other problems of significance. Biometrika 38, 141–159.

Fuji, S., Mitobe, I., Nakamae, H., Natsuaki, K.T., 1999. Nucleotide sequence of coat protein gene of yam mild mosaic virus, isolated in Papua New Guinea. Arch. Virol. 44, 1415–1419.

Fuji, S., Furuya, H., Naito, H., Natsuaki, K.T., 2001. Detection of potyviruses in yam plants cultivated in Japan. Jpn. J. Phytopathol. 67, 261–263.

Garcia-Arenal, F., Frail, A., Malpica, J.M., 2001. Variability and genetic structure of plant virus population. Annu. Rev. Phytopathol. 39, 157–175.

Gibbs, A., 1999. Evolution and origins of tobamovirus. Proc. Phylo. Trans. Soc. Lond. B. 354, 517–685.

Glasa, M., Kudela, O., Marie-jeanne, V., Quiot, J.B., 2001. Evidence of naturally occurring recombinant isolate of Plum pox virus from Slovakia. Plant Dis. 85, 920.

Gojobori, T., Moriyama, E.N., Kimura, M., 1990. Molecular clock of viral evolution and the neutral theory. Proc. Nat. Acad. Sci. U.S.A. 87, 10015–10018.

Goudou-Urbino, C., Givord, L., Konate, G., Boeglin, M., Quiot, J.B., Dubern, J., 1996. Differentiation of yam viruses isolates by using symptomatology, western-blot and monoclonal antibodies. J. Phytopathol. 144, 235–240.

Hahn, S.K., 1987. Yams. In: Evolution of Crop Plants. Smartt and Simmonds, Essex, England, pp. 112–120.

Ina, Y., 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J. Mol. Evol. 40, 190–226.

Kimura, M., 1980. A simple method for estimating rate of base substitutions through comparative studies of nucleotide sequence. J. Mol. Evol. 16, 111–120.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, AZ, USA.

Lebot, V., 1999. Biomolecular evidence for plant domestication in Sahul. Genet. Resour. Crop Evol. 46, 619–628.

Li, W.H., 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitution. J. Mol. Evol. 36, 96–99.

Li, W.H., 1997. Molecular Evolution. Sinauer, Sunderland, MA, 487 pp.

Li, W.H., Wu, C.I., Luo, C.C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2, 150–174.

Martin, F.W., Rhodes, A.M., 1977. Intra-specific classification of *Dioscorea alata*. Trop. Agric. (Trinidad) 5, 1–13.

Martin, D.P., Rybicki, E.P., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Mehta, C., Patel, N., 1995. StatXact3. Cytel Software Corporation Edit., 788 pp.

Moury, B., Morel, C., Johansen, E., Jacquemond, M., 2002. Evidence for diversifying selection in Potato virus Y and in the coat protein of other potuviruses. J. Gen. Virol. 83, 2563–2573.

Mumford, R.A., Seal, S.E., 1997. Rapid single-tube immunocapture RT-PCR for the detection of two yam potyviruses. J. Virol. Meth. 69, 73–79.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nei, M., Kumar, S., 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford, 333 pp.

Nielsen, R., 1999. Changes in d_s/d_n in the HIV-1 *env* gene. Mol. Biol. Evol. 165, 711–714.

Novella, I.S., Elena, S.F., Moya, A., Domingo, E., Holland, J.J., 1995. Size of genetic botlenecks leading to virus firtness loss is determined by mean of initial population fitness. J. Virol. 69, 2869–2872.

Odu, B.B., d'A Hughes, J., Shoyinka, S.A., Dongo, L.N., 1999. Isolation, characterization and identification of a potyvirus from *Dioscorea alata* L., (water yam) in Nigeria. Ann. Appl. Biol. 134, 65–71.

Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., Sano, T., Azyhata, F., Walsh, J.A., Fletcher, J., Chen, J., Gera, A., Gibbs, A., 2002. Molecular evolution of turnip mosaic virus: evidence of host adaptation, genetic recombination and geographical spread. J. Gen. Virol. 83, 1511–1521.

Orkwor, G.C., 1998. The importance of yams. In: Orkwor, G.C., Asiedu, R., Ilkanayake, I.J. (Eds.), Food Yams: Advances in Research, IITA-NRCRI, Ibadan, Nigeria, pp. 1–12.

Pamelo, P., Bianchi, N.O., 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. 10, 271–281.

Porth, A., Lesemann, D.E., Vetten, H.J., 1987. Characterization of potyvirus isolates from West African yams (*Dioscorea* spp.). J. Phytopathol. 120, 160–183.

Posada, D., Crandall, K., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulation. Proc. Natl. Acad. Sci. 24, 13757–13762.

Quiot, J.B., Labonne, G., Boeglin, M., Adamolle, C., Renaud, L.Y., Candresse, T., 1995. Behavior of two isolates of plum pox virus inoculated on peach and apricot trees: first results. Acta Hortic. 386, 290–297.

Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., Dunez, J., 1996. Frequent occurrence of recombinant potyvirus isolates. J. Gen. Virol. 77, 1953–1965.

Roossinck, M.J., 1997. Mechanisms of plant virus evolution. Annu. Rev. Phytopathol. 35, 191–209.

Sala, M., Wain-Hobson, S., 1999. Drift and conservatism in RNA virus evolution: are they adapting or merely changing? In: Domingo, E., Webster, R., Holland, J. (Eds.), Origin and Evolution of Viruses, Academic Press, London, UK.

Sawyer, S.A., 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–556.

Schneider, W.L., Roossinck, M.J., 2001. Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. J. Virol. 74, 3130–3134.

Shukla, D.D., Ward, W.W., Brunt, A.A., 1994. The *Potyviridae*. CAB International, Wallingford, UK.

Simon, A.E., Bujarski, J.J., 1994. RNA–RNA recombination and evolution in virus infected plants. Annu. Rev. Phytopathol. 32, 337–362.

Strimmer, K., Von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Swofford, D.L., 1998. PAUP: Phylogenetic Analysis Using Parsimony, Version 4, Sinaur Associated, Sunderland, MA, USA.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

Thouvenel, J.C., Fauquet, C., 1979. Yam mosaic, a new potyvirus infecting *Dioscorea cayenensis* in the Ivory Coast. Ann. Appl. Biol. 93, 279–283.

Tibayrenc, M., 1999. Toward an integrated genetic epidemiology of parasitic protozoa and other pathogens. Annu. Rev. Genet. 33, 449–477.

Turner, E., Elena, S.F., 2000. Cost radiation in an RNA virus. Genetics 156, 1465–1470.

Weiller, G.F., 1998. Phylogenetic profiles: a graphical method for detecting genetic recombination in homologous sequences. J. Mol. Evol. 15, 326–335.