Molecular Ecology (2002) **11**, 1219–1229

# A distinct *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda

J. P. LEGG,*† R. FRENCH,‡ D. ROGAN,§ G. OKAO-OKUJA* and J. K. BROWN§
*International Institute of Tropical Agriculture, Eastern and Southern Africa Regional Centre, PO Box 7878, Kampala, Uganda; †Natural Resources Institute, Chatham Maritime, Chatham ME4 4TB, UK; ‡USDA, ARS, Lincoln, NE 68583, USA; §University of Arizona,. Tucson, AZ 85721, USA

**Abstract**

**During the 1990s, an epidemic of cassava mosaic virus disease caused major losses to cassava production in Uganda. Two factors associated with the epidemic were the occurrence of a novel recombinant begomovirus, EACMV-Ug, and unusually high populations of the whitefly vector, *Bemisia tabaci*. Here we present molecular evidence for the occurrence of two cassava-colonizing *B. tabaci* genotype clusters, Ug1 and Ug2, one of which, Ug2, can be consistently associated with the CMD epidemic in Uganda at the time of collection in 1997. By contrast, a second genotype cluster, Ug1, only occurred 'at' or 'ahead of' the epidemic 'front', sometimes in mixtures with Ug2. Comparison of mitochondrial cytochrome oxidase I gene sequences for Ug1 and Ug2 and well-studied *B. tabaci* reference populations indicated that the two Ugandan populations exhibited ≈ 8% divergence, suggesting they represent distinct sub-Saharan African lineages. Neither Ugandan genotype cluster was identified as the widely distributed, polyphagous, and highly fecund B biotype of Old World origin, with which they both diverged by ≈ 8%. Within genotype cluster divergence of Ug1 at 0.61 ± 0.1% was twice that of Ug2 at 0.35 ± 0.1%. Mismatch analysis suggested that Ug2 has undergone a recent population expansion and may be of non-Ugandan origin, whereas Ug1 has diverged more slowly, and is likely to be an indigenous genotype cluster.**

*Keywords*: cassava mosaic geminiviruses, CMD, EACMV-Ug, molecular markers, *mtCO1*, whiteflies

*Received 2 November 2001; revision received 20 February 2002; accepted 27 February 2002*

## Introduction

Cassava (*Manihot esculenta* Crantz) is a key food security crop throughout much of sub-Saharan Africa. The most widespread and economically important factor constraining cassava production in Africa is cassava mosaic virus disease (CMD). CMD is caused by cassava mosaic geminiviruses (CMGs) (family: *Geminiviridae*, genus: *Begomovirus*), which are transmitted either by the whitefly vector, *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae), or through vegetative propagation in the planting of infected cuttings (Bock & Woods 1983). In much of Africa, CMD status is stable and most of the incidence is derived from farmers planting infected cuttings when establishing a new crop, although less frequently, where vector numbers are high, the disease spreads more rapidly and epidemics may occur (Thresh *et al*. 1997). Overall losses have been estimated at 12–25% of the total African crop (Thresh *et al*. 1997).

In the late 1980s, a severe outbreak of CMD was reported from northern Uganda (Otim-Nape 1988; Otim-Nape *et al*. 1994). In subsequent years and following more intensive

Correspondence: J.P. Legg, c/o Lambourn and Co. Ltd, Carolyn House, 26 Dingwall Road, Croydon CR9 3EE, UK. Fax: +44 7092 167675; E-mail: jlegg@infocom.co.ug

study it became apparent that an epidemic of CMD was spreading southwards through the central, southern and eastern parts of the country at a rate of 20–30 km per year (Otim-Nape *et al*. 1997; Legg & Ogwal 1998). The CMD associated with the epidemic was unusually severe (Gibson *et al*. 1996), spread more rapidly than previously occurring forms (Legg & Ogwal 1998) and was associated with raised populations of the whitefly vector (Legg & Ogwal 1998; Colvin *et al*. 1999). In addition, studies of CMGs occurring in virus-diseased plants collected throughout Uganda revealed the occurrence of a recombinant CMG in association with the epidemic (Zhou *et al*. 1997). A single virus species, African cassava mosaic virus (ACMV), occurred outside epidemic-affected areas, whereas within epidemic-affected zones ACMV occurred together with a novel strain of East African cassava mosaic virus (EACMV) (Zhou *et al*. 1997). This novel isolate appeared to be a recombinant for which most of the viral coat protein gene was derived from ACMV, whereas the 3′- and 5′-ends of the Cp ORF and the remainder of DNA-A were derived from EACMV (Deng *et al*. 1997; Zhou *et al*. 1997). In view of the > 90% overall homology of the DNA-A of this novel virus with EACMV, it was designated EACMV-Ug (Deng *et al*. 1997). EACMV-Ug was shown to cause more severe symptoms in cassava than ACMV, but the most severe symptoms were produced by a mixture of ACMV and EACMV-Ug, a phenomenon attributed to synergism between the two viruses (Harrison *et al*. 1997).

During the latter part of the decade, the 'pandemic' expanded to neighbouring countries, including Kenya in 1995 (Gibson 1996), southern Sudan in 1997 (Harrison *et al*. 1997) and Tanzania in 1998 (Legg & Okao-Okuja 1999), and it continues to spread westwards in central Africa and southwards through Rwanda and Tanzania (Legg 1999; Legg *et al*. 2001).

Convincing evidence was provided for a link between the emergent, recombinant begomovirus, EACMV-Ug and the expansion of the epidemic (Harrison *et al*. 1997) but little attention was given to the hypothesis that the genetic structure of the whitefly vector populations associated with the epidemic may have shifted, and thereby become a second contributing factor. Indeed, it was noted that vector populations were extraordinarily fecund compared with their status prior to the epidemic (Otim-Nape *et al*. 1997; Colvin *et al*. 1999). In the southwestern United States of America, the increased abundance and unprecedented economic impact of *B. tabaci* which occurred there in the early 1990s were the result of the introduction and establishment of the invasive, and highly fecund B biotype of *B. tabaci* (Brown *et al*. 1995) and the associated viral diseases which subsequently emerged (Brown 1994).

In Uganda, the rate of vector population growth is primarily governed by temperature (Legg 1995), a result congruent with earlier studies carried out in Ivory Coast (Fargette *et al*. 1993a,b). Consequently, differences in vector densities between sites where temperature and weather patterns are identical cannot be attributed to this parameter. Preliminary assessments of the biochemical variability of *B. tabaci* on cassava and other important crop hosts in Ivory Coast using isozyme analyses provided evidence for cassava-associated and noncassava-associated biotypes (Burban *et al*. 1992), and similar results were subsequently obtained from Uganda (Legg *et al*. 1994). Most recently, mating studies between *B. tabaci* populations collected from cassava within and outside the CMD epidemic zone in Uganda indicated that they produced fertile offspring and that there was no reproductive isolation (Maruthi *et al*. 1999).

Two hypotheses have been presented for the superabundance of *B. tabaci* populations associated with the CMD epidemic. The first proposes that a change in host quality elicited by the infection of cassava by the novel epidemic-associated EACMV-Ug boosts the rate of *B. tabaci* reproduction. The second suggests that increased *B. tabaci* populations are associated with the occurrence of a novel more fecund *B. tabaci* biotype in areas affected by the epidemic. Preliminary evidence has been presented in support of the first hypothesis. The observation was made that *B. tabaci* colonizing the CMD-susceptible cassava cultivar 'Ebwanateraka' infected with EACMV-Ug reproduced much more rapidly than *B. tabaci* from the same source but raised on virus-free plants of the same variety (Colvin *et al*. 1999). Changes in the relative composition of amino acids resulting from severe CMD infection were proposed as a possible cause of this apparently synergistic interaction (Colvin *et al*. 1999). An apparent contradiction to this hypothesis, however, is the widely reported superabundance of *B. tabaci* on largely CMD-free plantings of resistant cassava cultivars planted as part of the pandemic management programme in previously affected areas of Uganda (J.M. Thresh, personal communication). No clear evidence has yet been presented either to confirm or refute the second hypothesis that a distinct more fecund biotype may also be associated with the epidemic. The objective of this study was therefore to use molecular techniques to identify whether *B. tabaci* in epidemic-affected areas were distinct from nonepidemic populations. To investigate this possibility, *B. tabaci* samples were collected along transects made across epidemic and nonepidemic zones in Uganda during 1997, and additional samples were collected in 1999. Mitochondrial cytochrome oxidase I gene (*mtCOI*) sequences were then determined for the samples collected. In addition, it was hoped that a molecular comparison of *B. tabaci* from cassava in Uganda with populations collected from other countries and other crops would provide clues as to the evolutionary relationships between them and to recent patterns of population expansion.

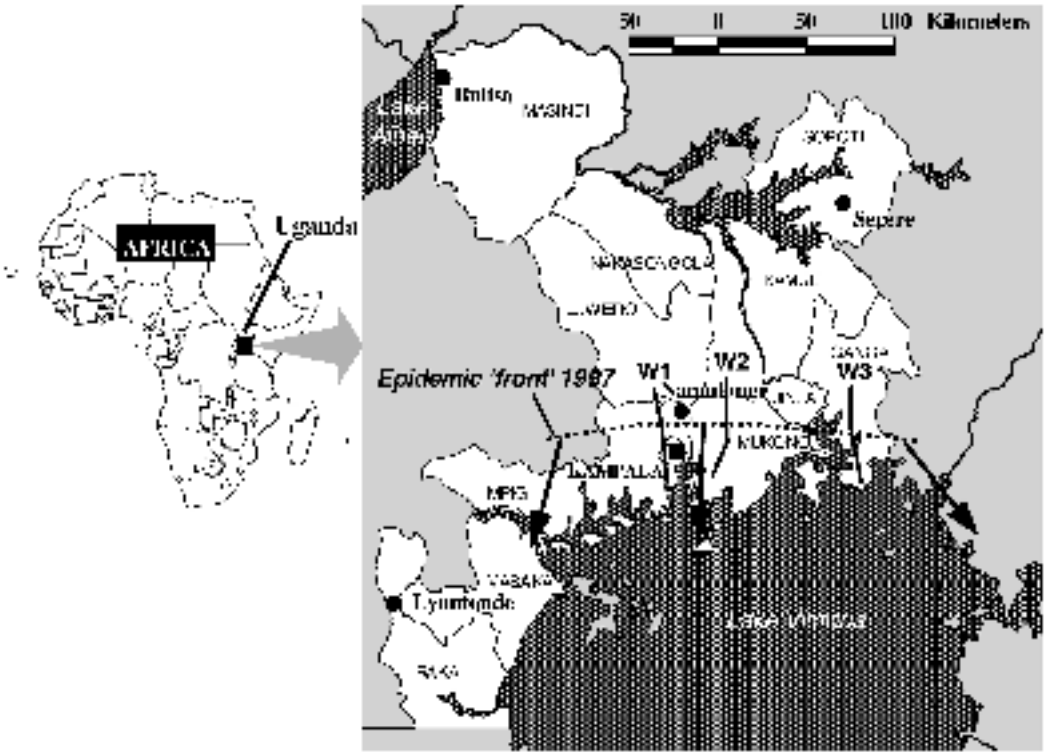


**Fig. 1** Map of central–southern Uganda illustrating sites of cassava *Bemisia tabaci* collection (1997, 1999), the 1997 *B. tabaci* collection transects (W1–W3) and the 1997 position of the north to south moving CMD epidemic front (arrows indicate the direction of epidemic expansion).

## Materials and methods

### *Whitefly collections*

Adult stages of *Bemisia tabaci* were collected from a range of sites across Uganda in 1997 and 1999 and preserved in 80% alcohol prior to molecular characterization.

*1997 Collections*. *Bemisia tabaci* adults were collected along three 50-km long north–south transects in southern Uganda (W1, W2, W3, Fig. 1). The western-most transect (W1) ran from Namagongolo (north) to Kitende (south), the central transect (W2) from Nakanyonyi (north) to Ntenjeru (south) and the most easterly (W3) from Iganga (north) to Bukoba (south). *B. tabaci* collections were made from farmers' cassava fields at the northern limit, southern limit and central point of each transect. Each of the transects ran perpendicular to the epidemic 'front' of severe CMD, and the central point lay on the front. Attributes of the CMD epidemic in Uganda have been thoroughly characterized elsewhere (Otim-Nape *et al*. 1997; Legg & Ogwal 1998) and the following epidemic zone descriptions are based on these characterizations. Epidemic zone descriptions developed through these studies were based on assessments of cassava in farmers' fields along north–south transects similar to the ones used in the study described in this paper. Disease severity is assessed using a 1–5 scoring range in which 1 = symptom free, 2 = mildly diseased, 3 = moderately diseased, 4 = severely diseased and 5 = very severely diseased (Hahn *et al*. 1980). Source of infection can be determined from a visual assessment of CMD diseased plants. Plants infected through the cutting bear symptoms on the lowermost first-formed leaves, in contrast to plants infected by the whitefly vector in which symptoms are absent on these lowermost leaves. Clear evidence has also been adduced to show that the novel CMG recombinant, EACMV-Ug, occurs only in areas affected by the epidemic and is not present elsewhere in Uganda (Harrison *et al*. 1997). Epidemic (behind the front) and nonepidemic (ahead of the front) zones are defined by the characteristics and pattern of spread of CMD, and the epidemic front (at the front) represents the interface between the two. Epidemic zone characteristics were considered as incidence of CMD > 70%, EACMV-Ug and ACMV present, evidence of rapid CMD spread, most disease derived from planting infected cuttings, CMD symptoms severe, major yield losses. Epidemic front zone characteristics were CMD incidence > 30% but < 70%, EACMV-Ug and ACMV present, rapid CMD spread, most disease from current-season whitefly-borne infection, mix of severe (whitefly infected) and mild to moderate (cutting infected) CMD symptoms, and moderate yield losses. 'Nonepidemic zone' characteristics

**Table 1** *Bemisia tabaci* collection transects and localities in southern Uganda, 1997 and sample codes used in Fig. 2

| | Disease zone Behind front (North) | | At front (Central) | | Ahead of front (South) | |
|---|---|---|---|---|---|---|
| Transect | Locality | Samples | Locality | Samples | Locality | Samples |
| Western (Bombo) (W1) | Namagongolo | 5BNama, 6BNama, 7BNama, 48BNama, 49BNama, 50BNama, 51BNama | Matuga | 2BMatu, 3BMatu, 4BMatu, 8BMatu, 9BMatu, 10BMatu, 52BMatu, 54BMatu, 59BMatu, 60BMatu, 62BMatu, 63BMatu, 64BMatu | Kitende | 65BKite |
| Central (Mukono) (W2) | Nakanyonyi | 24MNaka, 25MNaka, 26MNaka | Kabembe | 22MKabe, 23MKabe | Ntenjeru | 27MNten, 28MNten, 29MNten, 30MNten, 31MNten, 44MNten, 45MNten, 46MNten, 47MNten |
| Eastern (Iganga) (W3) | Iganga | 13IIgan, 14IIgan, 15IIgan | Ikulwe | 16IIkul, 17IIkul, 18IIkul, 19IIkul, 20IIkul, 69IIkul, 72IIkul, 73IIkul, 74IIkul, 75IIkul | Bukoba | 11IBuko, 12bIBuk, 12IBuko, 53IBuko, 61IBuko |

were < 30% CMD incidence, ACMV present, EACMV-Ug absent, slow CMD spread, virtually all disease inoculum from planting infected cuttings, CMD symptoms mild to moderate and minor yield losses. In this study, these three zones with contrasting patterns of CMD epidemiology are referred to as 'disease zones'.

At each of the behind the front (northernmost), at the front (central) and ahead of the front (southern) 'localities' along each transect (Table 1), *B. tabaci* adults were collected from four cassava fields which all lay within a radius of 5 km. The insects were collected from the undersides of the uppermost leaves of 3–8-month-old cassava crops at each site. Whiteflies collected from the four fields at a given locality were pooled, and this pooled set was considered as the collection for that locality. Additional collections (with sample codes) were made from Namulonge (32Namu, 56Namu, 66Namu) and Serere (34Sere, 55Sere) research stations and Bulisa (1Buli, 58Buli), on the shores of Lake Albert (Fig. 1), all of which were considered to lie within the epidemic zone. *B tabaci* adults were also collected from sweetpotato (*Ipomoea batatas* L) at Namulonge (33NamSP, 57NamSP, 67NamSP).

*1999 Collections*. Collections of *B. tabaci* adults were made from cassava at two locations. These were Lyantonde (43Lyan), Rakai district in western Uganda (Fig. 1) and Namulonge (36–42Namu, 70/71Namu). At this stage both of these sites were considered to be well behind the epidemic front, which was moving southwards into Tanzania (Legg 1999).

Identification and population structure of *B. tabaci* by mtCOI sequence analyses. Mitochondrial COI sequences (~615 bp) were determined from whiteflies collected in Uganda, and additional whitefly COI reference sequences were as described (Frohlich *et al*. 1999; Kirk *et al*. 2000; Brown 2001a). The mtCOI sequence was obtained by polymerase chain reaction (PCR) for three to four adult whiteflies from each field collection. The numbers of whiteflies sequenced for different transect localities varied (Table 1), but overall, the number of sequences generated for the different disease zones and transects was similar, and in all cases, three to four whitefly adults were used to produce each sequence. Amplicons were cloned and the DNA sequence was obtained in both forward and reverse directions for a minimum of three clones for each amplicon. Each sequence is the consensus of at least three cloned amplicons for each PCR product. Single adult whiteflies were prepared for PCR using the lysis procedure, as described by Frohlich *et al*. (1999). PCR amplification of the target cytochrome oxidase I gene (*mtCOI*), visualization by ethidium bromide staining and molecular cloning were as described by Frohlich *et al*. (1999). Primers used to amplify the mitochondrial COI gene fragment were C1-J-2195 (5′-TTGATTTTTTGGTCATCCAGAAGT-3′) and L2-N-3014 (5′-TCCAATGCACTAATCTGCCA-TATTA-3′) and were obtained from the UBC Insect Mitochondrial DNA Primer Oligonucleotide Set, compiled by B. J. Crespi and C. Simon (Simon *et al*. 1994). PCR products were cloned using a TA-cloning kit by insertion of PCR products into the plasmid pCR2.1 vector (Invitrogen, Carlsbad, USA) per manufacturer's instructions. Cloned mtCOI amplicons were sequenced at the Laboratory for Systematics and Molecular Evolution (University of Arizona, Tucson, USA).

Reference whitefly mtCOI sequences were obtained from previous studies carried out in the Arizona laboratory

and are available as accessions in the GenBank database. Reference populations and their Accession nos are as follows, Argentina: 2 Arg [AF340212, ARG1) and 3 Arg [AF340213, ARG2]; Arizona: 4 US AZ A, Arizona A biotype, AZA [AYO57122] and 69 US AZ B, Arizona B biotype AZB [AYO57123]; Benin: 1 Benin [AF110693, ABA]; Bolivia: 5 Boliv [AF342768]; California: 6 US CA A, California A biotype [AYO57124, CAL A]; Cameroon: 76 Cam [AF344247, Cam (3), 77 Cam [AF344258, Cam 9], 78 Cam [AF344256, Cam 15], and 79 Cam [AF344257, Cam 16]; China: 15 China [AF342777, HC, cotton]; El Salvador: 8 El Sal, weed [AYO57127]; Guatemala: 9 Guat, Esquintla cotton [AF342771], 10 Guat, tomato [AYO57129], 11 Guat, watermelon [AYO7130], 12 Guat, cucurbit [AYO57131], and 13 Guat, Chile [AYO57132]; Honduras: 16 Hond. melon 1 [AF342770] and 17 Hond. melon 2 [AYO57133]; Malawi: 21 Malaw, cassava [[AYO57215]; Mexico: 7 Mex Cul [AYO57125]; Morocco: 20 Moroc, tomato 1 [AF342773] and 21 Moroc, tomato 2 [[AYO57138]; Mozambique: 82 Moz [AF344278]; India (18 India) [AF110702], Nepal: 22 Nepal [AF342779, NEW], Pakistan: 23 Pakis [AF342778, PC91], 24 Pakis [AY057582, PC92] and 25 Pakis [AY057583 PC95]; Puerto Rico: 19 PR Jat [AF110705, Jatropha biotype] and (27 PR Sida) [AYO57134, Sida biotype]; South Africa-Lucia region: 81 SA Lucia [AF344259, Lucia 1]; Spain: 28 Spain [AF342775, SP92], 29 Spain [AYO57139, Sp95], and 30 Spain [AF342769, SP99]; Sudan: 26 Sudan [AF110706, SC]; Swaziland: 80 Swa [AF344269, Swa 1]; Turkey: 31 Turk [AF342776, TC91]; Zambia: 85 Zam [AF344280, Zam 1], 86 Zam [AF344281, Zam 2], and 87 Zam [AF344282, Zam 3]; Zimbabwe: 83 Zim [AF344285, Zim 1] and 84 Zim [AF344286, Zim 2]; and outgroups: 97 B afer (*Bemisia afer*) [AYO57218] and 98 B ber (*Bemisia berbericola*) [AYO57219]. Master sequences were edited and manually aligned using the SEQUENCE ALIGNMENT EDITOR (Se-Al) (Version 1.0 alpha 1, Copyright © 1996; A. Rambaut, Department of Zoology University of Oxford, UK). Neighbour-joining trees were made using CLUSTAL X (Thompson *et al*. 1997).

Hierarchical analysis of molecular variance (AMOVA; Excoffier *et al*. 1992) was performed using ARLEQUIN, Version 2.0 (Schneider *et al*. 1999). AMOVA provides an analysis of variance in gene frequencies, while also taking into account the number of mutations between molecular haplotypes. Standard variance components were generated at each of three levels of subdivision shown in Tables 1, i.e. among geographical regions (either Disease Zone or Transect), among populations (i.e. sample localities) within regions, and among individuals. The significance of the variance components and associated *F*-statistics (Excoffier *et al*. 1992) for different levels of genetic structure were tested using 1023 random permutations for total data.

Mismatch analysis was used to estimate past population size changes in the history of the two Uganda whitefly lineages. Estimates of Tau, time of population expansion, expressed in units of mutational time and the population size parameters before (Theta0) and following (Theta1) expansion were obtained by minimizing the sum of square deviation (SSD) of the observed data to a step-wise model of growth (Schneider & Excoffier 1999). Confidence intervals for those parameters were obtained by comparing the observed SSD value with SSDs resulting from Monte Carlo simulations of 1000 random samples using the coalescent algorithm of Hudson (1990) as implemented in ARLEQUIN, Version 2.0. The distribution of pairwise differences between two sequences drawn at random, given the derived values of Tau, Theta0, and Theta1, was calculated and plotted along with the observed values. Other population test statistics calculated were the raggedness index (Harpending 1994), Tajima's D (Tajima 1989) and *F* values (Fu 1997). Statistical significance estimates for these were obtained by coalescent simulations.

## Results

### *Phylogenetic and population analyses for whitefly mtCOI sequences*

A consensus mtCOI sequence was obtained for individual adult whitefly samples for the field locations noted, and sequences were deposited in GenBank under Accession nos AYO57141-AYO57216.

Analysis of whitefly mtCOI sequences revealed that two distinct mtDNA haplotypes, Uganda 1 (Ug1) and Uganda 2 (Ug2), were present on cassava in Uganda when transect collections were made in 1997. Both are sister taxa in a large, monophyletic clade comprising haplotypes unique to the African continent (Fig. 2). Closest (available) relatives to the Ug1 haplotype were identified from southern Africa in Zambia [Zam (85), Zam (86) and Zam (87)] and the Lucia region of South Africa [St Lucia (81)]. Divergence of nucleotide sequences between Ug1 and these southern African haplotypes ranged from 0.3 to 2.4%. In contrast, Ug2 sequences showed greatest similarity to those of *Bemisia tabaci* from Cameroon [Cam (76)], in central-west Africa, differing by only 1.0%. Ug1 and Ug2 differ by $\approx 7.8\%$ per nucleotide in the COI region sequenced, whereas per nucleotide diversity within haplotype Ug1 was $0.61 \pm 0.1\%$ and diversity within haplotype Ug2 was $0.35 \pm 0.1\%$. Between-haplotype divergence was therefore > 10-fold higher than genetic variation found within either haplotype lineage. *B tabaci* collected from sweetpotato at Namulonge, Uganda, had mtCO1 sequences > 16% divergent from those of sympatric cassava *B. tabaci* collected at the same time and from the same location.

The distribution of all pairwise comparisons within Ug1 and Ug2 is shown in Fig. 3. Both haplotype populations have unimodal distributions. Mismatch distributions such as these can provide clues about the demographic history

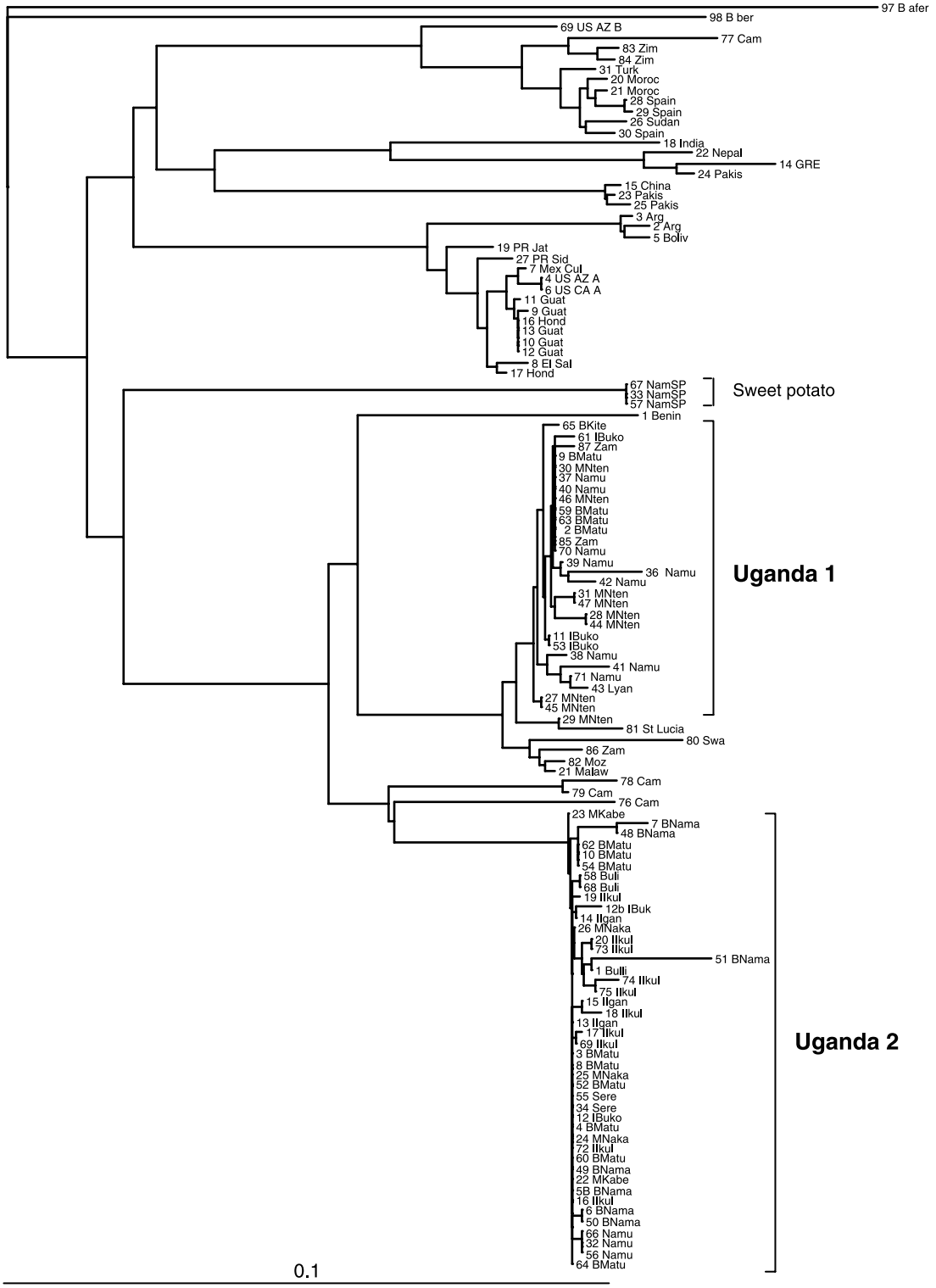


**Fig. 2** Neighbour-joining phylogeny of mtDNA COI sequences examined here together with sequences for other representative *Bemisia tabaci* populations, worldwide.

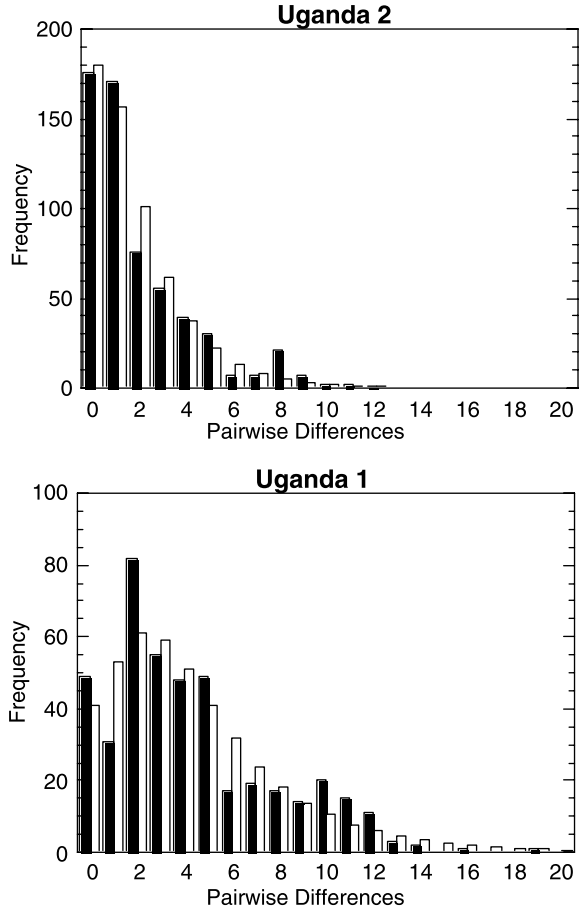


**Fig. 3** Mismatch distribution of Uganda2 and Uganda1 haplotypes. Dark bars are observed frequencies of pairwise differences and white bars are those expected for a model of uniform population expansion, given values of Tau, Theta1, and Theta2 in Table 2.

of a population. Neutrally evolving, stable populations tend to have multimodal, ragged mismatch distributions, whereas populations undergoing sudden population expansions or exponential growth are expected to produce smooth and unimodal mismatch distributions (Slatkin & Hudson 1991; Rogers & Harpending 1992; Harpending 1994; Rogers 1995). Historical population expansion model data for haplotypes Ug1 and Ug2 were consistent with population growth in both cases (Table 2). The observed mismatch frequencies are close to those expected (Fig. 3) indicating a good fit to the growth model ($P > 0.58$ and $P > 0.78$, respectively). Harpending's raggedness index (Harpending 1994) for the Ug1 haplotype data was 0.031 ($P > 54\%$) and for the Ug2 data the raggedness index was 0.030 ($P > 94\%$), also suggesting a very good fit of the empirical data to the growth model. The estimated time of expansion of Ug2 is near 0, suggesting a much more recent expansion for it than for Ug1 (Table 2). Finally, a hypothesis of a stationary population model for either Ug1 or Ug2 was not supported by Tajima's (Tajima 1989) *D*-values (–1.919; $P < 0.02$, and –2.011; $P < 0.01$, respectively) or by Fu's (Fu 1997) $F_s$-statistic (–4.540; $P < 0.02$, and –6.235; $P < 0.001$, respectively).

The geographical distribution of the two haplotypes differed among the three disease zones with Ug1 (local vector) occurring primarily 'ahead of the front' and Ug2 (invader) being present 'at' or 'behind the front'. Frequencies of occurrence of the two types in the different disease zones were: (ahead of front/at front/behind front) Ug1 (13/4/0), Ug2 (2/21/13). AMOVA analysis of the molecular variation among whiteflies sampled along the study transects revealed that over half (56%) of the variability was due to disease zone with a statistical confidence level > 99%. There was little variation (6%) attributed to localities in the same zone. A similar AMOVA using Bombo, Mukono and Iganga transects as the highest level groupings showed that none of the molecular diversity correlated with these Western, Central or Eastern transect locations. Additional samples collected from the more widely distributed epidemic-affected locations of Bulisa, Serere and Namulonge were all grouped within the Ug2 haplotype.

Ten additional *B. tabaci* samples were collected two years later at two localities (Lyantonde and Namulonge), which were both at that time behind the front (Legg 1999). In contrast to the 1997 collections from this disease zone, which were exclusively of the Ug2 haplotype, all 10 1999 samples were determined to have the Ug1 COI haplotype.

**Table 2** Population growth parameters derived from analysis of the mismatch distributions shown in Fig. 3

| Popn | Time of popn growth, Tau* | 95% CI [lower, upper bounds] | Theta0† | 95% CI [lower, upper bounds] | Theta1† | 95% CI [lower, upper bounds] | *P* value‡ |
|---|---|---|---|---|---|---|---|
| Uganda2 | 0.27 | [0.00, 3.52] | 1.53 | [0.00, 5.66] | 92.38 | [3.77, 5407] | 0.78 |
| Uganda1 | 2.03 | [0.31, 11.7] | 2.51 | [0.00, 8.30] | 10.97 | [4.63, 5000] | 0.58 |

*Units of t generations multiplied by the mutation rate per gene per generation.
†Twice the effective population size times the mutation rate per generation.
‡Probability that the observed data fit the growth model as well or better than 1000 simulated samples based on the population growth parameters (Schneider & Excoffier 1999).

**Table 3** Analysis of molecular variance to test for evidence of population structure, as implemented in arlequin (Schneider *et al*. 1999). The population structure tested here grouped sequences by disease zone (behind the front, at the front and ahead of the front) at the highest level, then by localities. The percentage column indicates the amount of total variance explained by each of the hierarchical levels. F-Statistics identify the correlation of alleles at each of the hierarchical levels. P-Values were obtained by comparing observed values to those calculated from 1000 random permutations of the data matrix and represent the probabilities of obtaining these parameter values if there were no population structure

| Source of variation | d.f. | Variance components | % of variation | *F*-Statistic | *P*-Value |
|---|---|---|---|---|---|
| Among groups | 2 | 8.10 | 55.7 | 0.6178 | < 0.0001 |
| Among populations within groups | 6 | 0.89 | 6.1 | 0.1375 | < 0.025 |
| Within populations | 44 | 5.55 | 38.2 | 0.5569 | < 0.0001 |
| Total | 52 | 14.54 | | | |

## Discussion

### *Aetiology and dynamics of the CMD pandemic*

The CMD pandemic has had a major economic impact on cassava production in East Africa and it is therefore vital to understand the factors associated with its intensity and pattern of spread. Clear evidence has been presented for the association of the novel virus recombinant, EACMV-Ug, with the epidemic (Harrison *et al*. 1997) but the role of the whitefly vector, *Bemisia tabaci*, has hitherto been poorly understood.

Previous studies have characterized an epidemic front for CMD in Uganda (Otim-Nape *et al*. 1997; Legg & Ogwal 1998), which was described as the interface between the epidemic affected zone to the north and the as yet unaffected zone to the south. These studies provided the first quantitative information suggesting that *B. tabaci* was more abundant at and behind the epidemic front than ahead of it (Legg & Ogwal 1998). Two hypotheses have been advanced in an attempt to explain these changes. The first suggests that synergistic interactions between virus-diseased cassava plants and *B. tabaci* give rise to increases in fecundity and promote plant-to-plant and crop-to-crop movement, thereby enhancing spread (Colvin *et al*. 1999). The second envisages the expansion in geographical range of a novel, more highly fecund *B. tabaci* genotype that transmits EACMV-Ug with equal or greater efficiency than the local genotypes, which is discussed here. Although data have been presented in support of the first hypothesis (Colvin *et al*. 1999), here we present additional and possibly complementary evidence for the association of a distinct *B. tabaci* genotype cluster with the CMD epidemic.

### *Association of* B. tabaci *genotypes with the CMD pandemic*

Two *B. tabaci* genotype clusters were recorded colonizing cassava in Uganda, referred to as Ug1 and Ug2. The putative invader genotype cluster, Ug2, was detected almost exclusively at localities behind the epidemic front and at the front, although a small proportion was also found in collections ahead of the front together with Ug1. Ug2 was the sole haplotype at the more distant behind the front locations of Bulisa, Serere and Namulonge (Fig. 1). By contrast, the putative local genotype cluster, Ug1, was most frequently detected ahead of the front, and to some extent at the front, although it was never identified in collections from behind the front. Collectively, these data indicate a strong association between Ug2 and the severe virus epidemic. This is the first demonstration of an association between a *B. tabaci* genotype cluster and the recent epidemic in Uganda.

### *Phylogenetic relationships within and between Ugandan and other* B. tabaci *populations*

Ug1 and Ug2 were ≈ 8% divergent, suggesting that they are distinct genotype clusters of *B. tabaci* (Frohlich *et al*. 1999; Brown 2001a). Within genotype cluster divergence for Ug1 was $0.61 \pm 0.1$ and $0.35 \pm 0.1$ for Ug2 revealing twice the level of within-population diversity for Ug1. These two genotype clusters differ from one another to about the same extent as they differ individually from the B biotype, further suggesting Ug1 and Ug2 are unique lineages. Both are clearly African genotype clusters, neither of which share a close genetic affinity to the B biotype as evidenced by an association with a cluster separate from that containing the well-characterized AZ B, a B type from Turkey, along with other B biotypes from different locations worldwide. Among the available reference sequences, the closest relatives of Ug1 were found to have an origin in Swaziland (Swa 80), Zambia (Zam 85–87), and the Lucia region of South Africa (St Lucia 81), all populations from southern Africa, and from which Ug1 sequences diverged by ≈ 2%. In contrast, the closest relative identified for Ug2 is from Cameroon (Cam 76–78) with which it is 1% divergent. *Bemisia tabaci* collected from

sweetpotato in Uganda were clearly separated from all African cassava populations, a result which is consistent with biological evidence for the host restriction of sympatric cassava and noncassava *B. tabaci* populations (Burban *et al*. 1992; Legg 1996).

### *Evolutionary perspectives*

Results collectively indicate that Ug2 was possibly introduced or upsurged in Uganda rather recently, whereas Ug1 appears to constitute a local or indigenous genotype cluster that has resided in the area for some time, owing to its distribution primarily 'ahead of the front'. It is also possible that the Ug2 genotype cluster occurred locally, was reproductively isolated and upsurged only recently, owing to unknown factors. The latter possibility seems less likely than an invasion hypothesis, because results indicate that certain collections 'at the front' contained mixtures of the two genotype clusters, hence, they are sympatric on cassava and gene flow would be expected if they are capable of interbreeding in the field, a possibility which seems to be supported by cage-based mating studies (Maruthi *et al*. 1999). If hybrids were produced that were more fit than one or both parental genotypes, they should constitute the majority, if not all, of the current population. An additional observation that supports this view, is the occurrence of the Ug1 haplotype alone in the smaller number of 1999 collections made from locations affected by the epidemic in the mid-1990s, coupled with the absence of Ug2. It seems likely that interbreeding between Ug1 and Ug2 populations may over time have resulted in the disappearance of the Ug2 haplotype and the emergence of Ug1/Ug2 hybrids with the Ug1 haplotype. To evaluate these hypotheses, extant populations must now be examined using a nuclear marker. Within-genotype cluster genetic diversity estimates are also indicative of a larger gene pool for the indigenous *B. tabaci* compared with the putative invasive genotype cluster, and are thereby consistent with Ug1 being widely distributed and a predominant genotype cluster in cassava in Uganda prior to the appearance of Ug2 in the early 1990s.

Results of expansion model analysis indicate population growth dynamics for both Ug1 and Ug2. The timing of growth events differs between the two, however. The 'invader' Ug2 experienced more recent population growth, which is consistent with its appearance coincident with the epidemic. In contrast, the mismatch analysis of Ug1 mtDNAs implies that its ancestors underwent a period of growth but much earlier than Ug2 such that Ug1 has had sufficient evolutionary time to diverge by almost 1% following its population expansion. Thus, Ug1 might have invaded the area in the distant past but has been well established ever since. The lack of evidence for contemporary growth of Ug1 is consistent with a nonupsurgent phenotype.

## Bemisia tabaci *and geminivirus epidemics in Uganda and elsewhere*

In the field, the greater abundance of whiteflies in epidemic-affected zones suggests the possibility that the epidemic-associated Ug2 may be a more highly fecund genotype cluster and could feasibly have been a major driving force behind the epidemic. However, the association between Ug2 and increased fecundity remains speculative and additional experiments are needed to confirm or refute this. Moreover, data presented here are not in contradiction with the hypothesis proposing synergism between *B. tabaci* and CMD-diseased cassava as the factor driving population increase. It is also possible that synergistic interactions between Ug2 and EACMV-Ug, leading to increased population growth rates, have given rise to the rapid expansion of Ug2 and its dominance in epidemic affected areas.

Another unanswered question concerns the potential for *B. tabaci* populations to upsurge suddenly and either displace existing populations or decline and disappear altogether. For many years, this population dynamic has been alluded to yet rarely has the phenotype been associated with a particular genotype. Indeed, the B biotype of *B. tabaci* has similar characteristics to the Uganda populations, and in addition, is highly polyphagous, whereas, the cassava-colonizing *B. tabaci* from Africa are thought to be host restricted (Burban *et al*. 1992; Legg 1996). Interestingly, the B biotype also fits the expansion model for recent and rapid population growth (R. French & J. Brown, unpublished), and as its recent upsurgence and invasiveness worldwide has been documented (Frohlich *et al*. 1999), it may serve as an excellent baseline for a comparison of the two Uganda genotype clusters. Invasions in the New World by the B biotype were immediately followed by the emergence of new begomoviruses that, to date, appear to be local in origin, with the exception of one introduction that was recently documented (Brown 1994, 2001b).

Studies are in progress to track the CMD pandemic with respect to patterns of whitefly vector and viral genotypes as the disease spreads into neighbouring countries in Africa. Additional studies are expected to reveal important characteristics about 'invasive' and 'noninvasive' *B. tabaci* phenotypes and the potential and basis for their impact on the emergence of new begomoviruses in the short- and long-term. The improved understanding of host/virus/vector interactions and dynamics of the CMD epidemic that these studies will provide will be a vital component in the overall effort to control a disease phenomenon that continues to be one of the most important current threats to the food security of millions of people in East and Central Africa.

## Acknowledgements

Funding for this work was provided by the Office for Foreign Disaster Assistance (OFDA) of the United States Agency for International Development (USAID). OFDA funds were used for the implementation of an Emergency CMD Management Project, of which this study was a part. This project was set up within the framework of the Whitefly IPM Project of the Systemwide Programme for Integrated Pest Management (SP-IPM), an intercentre programme established by the Consultative Group for International Agricultural Research (CGIAR). Professor Christine Rey, University of Witwatersrand, South Africa, is acknowledged for contributing reference *B. tabaci* collections from southern Africa. The National Agricultural Research Organization of Uganda, host to IITA-ESARC, is recognized for its on-going support for the work of IITA and its partners in Uganda and the wider East African region.

## References

Bock KR, Woods RD (1983) Etiology of African cassava mosaic disease. *Plant Disease*, **67**, 944–955.

Brown JK (1994) The status of *Bemisia tabaci* (Genn.) as a pest and vector in world agroecosystems. *FAO Plant Protection Bulletin*, **42**, 3–32.

Brown JK (2001a) Molecular markers for the identification and global tracking of whitefly vector–begomovirus complexes. *Virus Research*, **71**, 233–260.

Brown JK (2001b) The molecular epidemiology of begomoviruses. In: *Trends in Plant Virology* (eds Khan JA, Dykstra J), pp. 279–316. Haworth Press, New York.

Brown JK, Frohlich DR, Rosell RC (1995) The sweetpotato/silverleaf whiteflies: biotypes of *Bemisia tabaci* (Genn.), or a species complex? *Annual Review of Entomology*, **40**, 511–534.

Burban C, Fishpool LDC, Fauquet C, Fargette D, Thouvenel J-C (1992) Host-associated biotypes within West African populations of the whitefly *Bemisia tabaci* (Genn.), Homoptera: Aleyrodidae. *Journal of Applied Entomology*, **113**, 416–423.

Colvin J, Otim-Nape GW, Holt J *et al.* (1999) Factors driving the current epidemic of severe cassava mosaic disease in East Africa. Proceedings of the VIIth International Plant Virus Epidemiology Symposium, Aguadulce (Almeria), Spain, April 11–16, 1999. Abstract 76–77.

Deng D, Otim-Nape GW, Sangare A *et al.* (1997) Presence of a new virus closely associated with cassava mosaic outbreak in Uganda. *African Journal of Root and Tuber Crops*, **2**, 23–28.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fargette D, Jeger M, Fauquet C, Fishpool LDC (1993a) Analysis of temporal disease progress of African cassava mosaic virus. *Phytopathology*, **84**, 91–98.

Fargette D, Muniyappa V, Fauquet C, N'Guessan PN, Thouvenel J-C (1993b) Comparative epidemiology of three tropical whitefly-transmitted geminiviruses. *Biochimie*, **75**, 547–554.

Frohlich DR, Torres-Jerez I, Bedford ID, Markham PG, Brown JK (1999) A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Molecular Ecology*, **8**, 1593–1602.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Gibson RW (1996) The report of a survey monitoring the spread of the epidemic of African cassava mosaic virus from Uganda into western Kenya. Internal report, 17 pp. Natural Resources Institute, Chatham, UK.

Gibson RW, Legg JP, Otim-Nape GW (1996) Unusually severe symptoms are a characteristic of the current epidemic of mosaic virus disease of cassava in Uganda. *Annals of Applied Biology*, **128**, 479–490.

Hahn SK, Terry ER, Leuschner K (1980) Breeding cassava for resistance to cassava mosaic disease. *Euphytica*, **29**, 673–683.

Harpending H (1994) Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.

Harrison BD, Zhou X, Otim-Nape GW, Liu Y, Robinson DJ (1997) Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology*, **131**, 437–448.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology*, Vol. 7 (eds Futuyma D, Antonovics J), pp. 1–44. Oxford University Press, London.

Kirk AA, Lacey LA, Brown JK *et al.* (2000) Variation within the *Bemisia tabaci* s.1. species complex (Hemiptera: Aleyrodidae) and its natural enemies leading to successful biological control of *Bemisia* biotype B in the USA. *Bulletin of Entomological Research*, **90**, 317–327.

Legg JP (1995) The ecology of *Bemisia tabaci* (Gennadius) (Homoptera), vector of African cassava mosaic geminivirus in Uganda. PhD Thesis, University of Reading, UK.

Legg JP (1996) Host-associated strains within Ugandan populations of the whitefly *Bemisia tabaci* (Genn.) (Hom., Aleyrodidae). *Journal of Applied Entomology*, **120**, 523–527.

Legg JP (1999) Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Protection*, **18**, 627–637.

Legg JP, Gibson RW, Otim-Nape GW (1994) Genetic polymorphism amongst Ugandan populations of *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae), vector of African cassava mosaic geminivirus. *Tropical Science*, **34**, 73–81.

Legg JP, Ogwal S (1998) Changes in the incidence of African cassava mosaic geminivirus and the abundance of its whitefly vector along south–north transects in Uganda. *Journal of Applied Entomology*, **122**, 169–178.

Legg JP, Okao-Okuja G (1999) Progress in the diagnosis and epidemiological characterisation of cassava mosaic geminiviruses in East Africa. Proceedings of the VIIth International Plant Virus Epidemiology Symposium, April 11–16, 1999, Aguadulce (Almeria), Spain. Abstract 74–75.

Legg JP, Okao-Okuja G, Mayala R, Muhinyuza J-B (2001) Spread into Rwanda of the severe cassava mosaic virus disease pandemic and associated Uganda variant of East African cassava mosaic virus (EACMV-Ug). New Disease Reports. http://www.bspp.org.uk/ndr/.

Maruthi MN, Colvin J, Seal S (1999) The geographic distribution of cassava *Bemisia tabaci* biotypes in relation to the current epidemic of cassava mosaic disease in East Africa. Proceedings of the VIIth International Plant Virus Epidemiology Symposium, Aguadulce (Almeria), Spain, April 11–16, 1999. Abstract 88.

Otim-Nape GW (1988) *Cassava Situation in Luwero District. A Mission Report by the Uganda Root Crops Programme*. Namulonge Research Station, Kampala.

Otim-Nape GW, Bua A, Baguma Y *et al.* (1994) Accelerating the transfer of improved production technologies: controlling

cassava mosaic virus disease epidemics in Uganda. *African Crop Science Journal*, **2**, 479–495.

Otim-Nape GW, Bua A, Thresh JM *et al.* (1997) *Cassava Mosaic Virus Disease in Uganda: the Current Pandemic and Approaches to Control*. Natural Resources Institute, Chatham, UK.

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schneider S, Roessli D, Excoffier L (1999) *ARLEQUIN, Version 2.0: A Software for Genetic Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Thresh JM, Otim-Nape GW, Legg JP, Fargette D (1997) African cassava mosaic disease: the magnitude of the problem? Proceedings of the Cassava Biotechnology Network, Third International Scientific Meeting Kampala, Uganda, 26–31 August 1996. *African Journal of Root and Tuber Crops Special Issue: Contributions of Biotechnology to Cassava for Africa* (eds Thro AM, Akoroda MO), pp. 13–19.

Zhou X, Liu Y, Calvert L *et al.* (1997) Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology*, **78**, 2101–2111.

---

James Legg is a Senior Scientist working with IITA and NRI, UK. From a base in Kampala, Uganda, he pursues research interests in the characterization and management of virus/virus vector constraints to root crop production in East and Central Africa. Roy French is a Research Plant Pathologist for the USDA Agricultural Research Service. His research interests include virus evolution and gene function. Judith Brown is an Associate Professor in the Department of Plant Sciences at the University of Arizona, Tucson. Her research interests include the ecology and diversification of whitefly transmitted geminiviruses and whitefly vector biotypes, toward understanding virus–vector interactions and their coevolution.

---