

High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution

M. Bousalem,¹ E. J. P. Douzery² and D. Fargette¹

¹Laboratoire de Phytovirologie des Régions Chaudes, CIRAD/IRD, BP 5035, F-34032 Montpellier Cedex 1, France

²Laboratoire de Paléontologie, Paléobiologie et Phylogénie, Institut des Sciences de l'Évolution (UMR 5554/CNRS), Université Montpellier II, Place E. Bataillon, F-34095 Montpellier Cedex 5, France

To evaluate the genetic diversity and understand the evolution of *Yam mosaic virus* (YMV), a highly destructive pathogen of yam (*Dioscorea* sp.), sequencing was carried out of the C-terminal part of the replicase (NIb), the coat protein (CP) and the 3'-untranslated region (3'-UTR) of 27 YMV isolates collected from the three main cultivated species (*Dioscorea alata*, the complex *Dioscorea cayenensis*–*Dioscorea rotundata* and *Dioscorea trifida*). YMV showed the most variable CP relative to eight other potyviruses. This high variability was structured into nine distant molecular groups, as revealed by phylogenetic analyses and validated by assessment of the molecular evolutionary noise. No correlation was observed between the CP and 3'-UTR diversities and phylogenies. The most diversified and divergent groups included isolates from Africa. The remaining groups clustered in a single clade and a geographical distinction between isolates from the Caribbean, South America and Africa was observed. The role of the host in the selection of particular isolates was illustrated by the case of a divergent cultivar from Burkina Faso. Phylogenetic topological incongruence and complementary statistical tests highlighted the fact that recombination events, with single and multiple crossover sites, largely contributed to the evolution of YMV. We hypothesise an African origin of YMV from the yam complex *D. cayenensis*–*D. rotundata*, followed by independent transfers to *D. alata* and *D. trifida* during virus evolution.

Introduction

Potviruses (genus *Potyvirus*, family *Potyviridae*) constitute the largest and economically most important genus of plant viruses (Bos, 1992; Shukla *et al.*, 1994). They seem well adapted to the intensive, modern agriculture of temperate regions, but also flourish in crops cultivated in more traditional ways in the tropics. Thus, it is important to understand how potyviruses have evolved and the origin of the variability observed. Some elements of comprehension can be obtained by the study of *Yam mosaic virus* (YMV) as a potyvirus model.

YMV is a highly destructive pathogen of yam (*Dioscorea* sp.) (Thouvenel & Fauquet, 1979; Thouvenel & Dumont,

1990). YMV has been identified in all the areas of production (Africa, the Caribbean, Latin America and the South Pacific) and found in several species of *Dioscorea*. Yams are characterized by a very large diversity (603 species; Degras, 1993), including a large number of wild taxa (Hamon *et al.*, 1995). Yam domestication started at least 5000 years ago (Dumont, 1982) and the process is continuing, providing new sources of diversity. These characteristics, as well as its vegetative multiplication through tubers, its traditional cropping for centuries and the absence of sanitation, might favour virus evolution and diversification. Data on the evolution of YMV are scarce as variability analyses have been conducted on only a few samples collected from a restricted number of cultivars and species and as complete sequences of the coat protein (CP) and 3'-untranslated region (3'-UTR) are known for only one isolate (Duterme *et al.*, 1996; Aleman *et al.*, 1996; Aleman-Verdaguer *et al.*, 1997).

Author for correspondence: Mustapha Bousalem.

Fax +33 4 67 61 56 03. e-mail bousalem@mpl.ird.fr

RNA virus diversity results from the accumulation of mutations due to frequent errors in RNA synthesis (Domingo & Holland, 1994; Drake, 1993; Roossinck, 1997). Recombination events are also a major evolutionary factor for RNA plant viruses (Simon & Burjarski, 1994; Lai, 1995; Aranda *et al.*, 1997). However, recombination has rarely been observed in natural populations and the frequency of recombinants has been reported for only a few viruses (Aranda *et al.*, 1997; Fraile *et al.*, 1997; Garcia-Arenal *et al.*, 1997). Recombination has previously been demonstrated for potyviruses by analysis of nucleotide sequences retrieved from the GenBank database (Revers *et al.*, 1996), and one recombinant has been found in a natural population of *Plum pox virus* (PPV) (Cervera *et al.*, 1993).

In this paper, we assessed the genetic diversity of YMV and evaluated the relative importance of the accumulation of mutations and RNA recombination in the evolution of this potyvirus in relation to its hosts. With this aim, (i) we sampled 27 YMV isolates on representative hosts including the most widespread species *Dioscorea alata* L., which is of Asian origin, the specific complex *Dioscorea cayenensis* Lam.–*Dioscorea rotundata* Poir, which is of African origin and widely cultivated on this continent, and *Dioscorea trifida* L., which is restricted to its areas of origin, the Amazon and the Caribbean (Degras, 1993); (ii) we analysed the molecular variability of YMV isolates and compared it with that of other potyviruses, by sequencing the contiguous C-terminal part of Nlb (C-Ter), the complete CP gene and the 3'-UTR; (iii) we assessed the molecular phylogeny of YMV, in relation to yam host species and cultivars and the geographical origin of the samples; and (iv) we developed the systematic detection of recombinant isolates.

Methods

■ Virus isolates. Twenty-seven YMV isolates were collected from different yam species and cultivars in several regions of Africa, the Caribbean and French Guiana (Table 1). Our sampling took into account three criteria: (i) the world distribution of the three yam species (see Introduction), (ii) the current situation of yam cropping (90% of the total world production comes from Africa; Degras, 1993) and (iii) the prevalence of YMV on the three species (field prospecting has shown that *D. alata* and *D. cayenensis*–*D. rotundata* are not infected by YMV in French Guiana). Isolates were screened by direct double-antibody sandwich-ELISA with monoclonal antibodies raised against YMV strain 112 as performed by Goudou-Urbino *et al.* (1996). Tubers were cropped in insect-proof greenhouses. All isolates were propagated in *Nicotiana benthamiana* before analysis.

■ RT-PCR, cloning and DNA sequencing. Total RNA from infected *N. benthamiana* was extracted as described by Verwoerd *et al.* (1989). Reverse transcription was performed on extracted RNAs by using M-MLV reverse transcriptase with oligo(dT) primer. Nlb C-Ter and the complete CP and 3'-UTR region were amplified. PCRs were carried out on cDNA with 2 U Vent DNA polymerase in 100 µl, 100 pM of each primer and 0.2 mM of each dNTP. PCR was performed by first heating at 94 °C for 4 min, followed by 35 cycles at 94 °C for 30 s, 52–60 °C for

30 s and 72 °C for 1–2 min and then one cycle of elongation at 72 °C for 10 min. On the basis of the complete sequence of YMV 112 (Aleman *et al.*, 1996), primers specific to Nlb C-Ter were designed: P2/2YMV, 5' GAGCTCTCTACACTAATCAAGC 3', and P2/1YMV, 5' ATGGTC-ATGCTCGCCATG 3' (respectively positions 986–1004 and 1447–1469 of the Nlb gene). The downstream primer was an oligo(dT) primer or primer P4YMV (5' TTTTCTCCTCCGCCACATC 3') specific to the YMV 112 3'-UTR region. Two other primers that annealed in the CP (positions 450–471) were used occasionally: P3/2YMV, 5' GATGG-TGTGGTGCATAGAAAAC 3', and P1/2YMV, 5' GTTTTCTATGC-ACCACACCA TC 3'. PCR fragments were purified from agarose gels and sequenced either directly or after cloning into pCR-Script Amp SK(+) vector (Stratagene) with Epicurian coli XL-1 Blue MRF' Kan supercompetent cells used for transformation. Sequencing was carried out with the *Taq* terminator sequencing kit (Applied Biosystems) and analysed on an Applied Biosystems 373A sequencer. When direct sequencing provided unreadable sequences, probably due to mixed infection of the yam host, sequencing after cloning was performed for isolates C1/C3, CKA1/C11, B1/C1, CBE6b/C3, CAM1/C1, SOAAi/C1, CAM2/C31, GY/INRA/C11, CGU1/C18 and TRIFIDA/C5.

■ Phylogenetic analyses. A region of 1184 nucleotides (nt) was sequenced for 27 isolates, including the contiguous 3'-terminal region of Nlb (108 nt), the full CP (912 nt) and the 3'-UTR (164 nt). The complete Nlb-CP-3'-UTR DNA and CP amino acid (aa) sequences were aligned and analysed by using the MUST package (Philippe, 1993). Within CP, we distinguished (i) the N-terminal region (N-Ter), encoding the hyper-variable N-terminal extension of the CP polypeptide (nt 1–189, aa A¹–T⁶³), (ii) the conserved core (nt 190–879, aa H⁶⁴–E²⁹¹) and the C-Ter region (nt 880–912, aa R²⁹²–M³⁰³).

Phylogenetic reconstructions were obtained by two complementary methods: (i) maximum parsimony (MP) (PAUP 3.1.1; Swofford, 1993), with all molecular characters assessed as independent, unordered and equally weighted (Fitch, 1971); and (ii) maximum likelihood (ML) (Felsenstein, 1981), using the quartet-puzzling method (PUZZLE 4.0; Strimmer & von Haeseler, 1996), with the Tamura & Nei (1993) model of sequence evolution and with a four-category gamma distribution of parameter α to describe substitution-rate heterogeneities (Yang, 1996).

Robustness of the nodes of the phylogenetic trees was assessed by three different approaches: (i) the bootstrap (Felsenstein, 1985), yielding bootstrap percentages (BP) for each node, computed after 1000 resamplings followed by an MP reconstruction (bootstrap option in PAUP 3.1.1, with one random sequence addition per replicate and MAXTREES = 100); (ii) the Bremer support index (BSI), the number of extra nucleotide or amino acid substitutions required to collapse the corresponding node (Bremer, 1988), computed by the AUTODECAY 3.0 program (Eriksson, 1995); and (iii) the reliability percentage (RP), the number of times the group appears after 10000 ML puzzling steps (PUZZLE 4.0; Strimmer & von Haeseler, 1996).

To understand the pattern of nucleotide substitution in the CP gene better, the homoplasy and saturation of the two transitions and four transversions at each of the three codon positions were evaluated following the MP procedure of Hassanin *et al.* (1998). Accordingly, we measured the consistency index (CI) and the slope (S) of the regression between the number of observed nucleotide differences and the number of inferred nucleotide changes for 18 substitution types. When the CI and S values are close to 1 (or close to 0), the corresponding substitution type is almost not (or strongly) homoplastic and saturated.

■ Recombination analyses. Recombination events were suspected after performing comparisons of phylogenies reconstructed from the Nlb,

Table 1. Phylogenetic groups of YMV isolates in relation to country, species and cultivar of origin and date of sampling

Isolate	Collecting site	Date of sampling	<i>Dioscorea</i> species*	Cultivar	Accession no.
Group I					
C1/C3	Arbolle, Burkina Faso	1992	<u>cayenensis</u> –rotundata	Pilimpikou yam	AJ244053
BFC 51/C11	Arbolle, Burkina Faso	1992	<u>cayenensis</u> –rotundata	Pilimpikou yam	AJ244050
BFC 54	Arbolle, Burkina Faso	1992	<u>cayenensis</u> –rotundata	Pilimpikou yam	AJ244051
BFC 56	Samba, Burkina Faso	1992	<u>cayenensis</u> –rotundata	Pilimpikou yam	AJ244052
Group II					
112†	Adiopodoumé, Ivory Coast	1977	<u>cayenensis</u> –rotundata	Nandokaka	U42596
CID3/C12	Adiopodoumé, Ivory Coast	1994	<u>cayenensis</u> –rotundata	Kpan biankouma	AJ244058
CKA1/C11	Adiopodoumé, Ivory Coast	1994	<u>cayenensis</u> –rotundata	Gbangan	AJ244059
POGNON/C1	Guadeloupe (Fr)‡	1994	<u>cayenensis</u> –rotundata	Pognon	AJ244064
Group III					
B1/C1	Goua, Benin	1994	<u>cayenensis</u> –rotundata	Baniakpa	AJ244048
B14	Savalou, Benin	1994	<u>cayenensis</u> –rotundata	Laboko	AJ244049
CBE6b/C3	Bariba, Benin	1994	<u>cayenensis</u> –rotundata	Worou binsi	AJ244056
CAMI/C1	Bamileke, Cameroon	1994	<u>cayenensis</u> –rotundata	Shoufing	AJ244054
Group IV					
SOA AI/C1	Ouangolodougou, Burkina Faso	1992	<i>alata</i>	Florido	AJ244065
SOA2/C2	Mangodara, Burkina Faso	1992	<i>alata</i>	Florido	AJ244066
CAM2/C31	Dourou, Cameroon	1994	<u>cayenensis</u> –rotundata	Tii	AJ244055
174/C1	Dompago, Benin	1994	<u>cayenensis</u> –rotundata	Makpawa	AJ244046
Group V					
G5/C10	North-east French Guiana (Fr)‡	1992	<i>trifida</i>	Cousse couche violet	AJ244062
G13/C1	North-east French Guiana (Fr)	1992	<i>trifida</i>	Cousse couche violet	AJ244061
GY/INRA/C11	North-west French Guiana (Fr)	1991	<i>trifida</i>	Cousse couche violet	AJ244045
Group VI					
CGU1/C18	Guadeloupe (Fr)	1994	<u>cayenensis</u> –rotundata	Corossol	AJ244057
CGU2/C4	Guadeloupe (Fr)	1994	<u>cayenensis</u> –rotundata	Grosse caille	AJ244044
GR/SAVANE/C4	Guadeloupe (Fr)	1991	<u>cayenensis</u> –rotundata	Grande savane	AJ244063
TRIFIDA/C5	Guadeloupe (Fr)	1986	<i>trifida</i>	INRA 5-20	AJ238401
AID 10/5	Puerto Rico	1994	<i>alata</i>	Gémélos	AJ244043
Group VII					
608	Nigeria	1992	<u>cayenensis</u> –rotundata	TDR 608	AJ244047
Group VIII					
DIVIN	Guadeloupe (Fr)	1989	<i>alata</i>	Divin no. 1	AJ244060
Group IX					
CAM2	Dourou, Cameroon	1994	<u>cayenensis</u> –rotundata	Tii	AJ244042

* When distinction between the two components of the *D. cayenensis*–*D. rotundata* complex was possible, the relevant component is underlined (P. Hamon, personal communication).

† Isolate sequenced by Aleman *et al.* (1996).

‡ Guadeloupe and French Guiana are Départements d'Outre-Mer of France.

N-Ter CP, core CP and 3'-UTR regions. Three complementary statistical approaches were performed on the alignment of nucleotide sequences to maximize the probability of validation of the suspected recombination events. (i) The likelihood that genetic rearrangements had occurred was estimated by VTDIST (Sawyer 1989). This program omits conserved positions ('condensation') and defines a set of fragments between successive informative sites for each sequence pair. Values of the sum of squares for the condensed fragments (SSCF) and the maximum length of the condensed fragments (MCF) were then compared after 10 000 random permutations of sequences to provide an estimate of the

likelihood of genetic rearrangements. (ii) The recombination break-points were tentatively localized by using the maximum χ^2 approach described by Maynard-Smith (1992), as implemented in RECSITE (Revers *et al.*, 1996). (iii) The recombinant sequences and the location of recombination junctions were detected by using PHYLPRO (version beta 0.8) developed by Weiller (1998). This program determines the pairwise distances of all aligned sequences at each position within a split window (optimal size for our data was 60 columns). To reduce the weight of rare point mutations, only phylogenetically informative sites were used to calculate the pairwise distances (parsimonious option). The correlation between these

distances ('phylogenetic correlation') ranged from +1 (perfectly correlated) to 0 (unrelated). They were plotted for all positions in a single graph to obtain the 'phylogenetic profiles', where recombination signals appeared as single, sharp downward peaks.

Results and Discussion

Assessment of the molecular variability of YMV showed high diversity in the CP gene

We assessed the molecular variability of the CP gene and the 3'-UTR, which are commonly used as markers of genetic relatedness of potyviruses (Frenkel *et al.*, 1989, 1992; Shukla *et al.*, 1994). Sequence comparisons of the CP of 27 YMV isolates revealed divergence of 0–28.4% (nt) and 0–18.5% (aa). As with all potyviruses, this variability is most marked in the N-terminal region (Shukla *et al.*, 1994). N-Ter sequence comparisons revealed divergence of 11.6–45.0% (nt) and 15.9–57.1% (aa) and the core region showed divergence of 6.1–17.5% (nt) and 1.7–11.7% (aa).

This high level of variability of YMV CP was confirmed by a comparison between the intraspecies divergence of eight other potyviruses (177 CP sequences: Fig. 1). YMV was the potyvirus with the most variable CP, with an average divergence of 11.5% (aa).

Sequence comparisons of the 3'-UTR of 27 YMV isolates revealed divergence of 0–19.1%. The comparison of the intraspecies divergence of the YMV 3'-UTR with those of six other potyviruses (106 sequences) showed that the YMV 3'-UTR (mean divergence of 7.6%) was not the most variable among the potyviruses (Fig. 1). *Potato virus Y* (PVY) presented the largest diversity, followed by *Bean yellow mosaic virus* (BYMV) and YMV.

The high divergence values of the YMV CP placed some isolates between the strains of one potyvirus and closely related potyviruses in the tetramodal distribution of the family *Potyviridae* as described by Shukla *et al.* (1994) and Ward *et al.* (1995). Some PPV isolates also showed the same intermediate position, due to three individual isolates infecting particular host species, highlighting the role of the host in the selection of variants. In contrast to PPV, the high divergence of YMV is global and probably reflects a more complex evolutionary process, the major components of which are assessed below.

YMV variability is structured in distant phylogenetic relationships

To avoid any misleading conclusions due to the choice of phylogenetic reconstruction method, a 1184 nt region (Nlb C-Ter, CP and 3'-UTR) of 27 YMV isolates was analysed by two complementary approaches, the equally weighted MP and the ML. A preliminary analysis was carried out with two distinct and closely related potyviruses as outgroups for YMV [PVY N strain, as used by Aleman-Verdaguer *et al.* (1997), and *Turnip mosaic virus* QUE strain]. On the basis of 483 unambiguously

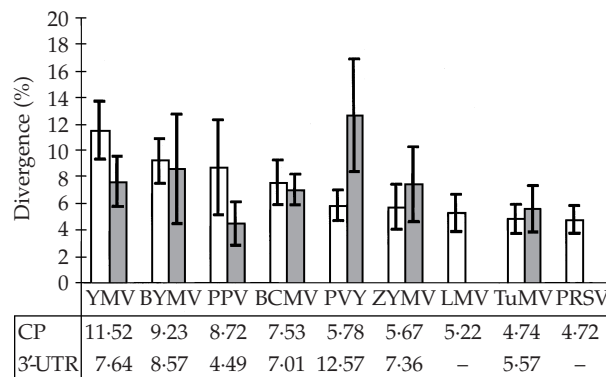


Fig. 1. High YMV variability shown by comparison between percentage divergence of CP amino acid sequences (open bars) and 3'-UTR nucleotide sequences (shaded bars) of nine distinct potyviruses. The mean pairwise percentage divergence between sequences of different potyvirus isolates is represented with its standard deviation. CP comparisons involved YMV (27 isolates), BYMV (9 isolates), PPV (26 isolates), *Bean common mosaic virus* (BCMV, 20 isolates), PVY (45 isolates), *Zucchini yellow mosaic virus* (ZYMV, 10 isolates), *Lettuce mosaic virus* (LMV, 11 isolates), *Turnip mosaic virus* (TuMV, 14 isolates) and *Papaya ringspot virus* (PRSV, 15 isolates). Comparisons based on the 3'-UTR involved YMV (30 isolates), BYMV (8 isolates), PPV (25 isolates), BCMV (11 isolates), PVY (14 isolates), ZYMV (9 isolates) and TuMV (9 isolates). The number of 3'-UTR sequences available for LMV and PRSV was too low to be included in the analysis. Accession numbers are available from the first author.

alignable nucleotides of the CP core region, this analysis suggested that BFC56, BFC51, BFC54 and C1/C3 clustered together into group I and that they are the sister group of all isolates. On the basis of these results, subsequent reconstructions used group I as the outgroup.

Six additional major groups (groups II–VI and IX) were then indicated by high branch support (BP = 100 and BSI ranging from +14 to +46) (Fig. 2a). The relative positions of isolates 608 and DIVIN were not well established and they were assigned to groups VII and VIII, respectively. The relationships between the major groups were not robustly resolved, except a basal position of CAM2 (group IX) and a sister-group position of group II relative to groups III–VIII.

The six previously identified groups were recovered after ML analysis (Fig. 2b) with significant RP, and taxa 608 and DIVIN clustered tentatively with groups VI (RP = 77) and V (RP = 86), respectively. Relative branch lengths also showed that groups I, II and IX were very divergent relative to the other isolates.

Phylogenetic relationships of YMV isolates based on the CP gene showed the same groups as for the complete Nlb-CP-3'-UTR sequences (Fig. 3). Analysis of CP amino acid sequences also yielded the same robust results (not shown). Pairwise sequence comparisons (Table 2) confirmed the great divergence of groups I, II and IX, which showed the highest intragroup (up to 3.4% nt and 4.5% aa) and intergroup (17.9–22.0% nt and 14.8–18.3% aa) divergence. One should note that analysis of the C-Ter region of Nlb detected these

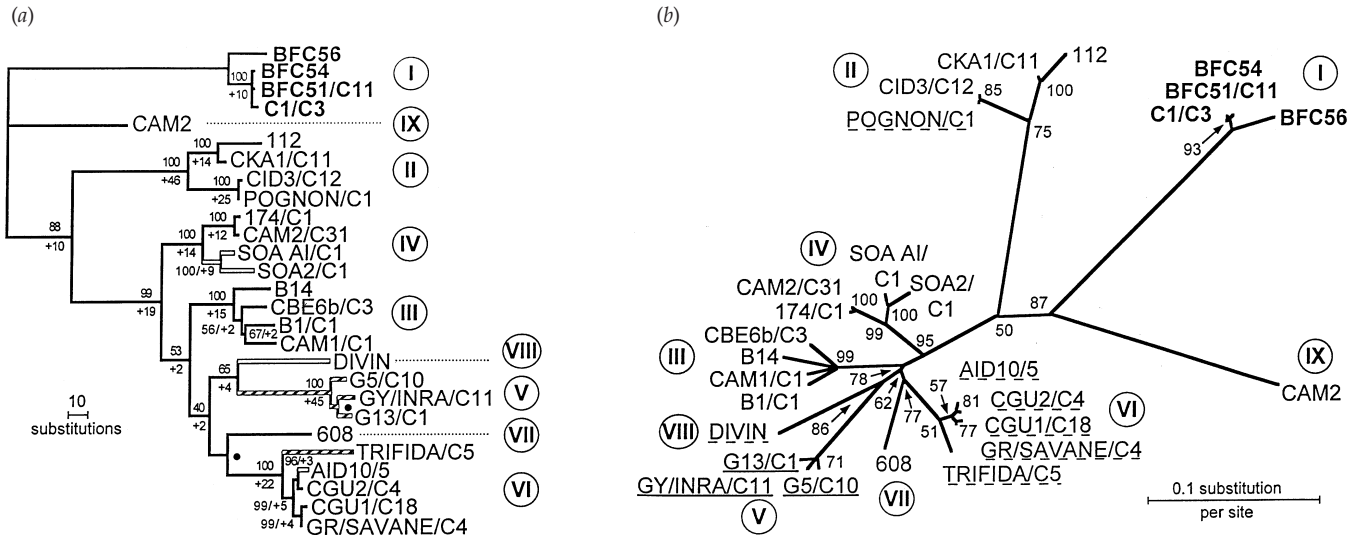


Fig. 2. (a) Phylogenetic evidence from MP of nine major groups of YMV (I–IX). One of the nine MP phylograms reconstructed from 1184 nucleotide positions of the N1b–CP–3′-UTR region of 27 YMV isolates is shown [492 variable and 372 phylogenetically informative sites; length = 1012 substitutions; CI (excluding uninformative characters) = 0.55, retention index = 0.80]. Values above branches are BP and values below are BSI (number of extra substitutions required to break the corresponding node). Nodes collapsed in the strict consensus of the nine trees are indicated by dots. Taxa used as the outgroup (i.e. group I) are indicated in bold. Branch lengths are proportional to the number of substitutions inferred. The host origin of the isolates is shown on the phylogram: *Dioscorea alata* (open bars), *D. trifida* (hatched bars) and the *D. cayenensis*–*D. rotundata* complex (filled bars). (b) Phylogenetic evidence from ML of nine major groups of YMV (I–IX). The highest-likelihood phylogram produced by the quartet puzzling ML approach on the same nucleotide matrix is shown ($\ln L = -6131.74$). Parameters used for ML calculations were 7.24 and 1.81 for substitution-rate ratios of transitions to transversions and pyrimidines to purines, respectively, and $\alpha = 0.29$ for rate heterogeneities. Values at the nodes are RP. Taxa used as the outgroup (i.e. group I) are indicated in bold. Branch lengths are proportional to the expected number of substitutions per site. All YMV are from Africa except those originating from the Caribbean (dotted underline) and French Guiana (underline).

three groups with BP close to 100, although the region analysed was short (108 nt; data not shown).

The nucleotide sequences of the N-Ter and core regions of the CP were then analysed separately. ML and MP phylogenies reconstructed from the N-Ter region robustly supported the monophyly of groups II, IV and VI and showed that CAM2 clustered robustly with group II (RP = 93, BP = 76). Phylogenetic analysis of the core region appear more related to the complete CP, and ML and MP analyses provide robust support for the monophyly of groups II–V (Fig. 3). Phylogenetic analysis of the 3′-UTR differed from that of the CP region, as it showed a lack of robust resolution of most nodes except monophyly of group VI, the latter without TRIFIDA/C5, which was tentatively clustered with group V.

The lack of congruence between CP and 3′-UTR can be attributed not only to the restricted length of the 3′-UTR (low phylogenetic signal), but also to the difference in selection pressure between the coding and the non-coding regions and also to recombination events (see below). Within most virus genera, different genes usually have the same phylogenies, indicating that their evolution has been linked, that they have experienced the same speciation events and have co-diverged (Gibbs *et al.*, 1997). In the case of the potyviruses, Aleman (1996) showed that all phylogenetic trees based on every cistron of the potyvirus genome displayed a similar pattern. In

the case of YMV, phylogenetic analysis carried out by Aleman-Verdaguer *et al.* (1997) on six isolates showed the same topology whichever gene was studied, HC, N1b, P1, P3 or the partial CP region. These data consolidate our phylogenetic analysis based on the CP. Our results indicate that, in contrast, assessing phylogenetic relationships of potyviruses with the 3′-UTR is not recommended.

Low molecular evolutionary noise in the CP of YMV validates variability and phylogenetic analyses

Pairwise sequence comparisons to estimate genetic diversity as well as phylogenetic reconstruction may be highly sensitive to the evolutionary noise brought about by convergences and reversals (i.e. homoplasy). In protein-coding genes, such as the CP, multiple substitutions are known to occur in the third but also in the second codon positions (Naylor *et al.*, 1995). To account for these homoplastic character-state changes and to evaluate their subsequent saturation through evolutionary time, weighting schemes have been proposed in order to reduce their effect in phylogenetic reconstruction (e.g. Hassanin *et al.*, 1998).

Because of the high rates of spontaneous mutation often reported for RNA genomes, it was especially important to measure the levels of homoplasy and saturation of each

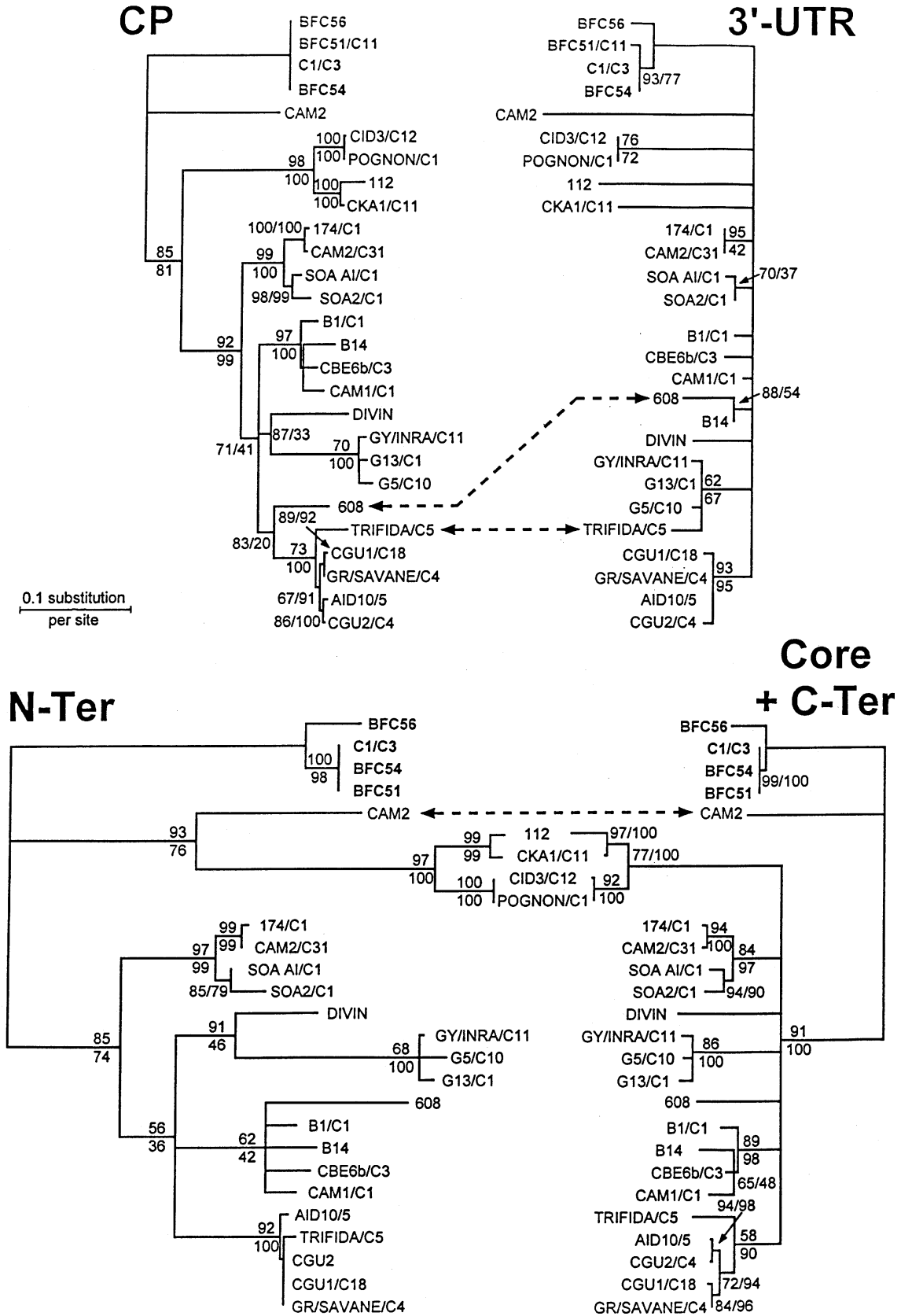


Fig. 3. For legend see facing page.

Table 2. Variability of nine groups of YMV estimated by intra- and intergroup CP sequence divergence comparison

Intergroup sequence divergence is indicated above (nucleotide) and below (amino acid) the diagonal. Intragroup divergence is indicated on the diagonal in bold (aa) and italics (nt).

Group	I	II	III	IV	V	VI	VII	VIII	IX
I	0·3–4 0–4·5	18·8–22	13·9–15·4	15·6–16·9	17·7–19·4	16·3–18·5	18·9	17·5–18·3	17·9
II	14·8–18·7	0·1–6·3 0·6–7·6	17·3–20·6	17·3–19·8	20·5–22·4	18·9–21·1	18·6–19·5	16·8–20·7	18·8–20·3
III	12·4–15·5	13·2–15·9	3–4·4 2·2–4	11·3–8·9	11·1–12·8	8·9–11·7	8·5–9·7	11·2–12·8	17·4–19·2
IV	13·4–15·2	12·3–15·6	6·3–9·3	0·2–4·4 0·1–3·3	12·3–14·3	9·9–12·7	11·2–12·3	11–11·9	20·5–20·9
V	15·9–17·6	15·2–18·5	10·3–13·2	10·6–11·6	0·2–1·2 2·3–3	11·2–13·3	11·2–12·9	12·2–12·8	17·8–19·9
VI	13·8–17·8	14·6–16·9	6·6–11·3	7·3–11·3	10·9–12·6	0·6–4·4 0·1–3·3	8·3–10·6	11–12	18·4–18·7
VII	15·5–18·9	13·9–16·2	6·3–7·6	7–7·9	10·9–11·9	6–9·6	–	10·3	18·3
VIII	15·6–17·2	15·6–18·3	10·3–11·6	9–9·6	14–15	8·6–11·3	9·6	–	20·7
IX	16·9–17·2	14·9–16·6	15·2–15·9	18·5–18·9	15·2–18·5	14·6–15·9	16·2	17·9	–

Table 3. Homoplasy and saturation of the two transitions and four transversions at each codon position in the CP coding sequence

Values separated by slashes correspond to the homoplasy indices (above the diagonal) and saturation indices (below the diagonal) at each of the three codon positions. The numbers of informative sites for the six substitutions were: A/C = 14/16/23, A/G = 45/33/97, A/T = 14/14/40, C/G = 12/11/18, C/T = 20/15/82 and G/T = 15/9/28, for the first, second and third codon positions. Homoplasy of the six different substitutions (for example A–G) was evaluated by using CI excluding uninformative characters, computed on the MP trees reconstructed from a recoded matrix (for example C and T coded as missing data in order to examine A–G substitutions). Additional details may be found in Hassanin *et al.* (1998). Saturation indices were measured by the slopes (S) of the linear regressions between pairwise numbers of observed changes against pairwise numbers of inferred substitutions by MP for the six different substitutions. CI and S close to 1 (or close to 0) indicate that the levels of homoplasy and saturation are very low (or very high).

Residue	A	C	G	T
A	–	0·94/1·00/0·96	0·60/0·64/0·52	0·94/1·00/0·91
C	0·98/1·00/0·99	–	1·00/1·00/1·00	0·55/0·69/0·46
G	0·63/0·77/0·66	1·00/1·00/1·00	–	1·00/1·00/0·96
T	0·99/1·00/0·99	0·48/0·69/0·50	1·00/1·00/0·97	–

substitution type in our 27 YMV CP sequences (Table 3). The substitutions that were most subject to multiple changes were C–T and A–G transitions in the third codon position and, to a

slightly lesser extent, in the first and second positions (range of CI and S values: 0·46–0·77). By contrast, all transversion types at all three codon positions were neither homoplastic nor

Fig. 3. ML phylograms reconstructed from the N-Ter, core plus C-Ter, CP and 3'-UTR regions of 27 YMV isolates suggesting recombination events between some isolates. Phylograms were reconstructed from the N-Ter (189 bp), core + C-Ter (723 bp), CP (912 bp) and 3'-UTR (164 bp) regions by the ML quartet puzzling method and rooted by BFC51, BFC54, BFC56 and C1/C3 (group I). All phylograms are drawn to the same scale to illustrate the increased substitution rates in the N-Ter domain. Robustness of the nodes was assessed by RP (above branches or to the left of the slash). MP analyses were conducted on the same nucleotide matrices and BP corresponding to the nodes revealed by ML are indicated (below branches or to the right of the slash). MP and ML analyses robustly supported different placement of 608, TRIFIDA and CAM2 isolates depending on the region under focus (dotted arrows).

saturated (range: 0.91–1.00). The overall pattern of substitution in the YMV CP showed that, according to the approach of Hassanin *et al.* (1998), there was no strong homoplasy or saturation in the data, thus validating the above-mentioned YMV variability (Table 2) and phylogenetic (Figs 2 and 3) analyses. Results of the MP analysis with character-state changes weighted by the product of CI and S values are not shown, but yielded similar branch support for the groups shown in Fig. 2(a), except for a more robust clustering of isolate 608 with group VI (BP = 73).

Structure of the YMV population in relation to geographical origin and yam host

The geographical origin and yam hosts sampled are recapitulated in Table 1 for the different groups of YMV and can be related to the phylogenies shown in Fig. 2(b) and (a), respectively.

The most divergent groups (I, II and IX; Fig. 2) included isolates from Africa that infect *D. cayenensis*–*D. rotundata*. However, as shown by the phylogenetic position of the POGNON isolate, members of the most diversified group, II (Table 2), are now present in Guadeloupe and illustrate the spread of the virus by human activity. A robustly supported pole included the six other groups, with isolates infecting the three yam species coming from Africa, the Caribbean (Puerto Rico and Guadeloupe) and South America (French Guiana) (Fig. 2b). In this clade, we observed a correlation between the clustering of the isolates and their geographical origin: groups III, IV and VII for Africa, groups VI and VIII for the Caribbean and group V for French Guiana.

A relationship between the phylogeny of the isolates and the yam species is not obvious, as members of the same phylogenetic groups (IV and VI) have been collected from two distinct hosts, *D. cayenensis*–*D. rotundata* and *D. alata* (Fig. 2a). We should also note that the latter yam species may be infected by isolates affiliated to group II, as shown by the inclusion of the CP N-Ter region of isolate AIA from Guadeloupe (Aleman-Verdaguer *et al.*, 1997) in our phylogenetic analysis (not shown). In the case of *D. trifida*, however, a correlation was established with group V, as G5/C10, GY/INRA/C11 and G13/C1 have only been observed on this species (Table 1 and Fig. 2a). Recent molecular and epidemiological data (not shown) obtained from the three yam species in the Caribbean and French Guiana confirmed this observation: isolates of group V collected in both these geographical areas were restricted to *D. trifida*. These results suggest the existence of a pathotype, which needs to be explored by biological assays. The case of isolate TRIFIDA/C5 will be discussed below.

To summarize, the phylogram of Fig. 2(a) suggests that the ancestral yam host for the 27 isolates studied is *D. cayenensis*–*D. rotundata* and that independent transfers to *D. alata* and *D. trifida* have occurred during YMV evolution. Owing to the

taxonomic controversy over the complex *D. cayenensis*–*D. rotundata*, we distinguished between its two components: the perennial and semi-perennial yams of *D. cayenensis* and the annual yams of *D. rotundata* (Hamon & Touré, 1990; Table 1). The three divergent African groups were found on either *D. cayenensis* (groups I and II) or *D. rotundata* (group IX), whereas the other African groups (III and IV) included isolates from both *D. cayenensis* and *D. rotundata*.

Role of the host in selection of particular isolates: the Pilimpikou cultivar in Burkina Faso

A relationship between cultivar and molecular group was only established for group I members. BFC56, BFC51, BFC54 and C1/C3 infect a divergent cultivar (Pilimpikou) of the *D. cayenensis*–*D. rotundata* complex, which may be related to the wild yam species *Dioscorea abyssinica*, *Dioscorea lecardii* and *Dioscorea sagittifolia* (R. Dumont & P. Hamon, unpublished data). In agreement with previous results (Duterme *et al.*, 1996; Aleman-Verdaguer *et al.*, 1997), the CP N-Ter region of these isolates lacked 12 aa (CP positions 37–48), but we could not conclude whether this was an ancestral or a derived deletion. Group I was also characterized by a mutation within the DAG block implicated in transmission by aphids. Three of the four isolates (BFC54, BFC51 and C1) showed a DTG motif instead of DAG. Analysis of other sequences of isolates from the same region (Duterme *et al.*, 1996; Aleman-Verdaguer *et al.*, 1997) confirmed the co-existence of populations with and without DAG. Different mutations in the DAG motif of potyviruses lead to the loss of transmissibility by aphids (Atreya *et al.*, 1995). Several transmission assays have shown the loss of this property from the Pilimpikou isolates (Urbino *et al.*, 1998). The region of Pilimpikou is isolated (200 km from other yam crops) and Pilimpikou is the sole yam cultivar cropped in this area. This situation reveals the adaptation of a natural virus population to a host and to particular ecological conditions.

Phylogenetic topological incongruence and statistical tests highlight recombination events

The comparison between phylogenies reconstructed from either CP and 3'-UTR or N-Ter and core plus C-Ter (Fig. 3) showed an incongruent position of three isolates, TRIFIDA/C5, CAM2 and 608, depending on the genomic area considered. This suggested the occurrence of recombination events during the evolution of YMV.

To evaluate the statistical significance of such events, to search for other recombinations not suggested by the topology of the trees and also to locate recombination break-points, complementary analyses were performed with VTDIST. The associated SSCF probabilities at the 5% level of significance are shown in Table 4. Several isolates showed significant associated SSCF probabilities in VTDIST analysis with the

Table 4. Statistical evaluation of recombinant YMV isolates by VTDIST analysis

In order to identify recombinant isolates, we compared the 27 YMV sequences with the consensus sequences representative of each of the nine groups as well as the sequence of the supposed recombinant TRIFIDA/C5 by using the pairwise option of VTDIST. For each comparison, the associated SSCF probability is shown (see Methods). ns, SSCF value not significant at the 5% level.

Isolate	Consensus sequence									TRIFIDA/C5
	I	II	III	IV	V	VI	VII	VIII	IX	
Group I										
C1/C3	NS	NS	NS	NS	NS	NS	NS	NS	0.0147	NS
BFC51/C11	0.0245	NS	NS	NS	NS	NS	NS	NS	0.0154	NS
BFC54	0.0239	NS	NS	NS	NS	NS	NS	NS	0.0195	NS
BFC56	0.0096	NS	NS	NS	NS	NS	NS	NS	0.0100	NS
Group II										
112	NS	0.0335	NS	0.0349	NS	NS	NS	0.0212	NS	NS
CID3/C12	NS	NS	0.0049	0.0139	0.0011	NS	0.0099	0.0124	NS	NS
CKA1/C11	NS	NS	0.0017	0.0282	0.0015	NS	0.0088	0.0013	NS	NS
POGNON/C1	NS	NS	0.0059	0.0184	0.0013	NS	0.0148	0.0169	NS	NS
Group III										
B1/C1	NS	0.0024	NS	NS	NS	NS	NS	NS	NS	NS
B14	NS	0.0023	NS	NS	NS	NS	NS	NS	NS	NS
CBE6b/C3	NS	0.0104	NS	NS	NS	NS	NS	NS	NS	NS
CAM1/C1	NS	0.0228	NS	NS	NS	NS	NS	NS	NS	NS
Group IV										
SOA Ai/C1	NS	NS	NS	NS	NS	NS	NS	0.0122	NS	NS
SOA2/C2	NS	0.0408	NS	NS	0.027	NS	NS	0.0004	NS	0.0498
CAM2/C31	NS	0.0479	NS	NS	NS	NS	0.0018	0.0420	NS	NS
174/C1	NS	0.0362	NS	NS	NS	NS	0.0007	0.0245	NS	NS
Group V										
G5/C10	NS	0.0007	NS	0.0100	NS	NS	0.0016	0.0003	NS	< 0.0001
G13/C1	NS	0.0022	NS	0.0393	NS	NS	0.0007	0.0005	NS	< 0.0001
GY/INRA/C11	NS	0.0029	NS	NS	NS	NS	0.0300	0.0084	NS	< 0.0001
Group VI										
CGU1/C18	NS	NS	NS	NS	NS	NS	NS	0.0135	NS	0.0001
CGU2/C4	NS	NS	NS	NS	NS	NS	NS	0.0479	NS	0.0006
GR/SAVANE/C4	NS	NS	NS	NS	NS	NS	NS	0.0175	NS	0.0002
AID 10/5	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.0020
Group VII										
608	NS	0.0489	NS	NS	0.0024	NS	–	0.0116	NS	NS
Group VIII										
DIVIN	NS	0.0051	NS	0.0107	0.0011	NS	0.0116	–	NS	0.0018
Group IX										
CAM2	0.0147	NS	NS	NS	NS	NS	NS	NS	–	NS
TRIFIDA/C5	NS	NS	NS	NS	< 0.0001	0.0003	NS	0.0018	NS	–

consensus sequence of group II (Table 4). RECSITE and PHYLPRO analysis did not confirm these results and so it was difficult to consider them as putative recombinants. In particular, in agreement with Candresse *et al.* (1997), we found that the weight of particular point mutations in the subset of very close sequences could sometimes induce unreliable results. We will now discuss in more detail only the recombination events that were highlighted clearly by all the tests used.

The recombinant TRIFIDA/C5 isolate is probably associated with severe epidemics on *D. trifida* from Guadeloupe

The TRIFIDA/C5 isolate showed a strikingly different phylogenetic position depending on the genomic region considered. It was found in group VI in the CP tree (Fig. 3: RP = 73, BP = 100), whereas it shifted to group V in the 3'-UTR

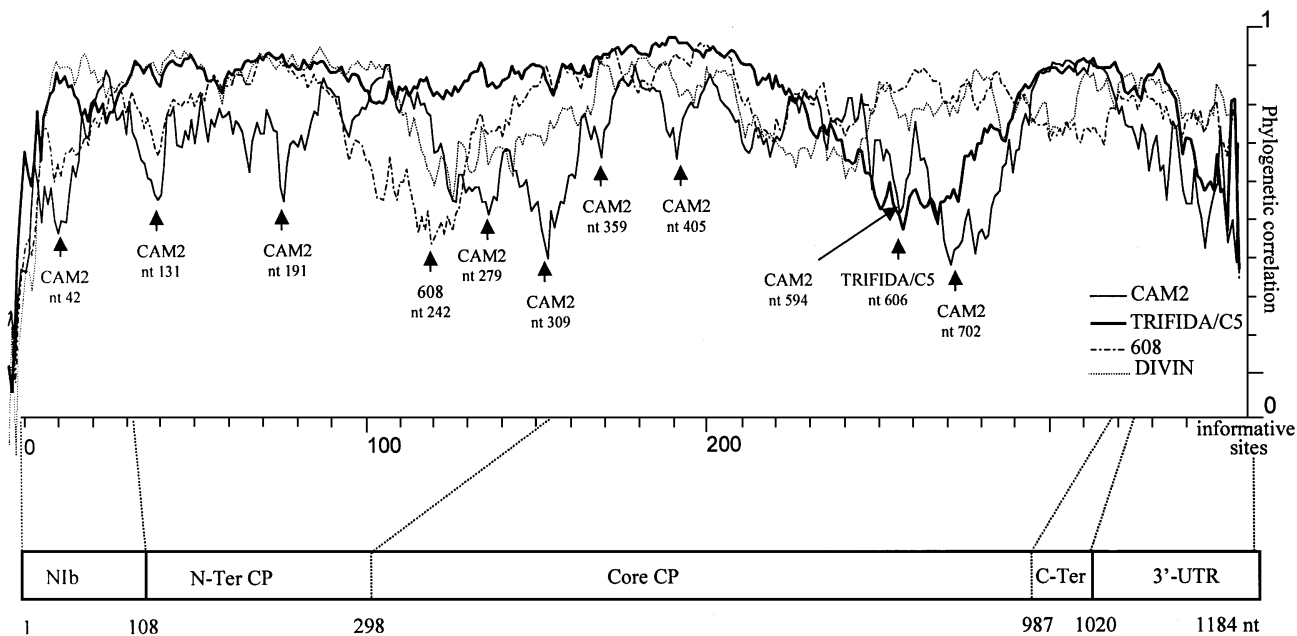


Fig. 4. Four significant recombinant YMV isolates are revealed by PHYLPRO phylogenetic profiles. PHYLPRO analysis was performed on all 27 YMV sequences. Only the phylogenetic profiles of the four significant recombinant isolates are shown. The phylogenetic correlation (y-axis) was obtained at each informative site from pairwise distance analysis of all aligned sequences and ranges from +1 (perfectly correlated) to 0 (unrelated). Recombination signals appear as areas of low phylogenetic correlation, visualized by single sharp downward peaks (see Methods). Their position is indicated along a scale using only the parsimonious informative sites (x-axis of PHYLPRO plot). Below the x-axis, the genomic map of the YMV sequence is shown using all 1184 nucleotides. The positions of the junctions between N1b, CP and 3'-UTR are shown as well as the different domains of the CP. The relative size of the N-terminal region is largest on the phylogenetic profile due to the parsimonious analysis and the greatest relative variability of this CP domain.

analysis (RP = 62, BP = 67). A recombination event was also suggested by VTDIST analysis ($P < 0.001$; Table 4). Pairwise analysis revealed that the parents of TRIFIDA/C5 belonged to groups V and VI, and the recombination break-points were identified with RECSITE ($P < 0.001$) at nt 558–592. PHYLPRO analysis clearly confirmed this result, with the observation of a single sharp downward peak at position 606 (Fig. 4). It also confirmed groups V and VI as the parents of TRIFIDA/C5 by comparison of the percentage divergence from other isolates downstream and upstream from this site.

This isolate was collected on the island of Guadeloupe (1986) during a severe epidemic on the 'Indian' native yam, *D. trifida* (J. B. Quiot, personal communication), which is in regression in the French Caribbean and French Guiana (Degras, 1993). As shown previously (e.g. Fig. 2a; see also discussion of the structure of the YMV population), different molecular strains infect this yam species. The isolates infecting *D. trifida* in French Guiana constitute a distinct phylogenetic group (V). The strain collected on *D. trifida* from Guadeloupe is a recombinant between this group and group VI from Guadeloupe infecting *D. cayenensis*–*D. rotundata* and *D. alata*. The analysis of 10 clones of this *D. trifida* isolate revealed only the recombinant sequence, which suggests a selective advantage for TRIFIDA/C5. Work in progress will allow the role of TRIFIDA/C5 in the severe epidemic on *D. trifida* to be specified.

First evidence of multiple crossover recombination events in natural populations of a plant virus

In the case of the CAM2 isolate, the results are less straightforward. The CAM2 isolate displayed incongruent positions between the N-Ter and core region of the CP trees: it clustered with group II in the N-Ter tree (Fig. 3: RP = 93, BP = 76), but an affinity with groups II–VIII was rejected in the core plus C-Ter tree (RP = 91, BP = 100). Analysis of the C-Ter region of N1b confirmed the latter result, with BP close to 100 (data not shown).

VTDIST results suggested a potential recombination event with isolates of group I, but not with isolates of group II (Table 4). PHYLPRO results showed a more complex situation, with at least nine major peaks (Fig. 4). One peak was located in N1b (nt 42), three peaks were located in the N-Ter region (nt 131, 191 and 279) and five in the core and C-Ter region (nt 309, 359, 405, 594 and 702). PHYLPRO analysis thus revealed a mosaic pattern of recombination. Two origins at least were clearly identified: group II in the N-Ter region of CP (upstream of nt 309) and group I in the core region of CP (upstream of nt 702).

The recombinant CAM2 represents the dominant population, as shown by direct sequencing. However, the sequences obtained after cloning revealed the co-existence of CAM2 in the same plant with the CAM2/C31 isolate, which is clustered

in a different phylogenetic group (Figs 2 and 3). This co-existence with other isolates is probably due to a secondary infection, which could be the source of multiple recombination events and could explain the difficulty in identifying the parents of the recombinant isolate.

To our knowledge, this is the first time that a mosaic pattern of recombination events has been observed in a natural population of a plant virus. This type of recombination has already been described in *Murine hepatitis virus* (MHV), a coronavirus, which contains a non-segmented RNA genome (Keck *et al.*, 1987). In potyviruses, a single and a double recombinant have been detected recently when squash plants were co-bombarded with mixtures of engineered truncated constructs of *Zucchini yellow mosaic virus* (Gal-On *et al.*, 1998).

Both the 608 and DIVIN isolates, with ambiguous phylogenetic status, are recombinant isolates

Isolate 608 clustered with group VI in the CP region (Fig. 3: RP = 83, BP = 20), whereas it joined isolate B14 (group III) in the 3'-UTR region (RP = 88, BP = 54), with contrasting robustness values depending on the phylogenetic approach. However, the results of VTDIST analysis showed evidence of recombination events between 608 and isolates of groups II, IV and V and isolate DIVIN. Recombination involving isolates 174/C1 (group IV) and DIVIN was confirmed by both PHYLPRO (Fig. 4) and RECSITE ($P < 0.001$), but with different positions for the break-points.

The DIVIN isolate was linked to group V in the CP and N-Ter regions (Fig. 3: RP = 87 and 91, respectively). However, phylogenetic conclusions about potential recombination could not be drawn, as the isolate was involved in a multifurcation in the trees derived from the 3'-UTR and core regions. The recombinant status of DIVIN was actually shown by pairwise analysis of VTDIST, with one parent belonging to group V ($0.0003 < P < 0.008$). The other parent could not be identified among our 27 sequences.

These results could explain the ambiguous phylogenetic status not only of isolate 608, but also of isolate DIVIN. An old recombination event followed by the accumulation of mutations on the 'recombinant' sequence might have blurred the molecular signatures of both parents, leading to phylogenetic individualization of the recombinant itself during the evolution process. These two recombinants represent the dominant population, as shown by direct sequencing.

New perspectives on YMV origin and evolution

As demonstrated by our phylogenetic analyses, the African groups I, II and IX were the first to arise during the evolution of the 27 YMV isolates examined and might constitute the major YMV genetic pool. These three groups have a different origin, including an adaptation to a specific host and a multiple recombination event. Our results imply an African origin for YMV from numerous wild species at the origin of the *D.*

cayenensis-*D. rotundata* complex. We can attribute the large diversity observed to the differential accumulation of mutations and the significant contribution of recombination events.

Data on the evolution and history of the yam are essential for a better understanding of the diversification of YMV. In West Africa, man began to gather yams for domestic use as early as 50 000 BC, but true yam-based agriculture started in approximately 3000 BC. The earliest movements of yam have been reported for Asiatic species (*D. alata*) in about 1500 BC from Malaysia to Africa, whereas the African species (*D. cayenensis*-*D. rotundata*) moved westwards as far as America (Coursey, 1976). For this purpose, a study of YMV in relation to wild yam species in Africa and *D. alata* in its area of origin will be developed.

The authors wish to thank Dr O. Le Gall for giving us the opportunity to use the RECSITE program as well as for helpful discussions, Dr G. F. Weiller for providing the PHYLPRO program, Dr A. Hassanin for the development of homoplasy and saturation analyses and Dr S. Dallot for fruitful discussions. We are grateful to Dr J. B. Quiot and Dr M. A. Mayo for their critical review of the manuscript and valuable suggestions.

This work was supported by the Institut de Recherche pour le Développement (IRD). The contribution of E. J. P. D. is no. 99-073 of the Institut des Sciences de l'Évolution de Montpellier, UMR 5554 – CNRS.

References

- Aleman, M. E. (1996). *Caractérisation moléculaire, diversité génétique et contrôle du virus de la mosaïque de l'igname (YMV)*. PhD thesis, École Nationale Supérieure Agronomique de Montpellier, France.
- Aleman, M. E., Marcos, J. F., Brugidou, C., Beachy, R. N. & Fauquet, C. (1996). The complete nucleotide sequence of yam mosaic virus (Ivory Coast isolate) genomic RNA. *Archives of Virology* **141**, 1259–1278.
- Aleman-Verdaguer, M.-E., Goudou-Urbino, C., Dubern, J., Beachy, R. N. & Fauquet, C. (1997). Analysis of the sequence diversity of the P1, HC, P3, NIb and CP genomic regions of several yam mosaic potyvirus isolates: implications for the intraspecies molecular diversity of potyviruses. *Journal of General Virology* **78**, 1253–1264.
- 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. *Journal of Molecular Evolution* **44**, 81–88.
- Atreya, P. L., Lopez-Moya, J. J., Chu, M., Atreya, C. D. & Pirone, T. P. (1995). Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. *Journal of General Virology* **76**, 265–270.
- Bos, L. (1992). Potyvirus, chaos or order? In *Potyvirus Taxonomy* (*Archives of Virology* Supplement 5), pp. 31–46. Edited by O. W. Barnett. Vienna and New York: Springer-Verlag.
- Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**, 795–803.
- Candresse, T., Revers, F., Le Gall, O., Kofalvi, S. A., Marcos, J. & Pallas, V. (1997). Systematic search for recombination events in plant viruses and viroids. In *Virus Resistant Transgenic Plants: Potential Ecological Impact*, pp. 20–24. Edited by M. Tepfer & E. Balazs. Versailles and Heidelberg: INRA and Springer-Verlag.
- Cervera, M. T., Riechmann, J. L., Martín, M. T. & García, J. A. (1993). 3'-Terminal sequence of the plum pox virus PS and óó isolates: evidence

- for RNA recombination within the potyvirus group. *Journal of General Virology* **74**, 329–334.
- Coursey, D. G. (1976).** Yams. *Dioscorea* sp. (*Dioscoreaceae*). In *Evolution of Crop Plants*, pp. 70–74. Edited by N. W. Simmonds. London: Longman.
- Degras, L. (1993).** *The Yam. A Tropical Root Crop*. Edited by R. Coste. London and Basingstoke: Macmillan/CTA.
- Domingo, E. & Holland, J. J. (1994).** Mutation rates and rapid evolution of RNA viruses. In *The Evolutionary Biology of Viruses*, pp. 161–184. Edited by S. S. Morse. New York: Raven Press.
- Drake, J. W. (1993).** Rates of spontaneous mutation among RNA viruses. *Proceedings of the National Academy of Sciences, USA* **90**, 4171–4175.
- Dumont, R. (1982).** Ignames spontanées et cultivées au Bénin et en Haute-Volta. In *Yams – Ignames*, pp. 31–43. Edited by J. Miège & S. N. Lyonga. Oxford: Clarendon Press.
- Duterme, O., Colinet, D., Kummert, J. & Lepoivre, P. (1996).** Determination of the taxonomic position and characterization of yam mosaic virus isolates based on sequence data of the 5'-terminal part of the coat protein cistron. *Archives of Virology* **141**, 1067–1075.
- Eriksson, T. (1995).** AutoDecay version 3.0. Distributed by the author. Department of Botany, University of Stockholm, Sweden.
- Felsenstein, J. (1981).** Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **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. *Systematic Zoology* **20**, 406–416.
- Frail, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M. & Garcia-Arenal, F. (1997).** Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *Journal of Virology* **71**, 934–940.
- Frenkel, M. J., Ward, C. W. & Shukla, D. D. (1989).** The use of 3' non-coding nucleotide sequences in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosaic virus-N. *Journal of General Virology* **70**, 2775–2783.
- Frenkel, M. J., Jilka, J. M., Shukla, D. D. & Ward, C. W. (1992).** Differentiation of potyviruses and their strains by hybridization with the 3' non-coding region of the viral genome. *Journal of Virological Methods* **36**, 51–62.
- Gal-On, A., Meiri, E., Raccach, B. & Gaba, V. (1998).** Recombination of engineered defective RNA species produces infective potyvirus *in planta*. *Journal of Virology* **72**, 5268–5270.
- Garcia-Arenal, F., Alonso-Prados, J. L., Aranda, M., Malpica, J. M. & Frail, A. (1997).** Mixed infection and genetic exchange occur in natural populations of cucumber mosaic cucumovirus. In *Virus Resistant Transgenic Plants: Potential Ecological Impact*, pp. 94–99. Edited by M. Tepfer & E. Balazs. Versailles and Heidelberg: INRA and Springer-Verlag.
- Gibbs, M. J., Armstrong, J., Weiller, G. F. & Gibbs, A. J. (1997).** Virus evolution: the past, a window on the future? In *Virus Resistant Transgenic Plants: Potential Ecological Impact*, pp. 1–19. Edited by M. Tepfer & E. Balazs. Versailles and Heidelberg: INRA and Springer-Verlag.
- Goudou-Urbino, C., Givord, L., Konate, G., Boeglin, M., Quiot, J. B. & Dubern, J. (1996).** Differentiation of yam virus isolates by using symptomatology, western-blot and monoclonal antibodies. *Journal of Phytopathology* **144**, 235–240.
- Hamon, P. & Touré, B. (1990).** The classification of the cultivated yams (*Dioscorea cayenensis-rotundata* complex) of West Africa. *Euphytica* **47**, 179–187.
- Hamon, P., Dumont, R., Zoundjihékpon, J., Touré, B. & Hamond, S. (1995).** *Les Ignames Sauvages d'Afrique de l'Ouest. Caractéristiques Morphologiques* (Wild yams in West Africa. Morphological characteristics). Paris: ORSTOM.
- Hassanin, A., Lecointre, G. & Tillier, S. (1998).** The 'evolutionary signal' of homoplasy in protein-coding gene sequences and its consequences for a priori weighting in phylogeny. *Comptes Rendus de l'Académie des Sciences de Paris, Life Sciences* **321**, 611–620.
- Keck, J. G., Stohman, S. A., Soe, L. H., Makino, S. & Lai, M. M. C. (1987).** Multiple recombination sites at the 5'-end of murine coronavirus RNA. *Virology* **156**, 331–341.
- Lai, M. M. C. (1995).** Recombination and its evolutionary effect on viruses with RNA genomes. In *Molecular Basis of Virus Evolution*, pp. 119–132. Edited by A. Gibbs, C. H. Calisher & F. Garcia-Arenal. Cambridge: Cambridge University Press.
- Maynard-Smith, J. (1992).** Analysing the mosaic structure of genes. *Journal of Molecular Evolution* **34**, 126–129.
- Naylor, G. J., Collins, T. M. & Brown, W. M. (1995).** Hydrophobicity and phylogeny. *Nature* **373**, 565–566.
- Philippe, H. (1993).** MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Research* **21**, 5264–5272.
- Revers, F., Le Gall, O., Candresse, T., Le Romancer, M. & Dunez, J. (1996).** Frequent occurrence of recombinant potyvirus isolates. *Journal of General Virology* **77**, 1953–1965.
- Roossinck, M. J. (1997).** Mechanisms of plant virus evolution. *Annual Review of Phytopathology* **35**, 191–209.
- Sawyer, S. (1989).** Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* **6**, 526–538.
- Shukla, D. D., Ward, W. W. & Brunt, A. A. (1994).** *The Potyviridae*. Wallingford: CAB International.
- Simon, A. E. & Bujarski, J. J. (1994).** RNA–RNA recombination and evolution in virus-infected plants. *Annual Review of Phytopathology* **32**, 337–362.
- Strimmer, K. & von Haeseler, A. (1996).** Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13**, 964–969.
- Swofford, D. L. (1993).** PAUP: phylogenetic analysis using parsimony, version 3.1.1. Champaign, IL: Illinois Natural History Survey.
- Tamura, K. & Nei, M. (1993).** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512–526.
- Thouvenel, J. C. & Dumont, R. (1990).** Pertes de rendement de l'igname infectée par le virus de la mosaïque de l'igname en Côte d'Ivoire. *L'Agronomie Tropicale* **45**, 125–129.
- Thouvenel, J. C. & Fauquet, C. (1979).** Yam mosaic, a new potyvirus infecting *Dioscorea cayenensis* in the Ivory Coast. *Annals of Applied Biology* **93**, 279–283.
- Urbino, C., Bousalem, M., Pinel, A., Fargette, D. & Dubern, J. (1998).** Les virus de l'igname: caractérisation immunologique et moléculaire du virus de la mosaïque de l'igname. In *L'Igname, Plante Séculaire et Culture d'Avenir, Actes du Séminaire International Cirad-Orstom-Coraf*, pp. 205–211. Edited by J. Berthaud, N. Bricas & J. L. Marchand. Montpellier, France: CIRAD.
- Verwoerd, T. C., Dekker, B. M. M. & Hoekema, A. (1989).** A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Research* **17**, 2362.
- Ward, C. W., Weiller, G. F., Shukla, D. D. & Gibbs, A. (1995).** Molecular systematics of the *Potyviridae*, the largest plant virus family. In *Molecular*

Basis of Virus Evolution, pp. 477–497. Edited by A. Gibbs, C. H. Calisher & F. Garcia-Arenal. Cambridge: Cambridge University Press.

Weiller, G. F. (1998). Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Molecular Biology and Evolution* **15**, 326–335.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution* **11**, 367–372.

Received 10 May 1999; Accepted 7 September 1999