Kraberger, S. et al. (2013). Evidence that dicot-infecting mastreviruses are particularly prone to inter-species recombination and have likely been circulating in Australia for longer than in Africa and the Middle East. Virology, 444(1-2): 282-291.
http://dx.doi.org/10.1016/j.virol.2013.06.024

Evidence that dicot-infecting mastreviruses are particularly prone to inter-species recombination and have likely been circulating in Australia for longer than in Africa and the Middle East

Simona Kraberger, Gordon W. Harkins, Safaa G. Kumari, John E. Thomas, Mark W. Schwinghamer, Murray Sharman, David A. Collings, Rob W. Briddon, Darren P. Martin and Arvind Varsani

Abstract

Viruses of the genus *Mastrevirus* (family *Geminiviridae*) are transmitted by leafhoppers and infect either mono- or dicotyledonous plants. Here we have determined the full length sequences of 49 dicot- infecting mastrevirus isolates sampled in Australia, Eritrea, India, Iran, Pakistan, Syria, Turkey and Yemen. Comprehensive analysis of all available dicot-infecting mastrevirus sequences showed the diversity of these viruses in Australia to be greater than in the rest of their known range, consistent with earlier studies, and that, in contrast with the situation in monocot-infecting mastreviruses, detected inter-species recombination events outnumbered intra-species recombination events. Consistent with Australia having the greatest diversity of known dicot-infecting mastreviruses phylogeographic analyses indicating the most plausible scheme for the spread of these viruses to their present locations, suggest that most recent common ancestor of these viruses is likely nearer Australia than it is to the other regions investigated.

Introduction

Throughout the agricultural regions of Australia, south and north-east Africa, the Middle East and India, mastreviruses are recognised as potentially important threats to chickpea (*Cicer arietinum*), lentil (*Lens culinaris*), bean (*Phaseolus vulgaris*) and tobacco (*Nicotiana tabacum*) production (Farzadfar et al., 2002; Hadfield et al., 2012; Halley-Stott et al., 2007; Horn et al., 1994, 1993; Kumari et al., 2004, 2008; Makkouk et al., 2003; Mumtaz et al., 2011; Nahid et al., 2008; Schwinghamer et al., 2010; Thomas et al., 2010). Besides being economically important export crops for countries such as Australia, pulses such as lentils, chickpeas and beans are key dietary staples in northern Africa, India, Pakistan and the Middle East with India alone producing around five million tonnes per annum over four decades upto 2005 (Knights et al., 2007). By influencing the yields of important food crops in these populous and often agriculturally marginal regions,

pathogens including mastreviruses threaten the food security of a substantial number of the world's most economically vulnerable people.

Mastreviruses (Family *Geminiviridae*) are single stranded DNA (ssDNA) viruses with "" 2.5–2.7 kb circular genomes that are encapsidated in twinned isometric viral particles (Harrison, 1985). Members of the mastrevirus genus include species which infect either dicotyledonous (dicot) or monocotyledonous (monocot) hosts that are transmitted by leafhoppers, and have genomes expressing four proteins – two encoded on the virion strand and two on the complementary strand. The movement protein and coat protein which are encoded on the virion strand are required for systemic spread and encapsidation, whereas two replication associated proteins, Rep and RepA are translated from alternatively spliced complementary sense transcripts and are required for replication (Dekker et al., 1991; Mullineaux et al., 1990; Schalk et al., 1989; Wright et al., 1997). These mastreviruses and related ssDNA viruses that replicate by a rolling circle mechanism (Jeske et al., 2001; Martin et al., 2011a) all have virion strand origins of replication containing a highly conserved nonanucleotide sequence (TAAT[A/G]TTAC in gemini-viruses) bounded by an inverted repeat sequence that is capable of forming a hairpin (Heyraud et al., 1993).

According to the most recent report by the ICTV Geminiviridae Study Group on mastrevirus classification there are six known species of dicot-infecting mastrevirues (Muhire et al., 2013). One species, *Chickpea chlorotic dwarf virus* (CpCDV), has been found only in the Middle East (including Turkey), Africa and India (Ali et al., 2004; Horn et al., 1993; Mumtaz et al., 2011; Nahid et al., 2008). All five of the other recognised species have only ever been found in Australia. These include *Chickpea redleaf virus* (CpRLV) (Thomas et al., 2010), *Chickpea yellows virus* (CpYV) (Hadfield et al., 2012), *Chickpea chlorosis virus* (CpCV) (Hadfield et al., 2012; Thomas et al., 2010), *Chickpea chlorosis Australia virus* (CpAV) (Hadfield et al., 2012) and *Tobacco yellow dwarf virus* (TYDV) (Hadfield et al., 2012; Morris et al., 1992). This known distribution of dicot-infecting mastrevirus species is less extensive than that of the monocot-infecting mastreviruses which have been identified in Africa, Europe, Asia, Indian Ocean islands, throughout the Pacific rim and, more recently, in the Caribbean (Muhire et al., 2013; Rosario et al., 2013).

Consistent with the notion that Australia is both the present centre of dicot-infecting mastrevirus diversity and is close to the region where these viruses first emerged, are that CpCDV is the only dicot-infecting mastrevirus species to be discovered outside of Australia, and phylogenetic evidence indicating that CpCDV forms a distinct monophyletic clade with high statistical support that is nested within a much larger clade that contains the five Australian species (Hadfield et al., 2012) It is, however, also possible that this view of dicot-infecting mastrevirus diversity has been biased by the fact that Australia is the site where these viruses have been most intensively sampled. It is entirely plausible that, as more dicot-infecting mastreviruses are sampled from elsewhere in the world, a completely different picture will emerge.

In order to get a better perspective of the extent of dicot-infecting mastrevirus diversity in other parts of the world we determined the full genome sequences of 30 isolates from symptomatic leaf material collected in north-east Africa, the Middle

East (including Turkey) and India between 1993 and 2005. We also determined the full genome sequences of 19 dicot-infecting mastrevirus isolates recovered from symptomatic plant samples collected in Australia between 2002 and 2011. This dataset was analysed together with all previously described monocot- and dicot-infecting mastreviruses and through this we identified six divergent strains of CpCDV.

Despite 10 years of effort sampling dicotyledonous plant species and using methods such as rolling circle amplification and next generations sequencing to identify and recover circular ssDNA viruses from infected plant material, the only regions of the world where dicot infecting mastreviruses have been conclusively identified are the Middle East, East Africa, Australia and South Africa. However, fragments of a dicot-infecting mastrevirus-like genome have been discovered through deep sequencing of small RNAs extracted from Peruvian sweet-potatoes (Kreuze et al., 2009) suggesting that the currently known distribution of these viruses is almost definitely an under-estimation of their geographical range. It is nevertheless possible for us to determine which of the regions where these viruses have been sampled is nearest to their geographical origin. Our results support the prevailing notion that the degree of dicot-infecting mastrevirus diversity outside of Australia is lower than that within Australia and that the dicot-infecting viruses discovered in the former regions most likely originated either in or near Australia.

Results and discussion

Classification of new dicot-infecting mastrevirus full genome sequences

Forty-nine dicot-infecting mastrevirus genomes (Table 1) were recovered from chickpea (n = 40), lentil (n = 4), faba bean (n = 2), field pea (n = 2) and bean (n = 1). These 49 viral genomes and 48 others available in GenBank were assembled into a single dataset and genome-wide pairwise identities between every possible pair of sequences were calculated (1 minus p-distance calculated with pairwise deletion of gaps; Fig. 1A) so as to assess the over-all genetic diversity of these viruses. Based on the recommendations of Muhire et al. (2013) 18 of the 19 Australian dicot-infecting mastrevirus genomes could be assigned to previously named species and strain groupings; TYDV (1/19), CpCAV (7/19), CpCV-A (3/19), CpCV-B (1/19), and CpCV-E (6/19). The one exceptional Australian dicot-infecting isolate was clearly a member of the species CpCV but was ~ 87% similar to any previously described CpCV isolate and was therefore assigned to a new strain of this species: CpCV-F. The 30 dicot-infecting mastreviruses from north- east Africa, the Middle East and the Indian subcontinent were all CpCDV isolates, either classifiable as members of the previously described CpCDV strains -A (11/30), and -D (2/30), or, because they shared ~ 94% identity to isolates in previously described strains, were assigned to new strains -F (8/30), -G (2/30), -H (1/30), -I (1/30), -J (1/30) and -K (4/30).

It is evident both from the identity scores of all pairs of available dicot-infecting mastrevirus sequences and the maximum identity scores of all pairs of isolates within individual species that even within individual species there is a greater diversity amongst the known Australian dicot-infecting mastrevirus isolates than there is amongst the CpCDV isolates found across north-east Africa, South Africa, the Middle East, Turkey, Pakistan and India combined (Fig. 1).

Complex patterns of inter- and intra-species recombination amongst dicot-infecting mastreviruses

As has been demonstrated previously with smaller datasets, recombination has played a major role in the evolution of dicot-infecting mastreviruses (Hadfield et al., 2012; Martin et al., 2011b). A total of 16 intra-species and 10 inter-species recombination events were detected. Although 12 of the recombination events were detected here were previously identified by Martin et al. (2011b) and Hadfield et al. (2012), the additional full genome sequences generated during this study has increased the resolution with which many of these recombination events can be characterised (Fig. 2).

Several groups of isolates apparently carry evidence of multiple independent recombination events. For example, the CpCV-F isolate has evidence of one intra-species recombination event involving the acquisition by an ancestral CpCV-E-like virus of a *cp* gene fragment from a CpCV-C-like virus (event 1 in Fig. 2). The ancestral CpCV-E-like sequence from which the ancestor of the CpCV-F sequences was likely derived was, as is the case with all contemporary CpCV-E and CpCV-A sequences, in turn carrying evidence of a likely much older inter-species recombination event involving the transfer of a *rep* gene fragment from a CpCAV-like sequence into the genome of a CpCV-B-like sequence (Event F in Fig. 2). More recently than the two previously discussed events detectable within the CpCV-F sequences, a small region of the SIR of a common ancestor of these sequences appears to have been derived by recombination from a currently unknown monocot-infecting mastrevirus species (Event G in Fig. 2). Similarly complex recombination patterns are detectable within the sampled CpCV-A and CpCDV-K genomes suggesting that such convoluted evolutionary histories might be fairly common amongst dicot-infecting mastreviruses.

Consistent with previous analyses of the monocot-infecting mastreviruses, we detected (1) that intra-species recombination events, in most cases, have tended to involve transfers of larger genome fragments (average of 22% ranging between 10% and 49% of the genome) than inter-species recombination events (average of 17% ranging between 10% and 30% of the genome; Martin et al., 2001; Varsani et al., 2009a, 2008b) and (2) that there are clear recombination breakpoint hotspots within the LIR and SIR genome regions (Martin et al., 2011b), and (3) a greater number of recombination breakpoints in the complementary sense genes than in the virion sense genes (Hadfield et al., 2012; Kraberger et al., 2012; Martin et al., 2011b; Owor et al., 2007; Varsani et al., 2009a, 2008b). The concentration of recombination breakpoints within the intergenic regions of these viruses enabled us to construct two relatively recombination-free datasets corresponding to the *cp* and *rep* gene regions of the full genome dataset – hereafter respectively referred to as the CP and Rep datasets.

Table 1
Host and country of origin details for all full dicot-infecting mastrevirus genomes deposited in GenBank, including those from this study. GenBank accessions in bold were are genomes determined in this study.

Species	Strain	GenBank no.	Country	Host common name	Host	Sampling year
CpCDV	**CpCDV-A**	FR687959	Syria	Chickpea	*Cicer arietinum*	2008
		KC172662	Turkey	Chickpea	*C. arietinum*	1996
		KC172663	Iran	Chickpea	*C. arietinum*	1996
		KC172655	Iran	Chickpea	*C. arietinum*	1999
		KC172653	Iran	Chickpea	*C. arietinum*	1999
		KC172654	Iran	Chickpea	*C. arietinum*	2002
		KC172656	Iran	Chickpea	*C. arietinum*	1999
		KC172657	Iran	Chickpea	*C. arietinum*	1999
		KC172658	Iran	Chickpea	*C. arietinum*	1999
		KC172659	Iran	Chickpea	*C. arietinum*	1999
		KC172660	Iran	Chickpea	*C. arietinum*	1999
		KC172661	Iran	Field Pea	*Pisum sativum*	1999
	CpCDV-B	Y11023	South Africa	Bean	*Phaseolus vulgaris*	1997
		DQ458791	South Africa	Bean	*P. vulgaris*	1997
		AM849096	Pakistan	Chickpea	*C. arietinum*	2005
	CpCDV-C	AM849097	Pakistan	Chickpea	*C. arietinum*	2005
		AM850136	Pakistan	Chickpea	*C. arietinum*	2007
		AM900416	Pakistan	Chickpea	*C. arietinum*	2007
	CpCDV-D	FR687960	Pakistan	Chickpea	*C. arietinum*	2008
		KC172664	India	Chickpea	*C. arietinum*	1993
		KC172665	India	Field Pea	*P. sativum*	1993
	CpCDV-E	AM933135	Sudan	Chickpea	*C. arietinum*	1997
		AM933134	Sudan	Chickpea	*C. arietinum*	1997
	CpCDV-F	**KC172666**	Pakistan	Lentil	*Lens culinaris*	1997
		KC172669	Yemen	Lentil	*L. culinaris*	1996
		KC172672	Yemen	Lentil	*L. culinaris*	1996
		KC172673	Yemen	Lentil	*L. culinaris*	1996
		KC172670	Yemen	Faba bean	*Vicia faba*	1996
		KC172671	Yemen	Faba bean	*V. faba*	1996
		KC172667	Syria	Chickpea	*C. arietinum*	2003
		KC172668	Syria	Chickpea	*C. arietinum*	1999
	CpCDV-G	**KC172674**	Eritrea	Chickpea	*C. arietinum*	2005
		KC172675	Eritrea	Chickpea	*C. arietinum*	2005
	CpCDV-H	**KC172676**	Eritrea	Chickpea	*C. arietinum*	2005
	CpCDV-I	**KC172677**	Eritrea	Chickpea	*C. arietinum*	2005
	CpCDV-J	**KC172678**	Eritrea	Chickpea	*C. arietinum*	2005
	CpCDV-K	**KC172679**	Eritrea	Chickpea	*C. arietinum*	2005
		KC172680	Eritrea	Chickpea	*C. arietinum*	2005
		KC172681	Eritrea	Chickpea	*C. arietinum*	2005
		KC172682	Eritrea	Chickpea	*C. arietinum*	2005
CpCV	**CpCV-A**	GU256530	Australia	Chickpea	*C. arietinum*	2002
		JN989413	Australia	Chickpea	*C. arietinum*	2002
		JN989414	Australia	Chickpea	*C. arietinum*	2002
		JN989415	Australia	Chickpea	*C. arietinum*	2002
		KC172685	Australia	Chickpea	*C. arietinum*	2010
		KC172683	Australia	Chickpea	*C. arietinum*	2002
		KC172684	Australia	Chickpea	*C. arietinum*	2002
	CpCV-B	GU256531	Australia	Chickpea	*C. arietinum*	2003
		KC172690	Australia	Chickpea	*C. arietinum*	2011
	CpCV-C	JN989416	Australia	Chickpea	*C. arietinum*	2002
		JN989417	Australia	Chickpea	*C. arietinum*	2002
	CpCV-E	JN989438	Australia	Bean	*Phaseolus sp.*	1984
		JN989426	Australia	Chickpea	*C. arietinum*	2002
		JN989437	Australia	Chickpea	*C. arietinum*	2002
		JN989429	Australia	Chickpea	*C. arietinum*	2002
		JN989434	Australia	Chickpea	*C. arietinum*	2002
		JN989428	Australia	Chickpea	*C. arietinum*	2002
		JN989430	Australia	Chickpea	*C. arietinum*	2002
		JN989431	Australia	Chickpea	*C. arietinum*	2002
		JN989432	Australia	Chickpea	*C. arietinum*	2002
		JN989433	Australia	Chickpea	*C. arietinum*	2002
		KC172699	Australia	Chickpea	*C. arietinum*	2002
		KC172698	Australia	Chickpea	*C. arietinum*	2002
		KC172694	Australia	Chickpea	*C. arietinum*	2002
		KC172695	Australia	Chickpea	*C. arietinum*	2002
		KC172696	Australia	Chickpea	*C. arietinum*	2002
		KC172697	Australia	Chickpea	*C. arietinum*	2002
	CpCV-F	**KC172700**	Australia	Chickpea	*C. arietinum*	2002
CpCAV		JN989418	Australia	Bean	*P. vulgaris*	2007
		JN989419	Australia	Chickpea	*C. arietinum*	2010
		JN989420	Australia	Chickpea	*C. arietinum*	2010
		JN989421	Australia	Chickpea	*C. arietinum*	2010

Table 1 (*continued*)

Species	Strain	GenBank no.	Country	Host common name	Host	Sampling year
		JN989422	Australia	Chickpea	*C. arietinum*	2002
		JN989423	Australia	Chickpea	*C. arietinum*	2003
		KC172691	Australia	Chickpea	*C. arietinum*	2011
		KC172693	Australia	Chickpea	*C. arietinum*	2011
		KC172692	Australia	Chickpea	*C. arietinum*	2011
		KC172689	Australia	Chickpea	*C. arietinum*	2002
		KC172686	Australia	Chickpea	*C. arietinum*	2003.
		KC172687	Australia	Chickpea	*C. arietinum*	2003
		KC172688	Australia	Chickpea	*C. arietinum*	2003
CpRLV		GU256532	Australia	Chickpea	*C. arietinum*	2003
CpYV		JN989439	Australia	Chickpea	*C. arietinum*	2002
TYDV		M81103	Australia	Tobacco	*Nicotinana sp.*	1992
		JN989440	Australia	Tobacco	*Nicotinana sp.*	1986
		JN989445	Australia	Tobacco	*Nicotinana sp.*	1985
		JN989446	Australia	Tobacco	*Nicotinana sp.*	2002
		JN989441	Australia	Bean	*P. vulgaris*	2010
		JN989442	Australia	Bean	*P. vulgaris*	2010
		JN989443	Australia	Bean	*P. vulgaris*	2010
		KC172702	Australia	Bean	*P. vulgaris*	2010
		JN989444	Australia	Chickpea	*C. arietinum*	2002

Contrary to recombination detected amongst the monocot- infecting mastreviruses (Monjane et al., 2011; Varsani et al., 2009a, 2008a, 2008b), and far more reminiscent of recombination patterns detectable in the dicot-infecting begomoviruses (Lefeuvre et al., 2007; Padidam et al., 1999), is the fact more inter-species recombination events are detectable in the dicot-infecting mastreviruses than intra-species recombination events (16 vs 10 events respectively). It is unclear why inter-species recombination might be more common in the dicot-infecting geminiviruses than it is in the monocot-infecting geminiviruses. It might suggest that the plant and/or geographical host ranges of dicot-infecting mastreviruses overlap, or at least have overlapped, more than those of the monocot mastreviruses, leading to more frequent co-infection of plants, a prerequisite for recombination to occur. Alternatively there may be a greater selection pressure against sequence changes in monocot-infecting which tend to have a narrow host range. As more sequence data accumulates for these groups and for the newly discovered monocot-infecting gemini-viruses, potentially representing a new genus (Varsani et al., 2009b) it will be interesting to see whether this pattern holds.

The geographical origin of the dicot-infecting mastreviruses

The WDV-rooted ML phylogenetic tree constructed from sequences with the tracts of recombinationally derived sequence removed, indicated that the MRCA of these viruses (the node at the root of the tree in Fig. 1) is probably Australian. Also, as has been suggested in previous analyses the diversity of dicot-infecting mastreviruses in Australia is clearly far greater than that seen amongst the currently sampled African, Middle-Eastern, Turkish and Indo-Pakistani sequences.

Given that the sequences examined here were sampled over a period of only 27 years (1984–2011) it was unsurprising that our three datasets yielded only weak support for the presence of a molecular clock signal (Path-O-Gen derived correlation coefficients ranging between 0.20 and 0.25). Since this indicated that the analysed datasets could not be productively used to estimate accurate nucleotide substitution rates, it was not possible for us to accurately date any of the historical dispersal events indicated by our phylogeographic analyses. Nevertheless, of the various molecular clock (strict and relaxed) and demographic (constant population

size, Bayesian skyline plot) models tested the constant population size+relaxed-clock model fitted the data best.

The maximum clade credibility (MCC) trees constructed using these models applied to the full genome, Rep, and CP datasets with sequences sampled from the Western Mediterranean (WM), Asian (AS) the Middle Eastern (ME), East African (EA), Southern African (SA) and Australian (AU) regions are presented in Fig. 3 (Supplementary Figs. S1,S2). For all of the analysed datasets Australia was indicated the most likely origin of the MRCA of all the analysed viruses (note the colour of the lines at the basal nodes of the trees in Fig. 3, Supplementary Figs. S1–S5). Specifically, Australia had 0.8735 posterior probability support as the root location state for the CP dataset, 0.8333 for the Rep dataset, and 0.6932 for the full genome dataset.

When the same data were analysed but with the sampling locations randomized amongst the analysed sequences the most probable root locations were inferred to be either East Africa for the CP dataset (P = 0.1789) or the Middle east for the Rep (P = 0.1697) and full genome dataset (P = 0.1701) suggesting that our results were not inherently biased in favour of identifying Australia as the location of the MRCA (Fig. 3, Supplementary Figs. S1 and S2).

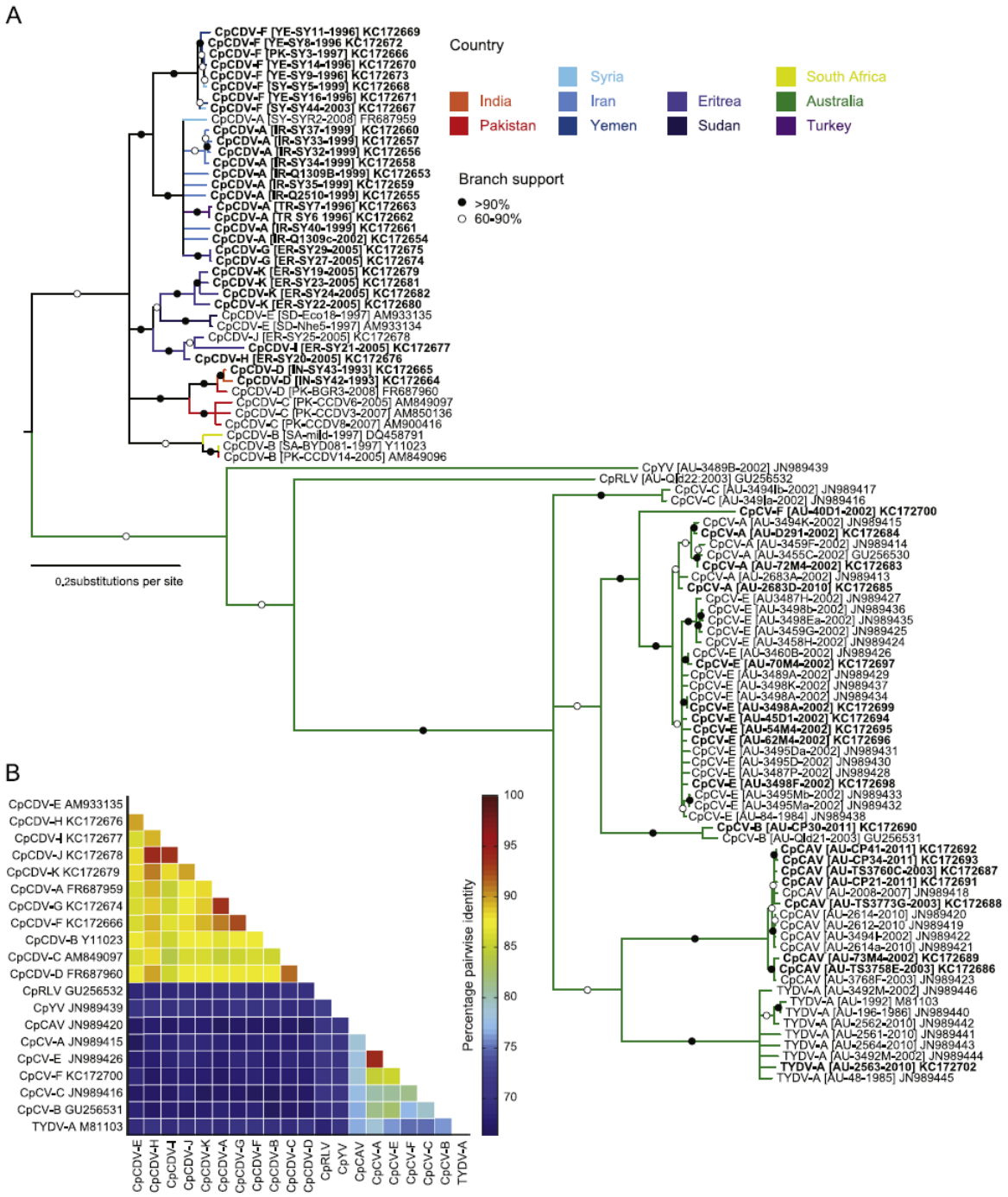
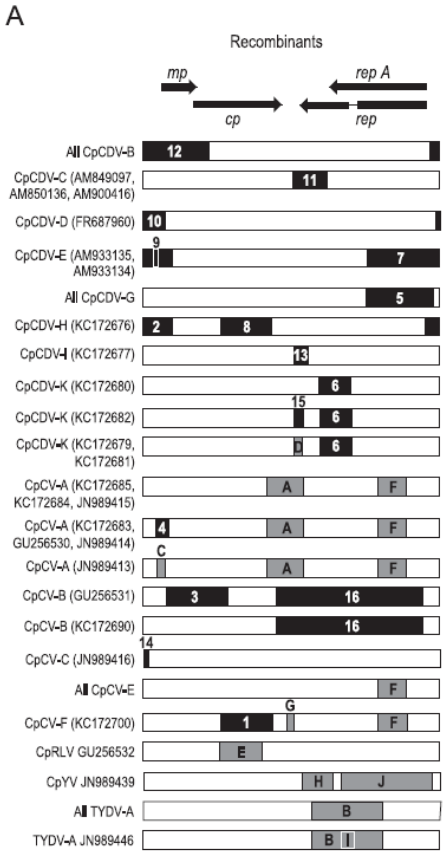


Fig. 1. (A) Maximum likelihood phylogenetic tree (constructed with the nucleotide substitution model GTR+G4) of all available dicot-infecting mastrevirus full genome sequences (with recombinant regions removed). The trees were rooted with WDV. Bootstrap support for branches is indicated by open (60–89%) and closed circles (> 90%), branches with less than 60% bootstrap support have been collapsed. Countries of origin are represented by colours shown in key. Viral isolate sequences determined in this study have accession numbers KC172653–KC172702. (B) Two dimensional percentage pairwise identity plot matrix of a representative dicot-infecting mastrevirus from each strain and species.



B

Event	Recombinant region	Potential major Parent	Potential minor Parent	Detection method	P-value
Intra-species recombination					
1	646-1127	All CpCV-E	CpCV-C (JN989416, JN989417)	RGMCT	$7,45x10^{-21}$
2	2465-212	CpCDV-J (KC172678)	Unknown	RGMCT	$2,24x10^{-12}$
3	141-697	CpCV-B (KC172690)	CpCV-A (KC172683, KC172684, KC172685, GU256530, JN989413, JN989414, JN989415), All CpCV-E, CpCV-F (KC172700)	RMCT	$1,18x10^{-11}$
4	94-170	CpCV-A (KC172684, KC172685), All CpCV-E	CpCV-F (KC172700)	RGT	$2,23x10^{-6}$
5	1942-2543	CpCDV-A (KC172653, KC172654, KC172655, KC172663, KC172657, KC172661FR687959)	All CpCDV-F	RGBMT	$1,71x10^{-27}$
6	1520-1798	CpCDV-H, CpCDV-I, CpCDV-J, CpCDV-E (AM933135, AM933135)	Unknown	RMCT	$2,71x10^{-27}$
7	1947-205	All CpCDV-K	Unknown	RMCT	$4,70x10^{-27}$
8	614-1083	CpCDV-J	All CpCDV-K	RMCT	$1,62x10^{-25}$
9	66-110	Unknown	CpCDV-H (KC172676), CpCDV-C (AM849097)	RMC	$7,23x10^{-23}$
10	2543-150	Unknown	CpCDV-C (AM849097, AM900416)	RGMCT	$6,64x10^{-28}$
11	1259-1580	All CpCDV-D	Unknown	RGMCT	$4,52x10^{-26}$
12	2555-504*	CpCDV-F (KC172666, KC172669)	CpCDV-K (KC172680)	RMCT	$5,36x10^{-25}$
13	1257-1364	CpCDV-J, CpCDV-H	Unknown	RGB	$9,09x10^{-20}$
14	4-57	CpCV-C (JN989417)	Unknown, All CpCV-E	RGB	$5,22x10^{-19}$
15	1279-1343	CpCDV-K (KC172680)	All CpCDV-G	RGB	$1,11x10^{-16}$
16	1158-2487	All CpCAV	Unknown	MCST	$1,90x10^{-11}$
Inter-species recombination					
A	1061-1366	All CpCV-E	All TYDV-A	RGMCT	$9,24x10^{-41}$
B	1447-2102	All CpCV-A, All CpCV-E, All CpCAV, All CpCV-B, CpCV-F, CpCV-C (JN989416, JN989417)	All CpCDV	RGMCT	$7,44x10^{-19}$
C	111-159	All CpCV-E, CpCV-F, CpCV-A (KC172684, KC172683, KC172685, GU256530, JN989414, JN989415)	All TYDV-A	RGT	$8,14x10^{-15}$
D	1279-1335	CpCDV-K (KC172680)	All TYDV-A, CpCV-A (GU256530)	RGBM	$6,16x10^{-13}$
E	634-1027	Unknown	CpYV	RMC	$1,64x10^{-23}$
F	2078-2345	All CpCV-B	All CpCAV	RMC	$1,87x10^{-26}$
G	1247-1297	All CpCV-E	Unknown	RBS	$5,23x10^{-25}$
H	1345-1634	CpCDV-E (AM933135, AM933134), All CpCDV-A, All CpCDV-F, CpCDV-H, CpCDV-I, CpCDV-J	All CpCAV	RMS	$3,35x10^{-24}$
I	1725-1828	TYDV-A (JN989443, KC172702, JN989440, JN989441, JN989444, M81103)	All CpCDV-A	RMCT	$4,72x10^{-23}$
J	1706-2509	CpCDV-C (AM850136, AM849097, AM900416)	All CpCAV	RBMCS	$3,58x10^{-29}$

RDP (R) GENCONV (G), BOOTSCAN (B), MAXCHI (M), CHIMERA (C), SISCAN (S) and 3SEQ (T)
* = The actual breakpoint position is undetermined.

Fig. 2. (A) Illustration of recombination events amongst all dicot-infecting mastrevirus isolates. Inter-species recombination events are represented in grey and have an associated letter code. Intra-species events are represented in black and have an associated number code. Arrows above the genome maps indicate the positions on these maps of the *mp* (movement protein), *cp* (coat protein), *repA* (replication associated A protein) and *rep* (replication associated protein) genes. (B) Details of all recombination events detected using RDP4. Major and minor parents are inferred based on genetic fragments they donated to the recombinant, with the major parent donating the larger fragment and the minor parent the smaller fragment. Methods used to detect recombination are as follows RDP (R) GENCONV (G), BOOTSCAN (B), MAXCHI (M), CHIMERA (C), SISCAN (S) and 3SEQ (T). The method with the most significant associated *p*-value is indicated in bold for each event.

Plausible routes of dicot infecting mastrevirus movement out of Australia

Collectively four statistically supported (BF_{log10} 45.0) virus movements between the six analysed locations were inferred from the three analysed datasets (Fig. 3). These involved initial movements out of Australia to both South Africa (BF_{log10} ¼ 179.8, 69.0, 26.9) and to the horn of Africa (BF_{log10} ¼ 21.8, 25.7, 5.2) with subsequent dispersal from the Middle East to Asia (BF_{log10} ¼ 550.1, 56.7, 363.1), and from horn of Africa to the Middle East (BF_{log10} ¼ 203.7, 435.4, 41.9) for the full genome, rep and CP datasets respectively.

Conclusion

Dicot-infecting mastreviruses have been identified in Australia, Africa, the Middle East and the Indian subcontinent as important crop pathogens. This study extends our current knowledge of the diversity of these viruses in these regions with the addition of 49 full genomes. Amongst these genomes are isolates of seven new divergent strains from two different species. Of particular interest is our recombination analysis which revealed a surprisingly high level of inter-species recombination events between dicot-infecting mastrevirus from two geographically distant regions, a pattern which while consistent with that found in dicot-infecting begomoviruses, is in contrast with that found in the monocot-infecting mastreviruses (Kraberger et al., 2012; Shepherd et al., 2010; Varsani et al., 2008a, 2008b). Such a high frequency of recombination events coupled with

evidence that recombination has likely contributed to the emergence of various geminiviruses as agricultural pests during modern times (Rocha et al., 2013; Varsani et al., 2008b), highlights the importance of continual surveillance to monitor for the presence and identities of these viruses in the environment so as to identify potentially new pathogens that may evolve to threaten agriculture.

Pulses were among the first cultivated plants, with some of the oldest archaeobotanical evidence indicating that the Middle East is one of the ancient centres of this practice (Mikić, 2012; Tanno and Willcox, 2006). Given that the Middle East and surrounding countries have such a long history of the cultivation of pulses in comparison with Australia it is surprising that Australia harbours a greater diversity of dicot-infecting mastreviruses than Africa, the Middle East and India combined. The corrective measures that we have taken to account for recombination and sampling biases strengthen our conclusion that the MRCA of the currently known dicot-infecting mastreviruses is most likely nearer Australia than the other sampling locations that were considered. It is nevertheless important to stress that Australia is merely the region amongst those that have been sampled where the MRCA of the analysed sequences originated. The MRCA of these sequences could have actually existed in any of the many regions of the world where samples have not been collected, with descendants of these sequences having simply passed through Australia enroute to the other geographical regions that have been considered here.

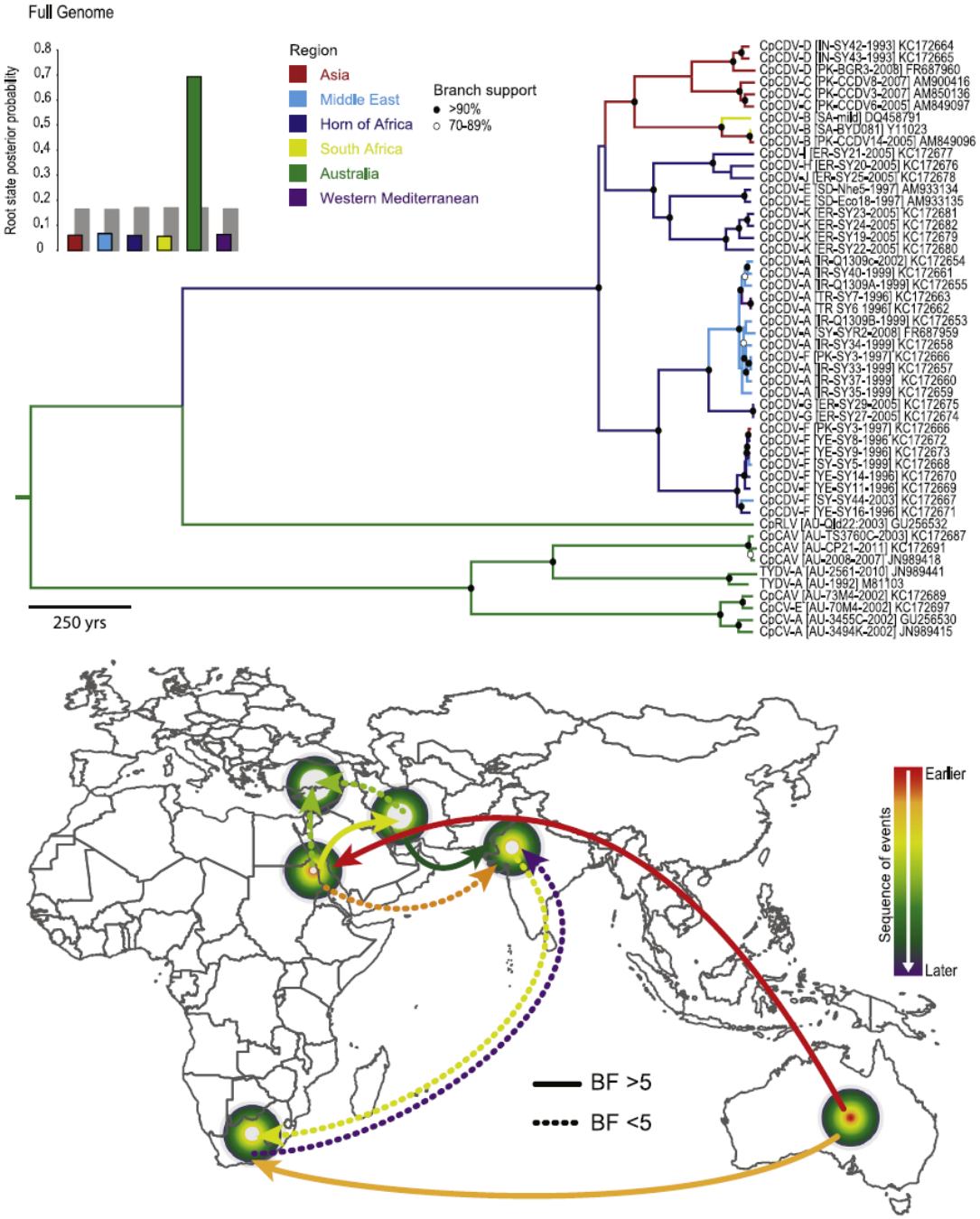


Fig. 3. (A) Maximum clade credibility tree constructed from the dicot-infecting mastrevirus full genome dataset under the GTR+G4 nucleotide substitution model, constant population size demographic model, a relaxed-clock evolutionary model and a discretised spatial diffusion phylogeographic model. This later model considered spatial diffusion between six geographic locations and included only a randomly selected subset of 10 of the Australian mastreviruses included in Fig. 1. Branches and taxon names are coloured according to the region where they were collected. Posterior support greater than 90% is indicated by a filled circle and greater than 70% by an open circle at the nodes. Probabilities obtained with randomisation of the tip locations are provided as grey bars for each location. (B) Plausible historical movement pathways of dicot-infecting mastreviruses inferred using the full genome dataset. The spatial dynamics of dicot-infecting mastreviruses movements were inferred using the discrete phylogeographic model considering only the six geographical regions from which the analysed viruses were sampled.

Similarly, the MRCA of the sequences considered here is not necessarily the MRCA of all the dicot-infecting mastreviruses currently circulating on Earth, and is almost certainly also not the "first" mastrevirus that infected dicotyledonous hosts. Given that fragments of a highly divergent virus genome resembling those of dicot-infecting mastreviruses has been detected in the Peruvian sweet potato germplasm collection (Kreuze et al., 2009), it is entirely plausible that the viruses considered here are part of a much more diverse, but currently undiscovered, global dicot-infecting mastrevirus population. Without much more intensive sampling of dicot-infecting mastreviruses, both in the regions considered here and across the vast areas of Asia, Africa, the Pacific

Rim and the Americas where these viruses have remained unsampled, we cannot yet hope to pinpoint the actual geographical origins of either the MRCA of all dicot-infecting mastreviruses or the location of the first dicot-infecting mastrevirus. With the application of modern molecular tools and new metagenomic approaches to mastrevirus discovery (Rosario et al., 2013), we anticipate that there will be a rapid increase in the diversity of known dicot-infecting mastreviruses that should greatly increase the resolution with which the movement pathways and geographically origins of these viruses can be determined.

GenBank accession♯: KC172653–KC172702

Materials and methods

Sample collection, virus isolation and genome cloning

Samples from 49 pulses [chickpea (*C. arietinum*), lentil (*L. culinaris*), faba bean (*Vicia faba*), field pea (*Pisum sativum*) and bean (*P. vulgaris*), collected in Syria (*n* = 2), Pakistan (*n* = 1), India (*n* ¼ 2), Turkey (*n* = 2), Eritrea (*n* = 9), Iran (*n* = 9), Yemen (*n* = 5) and Australia (*n* = 19)] which had previously been identified to be positive for mastreviruses either by PCR or ELISA were used in this study (Supplementary Table S1 details host species for each sample). Total DNA was extracted from plant sap or dried plant material using Epoch nucleic acid purification kits (Epoch Life Science, USA). Enrichment of circular viral DNA from total DNA was carried out using the Illustra TempliPhi Amplification Kit (GE Healthcare, USA) as previously described by Owor et al. (2007) and Shepherd et al. (2008). Viral DNA amplicons were then digested using the restriction enzymes *Hind*III or *Xmn*I which yielded "' 2.6 kb linearised unit length genomes. These were gel purified and ligated at either the *Hind*III or *Xmn*I sites of the cloning vector, pGEM3Zf+ (Promega Biotech, USA).

We used a polymerase chain reaction (PCR) amplification approach to recover viral genomes from 44 of the 49 TempliPhi enriched DNA samples for which we were unable to find a unique restriction enzyme. Degenerate back-to-back primers (dicot for- ward 5′-GAN TTG GTC CGC AGT GTA GA-3′, dicot reverse 5′-GTA CCG GWA AGA CMW CYT GG-3′), previously described by Hadfield et al. (2012) were used to amplify full length dicot-infecting mastrevirus genomes using Kapa HiFi HotStart DNA polymerase (Kapa Biosystems, USA) with the following thermocycling conditions: 94 1C for 3 min, 25 cycles of 98 1C (3 min), 52 1C (30 sec), 72 1C (2.45 min) and a final extension of 72 1C for 3 min. PCR amplicons were ligated into linearised pJET1.2 vector (CloneJET™ PCR cloning kit, Fermentas, USA). All plasmids with cloned viral genomes were sequenced at Macrogen (Korea) by primer walking.

Sequence assembly and pairwise sequence analyses

Viral genome sequences were assembled using DNAMAN (version 7; Lynnon Biosoft, Canada). Forty-eight dicot-infecting mastrevirus full genome sequences available in public databases on 24 October 2012 and the wheat dwarf virus sequence (AM040732; included as an outlier) were obtained and aligned with the sequences determined in this study using MUSCLE (Edgar, 2004). The nucleotide sequence alignment thus obtained was manually edited using MEGA5 (Tamura et al., 2011). Similarly, putative Rep, MP and CP encoding sequences of the 97 virus genomes were computationally

translated and aligned using MEGA5 with manual editing. Pairwise identities (1 – *p*-distance, with pairwise deletion of gaps) of the full dicot-infecting mastrevirus genomes were determined using SDT v1.0 (Muhire et al., 2013).

Recombination analysis and construction of mostly recombination-free datasets

Recombination analysis within the dicot-infecting mastreviruses was performed using RDP4 (Martin et al., 2010), with the following methods: RDP, GENECONV (Padidam et al., 1999), Bootscan (Martin et al., 2005), Maxchi (Smith, 1992), Chimera (Posada and Crandall, 1998), Siscan (Gibbs et al., 2000), and 3Seq (Boni et al., 2007). Potential recombination signals were accepted as being genuine evidence of actual recombination events when they were detected with three or more of the seven methods (with associated *p*-values of $\mathbf{O}\ 10^{-3}$) coupled with phylogenetic support for recombination having occurred.

Based on the recombination analysis two mostly recombination-free sequence alignments corresponding to a coat protein (CP) gene dataset and a Rep gene dataset were extracted from the full genome sequence alignments.

Phylogenetic analyses and identification of the likely origin of dicot-infecting mastreviruses

A maximum likelihood (ML) phylogenetic tree of the aligned full genome sequences, with recombinant region removed, was constructed using PHYML version 3 (Guindon et al., 2010) with 1000 non-parametric bootstrap replicates with GTR+G4 selected as the best fit nucleotide substitution model using RDP 4 (Martin et al., 2010) and rooted with *Wheat dwarf virus* (WDV). Branches with less that 60% bootstrap support were manually collapsed using MESQUITE (Version 2.75).

We opted to use Bayesian maximum clade credibility (MCC) trees produced using the computer program BEAST (Drummond et al., 2012) to evaluate the likely geographical origin of the dicot-infecting mastreviruses. These trees were time-calibrated based on sequence sampling times with the root location based on the most plausible dating of the most recent common ancestor (MRCA) of the analysed sequences. Each of the MCC trees produced by BEAST represented an entire distribution of similarly plausible trees and explicitly accounted for phylogenetic uncertainty during their inference. Furthermore, besides offering fully probabilistic models of sequence evolution, BEAST also implements phylogeographic models of sequence movement between discrete sampling locations (such as between cities, provinces, countries or other discrete geographical regions). These models have been employed previously to investigate the movement dynamics the monocot-infecting mastrevirus species, *Maize streak virus* (Monjane et al., 2011) and the begomovirus species, *Tomato yellow leaf curl virus* (Lefeuvre et al., 2010) and *East African cassava mosaic virus* (De Bruyn et al., 2012). The discrete phylogeography model used here to infer when and where the MRCA of the dicot-infecting mastreviruses existed considered geographic diffusion among six discrete sample locations: the Western Mediterranean (WM), Asia (AS) the Middle East (ME), East Africa (EA), Southern Africa (SA) and Australia (AU).

Since previous analyses have indicated that sampling biases can strongly influence the phylogeographic inference of ancestral sequence locations in BEAST (De Bruyn et al., 2012; Lefeuvre et al., 2010; Monjane et al., 2011) we took steps to both directly reduce the influences of these biases prior to analyses and to test for the effects of any biases after the analyses were concluded. Specifically, we randomly removed all but 10 of the Australian sequences from the full genome, CP and Rep datasets pre-analysis. Post-analysis we directly evaluated the effects of residual sampling biases on the inferred geographical location of the MRCA by randomly swapping sampling locations among the sequences followed by revaluation of the MRCA location state. This test would indicate that a sampling bias had influenced inference of the MRCA location if the same locations(s) were indicated for the MRCA in both the randomised and un-randomised analyses.

For each of the analysed datasets independent replicate runs of the Markov chain of 2×10^7 steps were performed using BEAST so as to achieve effective sample size (ESS) estimates for all relevant model parameters that were always 4200. The degree of clock-like evolution evident within the analysed sequence datasets (full genome, CP and Rep) was evaluated using root-to-tip genetic distance vs sampling date regression analyses based on inferred neighbor-joining trees using the computer program, Path-O-Gen (available from http://tree.bio.ed.ac.uk/software/pathogen/) (Drummond et al., 2003).

We used the computer program SPREADv1.0.4 (Bielejec et al., 2011) (available from http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html) to perform Bayes factor (BF) tests of potential epidemiological links between the analysed geographical regions revealed by the phylogeographic analyses performed by BEAST. In these tests we accepted BF_{log10} values greater than or equal to 5.0 as being indicative of significant statistical support for movement between pairs of geographical regions (where a BF_{log10} 4100 was taken to represent decisive support, a BF_{log10} 410.0 was taken to represent strong support and a BF_{log10} o 5.0 was taken to represent poor support.). SPREAD was then used to produce .kml formatted files containing information on BF test supported routes of virus movement. These files can be viewed using the computer program, Google Earth (available from http://earth.google.com).

Acknowledgments

The molecular work was supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand (UOC0903) awarded to Arvind Varsani. Field collections in 2002–2003 were funded by New South Wales Department of Primary Industries, The State of Queensland Department of Employment, Economic Development and Innovation (DEEDI, formerly Queensland Department of Primary Industries & Fisheries), and Grains Research and Development Corporation project DAN00023 awarded to Mark Schwinghamer and John E Thomas. Simona Kraberger was supported by a School of Biological Sciences (University of Canterbury, New Zealand) postgraduate scholarship.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.06.024.

References

Ali, M.A., Kumari, S.G., Makkouk, K.H., Hassan, M.M., 2004. Chickpea chlorotic dwarf virus, CpCDV naturally infects *Phaseolus* bean and other wild species in the Gezira region of Sudan. A. J. Plant Prot. 22, 96.

Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910–2912.

Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for
inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, M., Abdoul-Karime, A.L., Abdou-Chakour, C., Reynaud, B., Harkins, G.W., Varsani, A., Martin, D.P., Lett, J.M., 2012. East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. BMC Evololut. Biol. 12, 228.

Dekker, E.L., Woolston, C.J., Xue, Y., Cox, B., Mullineaux, P.M., 1991. Transcript mapping reveals different expression strategies for the bicistronic RNAs of the geminivirus wheat dwarf virus. Nucleic Acids Res. 19, 4075–4081.

Drummond, A., Pybus, O.G., Rambaut, A., 2003. Inference of viral evolutionary rates from molecular sequences. Adv. Parasitol. 54, 331–358.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Farzadfar, S., Pourrahim, R., Golnaraghi, A.R., Shahraeen, N., Makkouk, K.M., 2002. First report of sugar beet and bean as natural hosts of Chickpea chlorotic dwarf virus in Iran. Plant Pathol. 51 795–795.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hadfield, J., Thomas, J.E., Schwinghamer, M.W., Kraberger, S., Stainton, D., Dayaram, A., Parry, J.N., Pande, D., Martin, D.P., Varsani, A., 2012. Molecular characterisation of dicot-infecting mastreviruses from Australia. Virus Res. 166, 13–22.

Halley-Stott, R.P., Tanzer, F., Martin, D.P., Rybicki, E.P., 2007. The complete nucleotide sequence of a mild strain of Bean yellow dwarf virus. Arch. Virol. 152, 1237–1240.

Harrison, B.D., 1985. Advances in geminivirus research. Annu. Rev. Phytopathol. 23, 55–82.
Heyraud, F., Matzeit, V., Schaefer, S., Schell, J., Gronenborn, B., 1993. The conserved nonanucleotide motif of the geminivirus stem-loop sequence promotes replicational release of virus molecules from redundant copies. Biochimie 75, 605–615.
Horn, N.M., Reddy, S.V., Reddy, D.V.R., 1994. Virus-vector relationships of chickpea chlorotic dwarf geminivirus and the leafhopper Orosius orientalis (Hemiptera: Cicadellidae). Ann. Appl. Biol. 124, 441–450.
Horn, N.M., Reddy, S.V., Roberts, I.M., Reddy, D.V.R., 1993. Chickpea chlorotic dwarf virus, a new leafhopper-transmitted geminivirus of chickpea in India. Ann. Appl. Biol. 122, 467–479.
Jeske, H., Lütgemeier, M., Preiß, W., 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. EMBO J. 20, 6158–6167.
Knights, E.J., Açikgöz, N., Warkentin, T., Bejiga, G., Yadav, S.S., Sandhu, J.S., 2007. Area, production, and distribution. In: Yadav, S.S., Redden, R., Chen, W., Sharma, B. (Eds.), Chickpea Breeding and Management. CABI, Wallingford, pp. 167–178.
Kraberger, S., Thomas, J.E., Geering, A.D.W., Dayaram, A., Stainton, D., Hadfield, J., Walters, M., Parmenter, K.S., van Brunschot, S., Collings, D.A., Martin, D.P., Varsani, A., 2012. Australian monocot-infecting mastrevirus diversity rivals that in Africa. Virus Res. 169, 127–136.
Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R., 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. Virology 388, 1–7.
Kumari, S.G., Makkouk, K.M., Attar, N., Ghulam, W., Lesemann, D.E., 2004. First report of chickpea chlorotic dwarf virus infecting spring chickpea in Syria. Plant Dis. 88 424–424.
Kumari, S.G., Makkouk, K.M., Loh, M.H., Negassi, K., Tsegay, S., Kidane, R., Kibret, A., Tesfatsion, Y., 2008. Viral diseases affecting chickpea crops in Eritrea. Phytopathol. Mediter. 47, 42–49.
Lefeuvre, P., Martin, D.P., Harkins, G., Lemey, P., Gray, A.J., Meredith, S., Lakay, F., Monjane, A., Lett, J.M., Varsani, A., Heydarnejad, J., 2010. The spread of tomato yellow leaf curl virus from the middle east to the world. PLoS Pathogens 6, e1001164.
Lefeuvre, P., Martin, D.P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A., Becker, N., Reynaud, B., Lett, J.-M., 2007. Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. J. General Virol. 88, 3458–3468.
Makkouk, K.M., Rizkallah, L., Kumari, S.G., Zaki, M., Enein, R.A., 2003. First record of chickpea chlorotic dwarf virus (CpCDV) affecting faba bean (*Vicia faba*) crops in Egypt. Plant Pathol. 52 413–413.
Martin, D.P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., Varsani, A., 2011a. Recombination in eukaryotic single stranded DNA viruses. Viruses 3, 1699–1738.

Martin, D.P., Briddon, R.W., Varsani, A., 2011b. Recombination patterns in dicot-infecting mastreviruses mirror those found in monocot-infecting mastreviruses. Arch. Virol. 156, 1463–1469.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses 21, 98–102.

Martin, D.P., Willment, J.A., Billharz, R., Velders, R., Odhiambo, B., Njuguna, J., James,
D., Rybicki, E.P., 2001. Sequence diversity and virulence in Zea mays of maize streak virus isolates. Virology 288, 247–255.

Mikić, A., 2012. Origin of the words denoting some of the most ancient old world pulse crops and their diversity in modern european languages. PLoS ONE 7, e44512.

Monjane, A.L., Harkins, G.W., Martin, D.P., Lemey, P., Lefeuvre, P., Shepherd, D.N., Oluwafemi, S., Simuyandi, M., Zinga, I., Komba, E.K., Lakoutene, D.P., Mandakombo, N., Mboukoulida, J., Semballa, S., Tagne, A., Tiendrébéogo, F., Erdmann, J.B., van Antwerpen, T., Owor, B.E., Flett, B., Ramusi, M., Windram, O.P., Syed, R., Lett, J.-M., Briddon, R.W., Markham, P.G., Rybicki, E.P., Varsani, A., 2011. Reconstructing the history of maize streak virus Strain A dispersal to reveal diversification hot spots and its origin in Southern Africa. J. Virol. 85, 9623–9636.

Morris, B.A.M., Richardson, K.A., Haley, A., Zhan, X., Thomas, J.E., 1992. The nucleotide sequence of the infectious cloned dna component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants. Virology 187, 633–642.

Muhire, B., Martin, D., Brown, J., Navas-Castillo, J., Moriones, E., Zerbini, F.M., Rivera-Bustamante, R., Malathi, V.G., Briddon, R., Varsani, A., 2013. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family Geminiviridae). Arch. Virol. 158, 1411–1424.

Mullineaux, P.M., Guerineau, F., Accotto, G.-P., 1990. Processing of complementary sense RNAs of Digitaria streak virus in its host and in transgenic tobacco. Nucleic Acids Res. 18, 7259–7265.

Mumtaz, H., Kumari, S., Mansoor, S., Martin, D., Briddon, R., 2011. Analysis of the sequence of a dicot-infecting mastrevirus (family *Geminiviridae*) originating from Syria. Virus Genes 42, 422–428.

Nahid, N., Amin, I., Mansoor, S., Rybicki, E., van der Walt, E., Briddon, R., 2008. Two dicot-infecting mastreviruses (family Geminiviridae) occur in Pakistan. Arch. Virol. 153, 1441–1451.

Owor, B.E., Shepherd, D.N., Taylor, N.J., Edema, R., Monjane, A.L., Thomson, J.A., Martin, D.P., Varsani, A., 2007. Successful application of FTA((R)) classic card technology and use of bacteriophage phi 29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. J. Virol. Methods 140, 100–105.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new gemini-viruses by frequent recombination. Virology 265, 218–225.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rocha, C.S., Castillo-Urquiza, G.P., Lima, A.T., Silva, F.N., Xavier, C.A., Hora-Junior, B.T., Beserra-Junior, J.E., Malta, A.W., Martin, D.P., Varsani, A., Alfenas-Zerbini, P., Mizubuti, E.S., Zerbini, F.M., 2013. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. J. Virol.

Rosario, K., Padilla-Rodriguez, M., Kraberger, S., Stainton, D., Martin, D.P., Breitbart, M., Varsani, A., 2013. Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Epiprocta) from Puerto Rico. Virus Res. 171, 231–237.

Schalk, H.J., Matzeit, V., Schiller, B., Schell, J., Gronenborn, B., 1989. Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication. EMBO J. 8, 359–364.

Schwinghamer, M., Thomas, J., Schilg, M., Parry, J., Dann, E., Moore, K., Kumari, S., 2010. Mastreviruses in chickpea (*Cicer arietinum*) and other dicotyledonous crops and weeds in Queensland and northern New South Wales, Australia. Australas. Plant Pathol. 39, 551–561.

Shepherd, D.N., Martin, D.P., Lefeuvre, P., Monjane, A.L., Owor, B.E., Rybicki, E.P., Varsani, A., 2008. A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. J. Virol. Methods 149, 97–102.

Shepherd, D.N., Martin, D.P., Van Der Walt, E., Dent, K., Varsani, A., Rybicki, E.P., 2010. Maize streak virus: an old and complex 'emerging' pathogen. Mol. Plant Pathol. 11, 1–12.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2713–2739.

Tanno, K.I., Willcox, G., 2006. The origins of cultivation of *Cicer arietinum* L. and *Vicia faba* L: early finds from Tell el-Kerkh, north-west Syria, late 10th millennium B.P. Vegetation History Archaeobot. 15, 197–204.

Thomas, J., Parry, J., Schwinghamer, M., Dann, E., 2010. Two novel mastreviruses from chickpea (*Cicer arietinum*) in Australia. Arch. Virol. 155, 1777–1788.

Varsani, A., Monjane, A.L., Donaldson, L., Oluwafemi, S., Zinga, I., Komba, E.K., Plakoutene, D., Mandakombo, N., Mboukoulida, J., Semballa, S., Briddon, R.W., Markham, P.G., Lett, J.M., Lefeuvre, P., Rybicki, E.P., Martin, D.P., 2009a. Comparative analysis of Panicum streak virus and Maize streak virus diversity, recombination patterns and phylogeography. Virol. J. 6, e194.

Varsani, A., Oluwafemi, S., Windram, O., Shepherd, D., Monjane, A., Owor, B., Rybicki, E., Lefeuvre, P., Martin, D., 2008a. Panicum streak virus diversity is similar to that observed for maize streak virus. Arch. Virol. 153, 601–604.

Varsani, A., Shepherd, D.N., Dent, K., Monjane, A.L., Rybicki, E.P., Martin, D.P., 2009b. A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. Virol. J. 6, e36.

Varsani, A., Shepherd, D.N., Monjane, A.L., Owor, B.E., Erdmann, J.B., Rybicki, E.P., Peterschmitt, M., Briddon, R.W., Markham, P.G., Oluwafemi, S., Windram, O.P., Lefeuvre, P., Lett, J.M., Martin, D.P., 2008b. Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. J. General Virol. 89, 2063–2074.

Wright, E.A., Heckel, T., Groenendijk, J., Davies, J.W., Boulton, M.I., 1997. Splicing features in maize streak virus virion- and complementary-sense gene expression. Plant J. 12, 1285–1297.