Infection, Genetics and Evolution 29 (2015) 203–215

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Molecular diversity of *Chickpea chlorotic dwarf virus* in Sudan: High rates of intra-species recombination – a driving force in the emergence of new strains

Simona Kraberger [a], Safaa G. Kumari [b,*], Abdelmagid A. Hamed [c], Bruno Gronenborn [d], John E. Thomas [e], Murray Sharman [f], Gordon W. Harkins [g], Brejnev M. Muhire [h], Darren P. Martin [h], Arvind Varsani [a,i,j,*]

[a] School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Christchurch 8140, New Zealand
[b] Virology Laboratory, International Centre for Agricultural Research in the Dry Areas (ICARDA), Tunis, Tunisia
[c] Plant Pathology Research Program, Agricultural Research Corporation, P.O. Box 126, Wadmedani, Sudan
[d] Institut des Sciences du Végétal, CNRS, 91198 Gif sur Yvette, France
[e] Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Ecosciences Precinct, GPO Box 267, Brisbane, QLD 4001, Australia
[f] Department of Agriculture, Fisheries and Forestry, Ecoscience Precinct, GPO Box 267, Brisbane, QLD 4001, Australia
[g] South African National Bioinformatics Institute, SA Medical Research Unit for Bioinformatics Capacity Development, University of the Western Cape, Private Bag X17, Bellville, Cape Town 7535, South Africa
[h] Computational Biology Group, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
[i] Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, 7700, South Africa
[j] Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 21 October 2014
Received in revised form 17 November 2014
Accepted 21 November 2014
Available online 29 November 2014

Keywords:
Mastrevirus
Geminivirus
Chickpea chlorotic dwarf virus
Recombination
Sudan
Chickpea

ABSTRACT

In Sudan *Chickpea chlorotic dwarf virus* (CpCDV, genus *Mastrevirus*, family *Geminiviridae*) is an important pathogen of pulses that are grown both for local consumption, and for export. Although a few studies have characterised CpCDV genomes from countries in the Middle East, Africa and the Indian subcontinent, little is known about CpCDV diversity in any of the major chickpea production areas in these regions. Here we analyse the diversity of 146 CpCDV isolates characterised from pulses collected across the chickpea growing regions of Sudan. Although we find that seven of the twelve known CpCDV strains are present within the country, strain CpCDV-H alone accounted for ~73% of the infections analysed. Additionally we identified four new strains (CpCDV-M, -N, -O and -P) and show that recombination has played a significant role in the diversification of CpCDV, at least in this region. Accounting for observed recombination events, we use the large amounts of data generated here to compare patterns of natural selection within protein coding regions of CpCDV and other dicot-infecting mastrevirus species.

1. Introduction

The Middle East, North Africa and the Indian subcontinent are all major producers of chickpeas, lentils, faba beans and various other pulses. In the Sudan pulses are both an important food source for the Sudanese people and a cash crop. They are grown in the fertile regions along the banks of the Nile which runs through the middle of the country, from South Sudan towards Egypt in the north. A serious constraint of pulse production in general, and of chickpea farming in particular, is the pathogen *Chickpea chlorotic dwarf virus* (CpCDV, genus *Mastrevirus*, family *Geminiviridae*). CpCDV causes a variety of symptoms in chickpea including stunting, foliar yellowing or reddening and reduced seed production. Other important chickpea-infecting viruses are *Faba bean necrotic yellows virus* (genus *Nanovirus*, family *Nanoviridae*) and members of the genus *Polerovirus* (in the family *Luteoviridae*; Abraham et al., 2006; Kumari et al., 2008; Makkouk et al., 2003).

Globally there are seven known species of dicotyledonous plant-infecting mastreviruses (referred to here as dicot-infecting), five of which have only been documented in Australia: *Chickpea*

* Corresponding authors at: School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. Tel.: +64 3 366 7001x4667; fax: +64 3 364 2590 (A. Varsani).
E-mail addresses: s.kumari@cgiar.org (S.G. Kumari), arvind.varsani@canterbury.ac.nz (A. Varsani).

http://dx.doi.org/10.1016/j.meegid.2014.11.024
1567-1348/

chlorosis virus; CpCV (Hadfield et al., 2012; Kraberger et al., 2013; Thomas et al., 2010), *Chickpea chlorosis Australia virus*; CpCAV (Hadfield et al., 2012), *Chickpea redleaf virus*; CpRLV (Thomas et al., 2010), *Chickpea yellows virus*; CpYV (Hadfield et al., 2012) and *Tobacco yellow dwarf virus*; TYDV (Hadfield et al., 2012; Morris et al., 1992). The two species found outside of Australia are CpCDV (Ali et al., 2004; Horn et al., 1993; Kraberger et al., 2013; Kumari et al., 2004; Makkouk et al., 1995; Manzoor et al., 2014; Mumtaz et al., 2011; Nahid et al., 2008) and Chickpea yellow dwarf virus (CpYDV) (Kraberger et al., submitted for publication). Whereas CpCDV has been found in the Middle East (including Turkey), Africa and the Indian subcontinent, CpYDV has so-far only been found in Pakistan. With the exception of TYDV, all of these dicot-infecting mastrevirus species have predominately been found infecting chickpeas, although little is known about their potential host range.

CpCDV is transmitted by the leafhopper species *Orosius orientalis* (Matsumura; Horn et al., 1993) and *Orosius albicinctus* (Distant; Cicadellidae: Hemiptera; Akhtar et al., 2011; Kumari et al., 2004). Natural hosts identified in the field include chickpea (*Cicer arietinum*; Kraberger et al., 2013; Kumari et al., 2004; Mumtaz et al., 2011; Nahid et al., 2008), lentil (*Lens culinaris* Medik; Kraberger et al., 2013; Makkouk et al., 2002), faba bean (*Vicia faba*; Kraberger et al., 2013), field pea (*Pisum sativum*; Kraberger et al., 2013), French bean (*Phaseolus vulgaris* L.; Ali et al., 2004; Liu et al., 1997), sugar beet (*Beta vulgaris* L.; Farzadfar et al., 2008), *Sesbania bispinosa* (Jacq.; Nahid et al., 2008), pepper (*Capsicum annum* L.; Akhtar et al., 2013) and cotton (*Gossypium* sp.; Manzoor et al., 2014).

Recent studies have extended our current knowledge of CpCDV diversity (Kraberger et al., 2013; Manzoor et al., 2014; Mumtaz et al., 2011) and there are currently twelve identified strains of CpCDV (A–L). Despite this there is little information on the prevalence and diversity of CpCDV within the pulse growing regions of individual countries. Field surveys in Sudan between 1996 and 2000 used serological analysis (tissue blot immunoassays) to reveal a CpCDV incidence of 72% in chickpea crops, and therefore identified this virus as the most important threat to chickpea production in Sudan (Hamed and Makkouk, 2002). The antibodies used for serological testing of CpCDV are known to cross react with other dicot-infecting mastrevirus species and at the time no nucleotide sequence data was obtained for further analysis of these samples.

Along with 16 CpCDV genomes from pulse samples collected between 1997 and 2008, we obtained full-length CpCDV genomes from an additional 129 samples collected in Sudan between 2012 and 2014 and two available in GenBank in order to analyse the molecular diversity for the first time of a large representative CpCDV sample set at high resolution within the pulse-growing areas of an entire country. Besides discovering four new CpCDV strains, we found evidence that extensive inter- and intra-strain recombination has made a major contribution to the diversification of this species. Finally, we capitalised on the large amounts of data generated here to compare patterns of natural selection found in CpCDV to those found in other known dicot-infecting mastreviruses.

2. Material and methods

2.1. Sample collection, DNA isolation and full genome recovery

In the growing seasons 2012–2014 leaf material from pulse plants displaying foliar yellowing, mosaic/mottling patterns and/or stunting was collected from 275 individual plants located in the major pulse-growing regions of Sudan. Additionally, pulse plant samples collected in Sudan in 1997, 2006 and 2008 were obtained from Institut des Sciences du Végétal, CNRS in France. A total of 312 samples from Sudan were screened from Gezira state ($n = 166$), the River Nile state ($n = 141$) and the Northern state ($n = 5$). Also, opportunistic sampling of similarly symptomatic plants was undertaken in Morocco ($n = 18$) in 2013.

Total genomic DNA from dried plant material was extracted using the GF-1 nucleic acid extraction kit (Vivantis Technologies, Malaysia), according to manufacturer's specifications. Circular DNA was enriched from DNA extracts using rolling circle amplification with the Illustra TempliPhi Amplification Kit (GE Healthcare, USA) as previously described (Owor et al., 2007; Shepherd et al., 2008). Full viral genomes were amplified from 0.5 µl of enriched viral DNA using polymerase chain reaction (PCR). The PCR reaction comprised Kapa HiFi HotStart DNA polymerase (Kapa biosystems, USA) together with previously described degenerate back-to-back primers (dicot forward 5′-GAN TTG GTC CGC AGT GTA GA-3′, dicot reverse 5′-GTA CCG GWA AGA CMW CYT GG-3′) (Hadfield et al., 2012). This primer pair binds in a region which is highly conserved among most mastrevirus species. The thermal cycling protocol used was as follows: 94 °C for 3 min, 25 cycles [98 °C (3 min), 52 °C (30 s), 72 °C (2 min and 45 s)], 72 °C for 3 min. PCR products were purified using the Quick-spin PCR Product Purification Kit (iNtRON Biotechnology, Korea) and ligated into pJET1.2 vector using the CloneJET™ PCR Cloning Kit (Fermentas, USA). Resulting recombinant plasmids containing viral genomes were sequenced at Macrogen Inc. (Korea) using primer walking.

2.2. Sequence assembly, management and pairwise similarity calculation

Complete CpCDV genomes were assembled using DNA Baser Sequence Assembler V4 (Heracle BioSoft, Romania) and probable genes identified using DNAMAN V7 (Lynnon Biosoft, Canada). The complete CpCDV genome sequences determined in this study together with 115 publically available dicot-infecting mastrevirus sequences, and a single wheat dwarf virus (WDV) genome (accession number AM040732) in order to root the phylogenetic tree, (downloaded on 01 July 2014) were aligned using MUSCLE (Edgar, 2004). The resulting alignment was manually edited by eye with MEGA5.2 (Tamura et al., 2011). Pairwise sequence similarities between CpCDV genome sequences were determined using SDT V1.0, calculated as 1-p-distance, with pairwise deletion of gaps (Muhire et al., 2014).

2.3. Construction of phylogenetic trees

A full genome maximum likelihood phylogenetic tree was constructed using PHYML version 3 (Guindon et al., 2010) using the best fit substitution model (TN93+G+I510) identified using jModelTest (Darriba et al., 2012) and rooted with WDV. Branches with approximate likelihood ratio test (aLRT) support <80% were collapsed using Mesquite version 2.75 (Maddison and Maddison, 2011).

2.4. Recombination analysis

Two datasets, one containing all available dicot-infecting mastrevirus sequences and the other containing only the available CpCDV sequences were compiled and analysed for evidence of recombination using RDP4 (Martin et al., 2010) with the following methods: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), Bootscan (Martin et al., 2005), Maxchi (Smith, 1992), Chimera (Posada and Crandall, 2001), Siscan (Gibbs et al., 2000), and 3Seq (Boni et al., 2007). Recombination events were only accepted as credible when they were detectable by three or more of these methods with associated p-values $<10^{-3}$ and they had strong phylogenetic support (i.e. clustering of the identified recombinant(s) in different parts of phylogenetic trees constructed from regions of the alignment corresponding to the fragments of

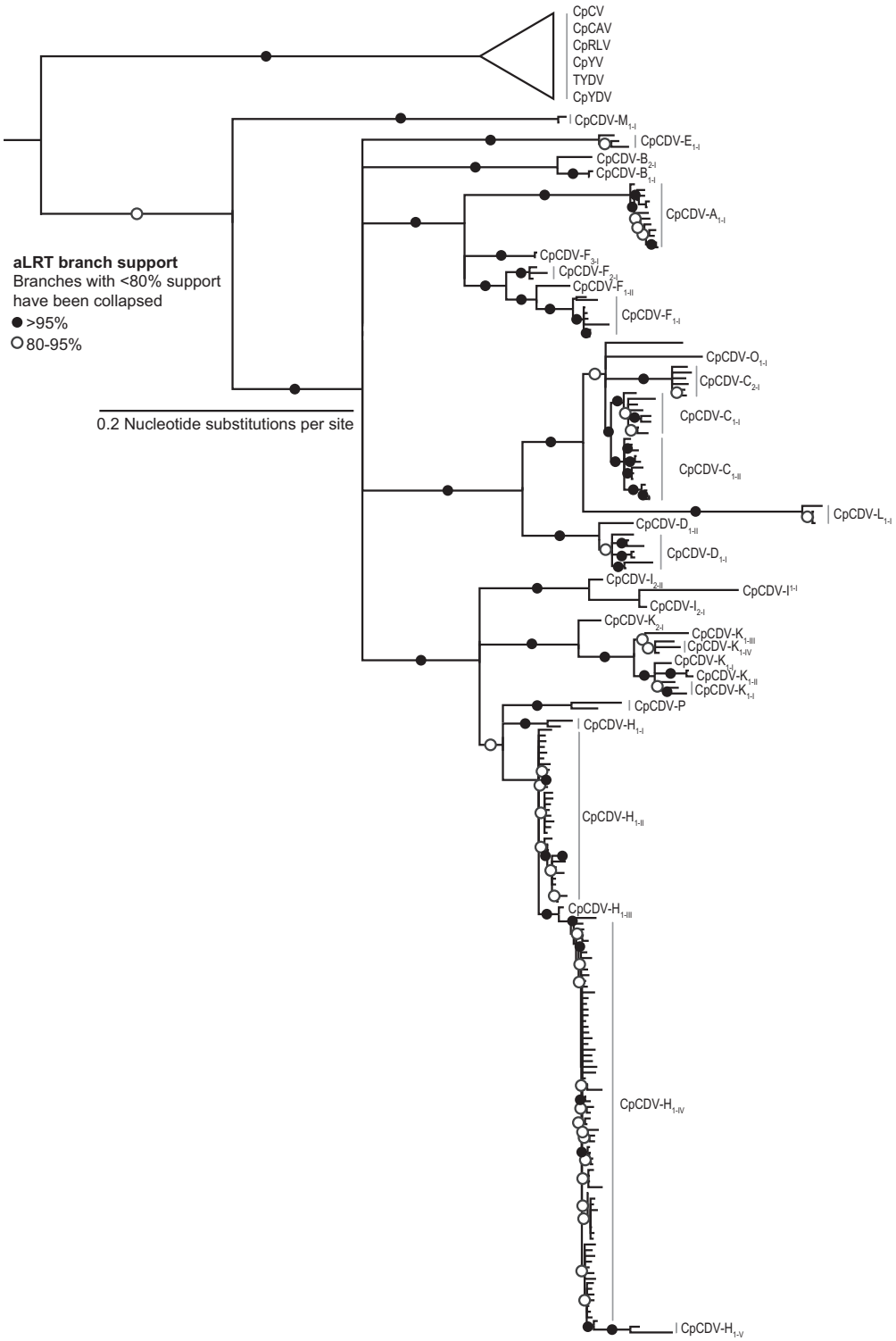

Fig. 1. Maximum likelihood phylogenetic tree indicating the relationships of the known dicot-infecting mastrevirus species. Full genome sequences of CpCDV from this study together with those available in GenBank, as well as the five other species of dicot-infecting mastreviruses from Australia and CpYDV from Pakistan are included. Australian dicot-infecting mastrevirus species and CpYDV clade has been collapsed and represented by a triangle. Branch support is indicated by the open and closed circles in the key. Branches with less than 80% aLRT support have been collapsed.

the recombinant identified as having been derived from each of its parents).

2.5. Selection analysis

The full genome dicot-infecting mastrevirus dataset was divided into movement protein (MP), capsid protein (CP) and replication-associated protein (Rep) coding regions and realigned using codon information with MUSCLE. From these alignments we extracted coding region (MP, CP, and Rep) datasets for CpCDV, CpCV, CpCAV and TYDV. There were too few sequences available for CpYDV, CpRLV and CpYV for us to perform selection analyses on these species. Accounting for recombination breakpoints identified with the GARD method (Kosakovsky Pond et al., 2006) these

12 datasets were separately analysed for evidence of selection acting on individual codon sites using the MEME (Murrell et al., 2012) and FUBAR (Murrell et al., 2013) methods implemented in the HyPhy package via the online DATAMONKEY server (http://www.datamonkey.org/) (Delport et al., 2010). The FUBAR method was used to identify individual codon sites evolving under either diversifying or negative selection throughout the entire evolutionary history of the sequences being analysed and the MEME method was used to identify individual codons evolving under episodic diversifying selection within individual sub-lineages within the analysed datasets.

3. Results and discussion

3.1. Analysis of CpCDV diversity

In this study a total of 145 full CpCDV genomes were recovered and sequenced from infected pulse plants collected in the Gezira state ($n = 104$), River Nile state ($n = 38$) and Northern state ($n = 3$) of Sudan and from one sample obtained from Morocco (Sup. Table 1). To our knowledge this is the first report of CpCDV in Morocco.

Prior to this study twelve CpCDV strains had been identified from Syria, Turkey, Iran, South Africa, Pakistan, India, Sudan, Yemen and Eritrea. Based on the mastrevirus classification guidelines outlined by Muhire et al. (2013) (i.e. isolate sequences with 78–94% pairwise identity are unique strains and isolates with >94% pairwise identity belong to the same strain) we were able to assign 140 of the CpCDV isolates from this study to seven of the 12 previously described CpCDV strains C ($n = 18$), D ($n = 3$), E ($n = 1$), F ($n = 3$), H ($n = 107$), I ($n = 1$) and K ($n = 7$). The remaining six isolates most likely represent four new strains which we have tentatively named CpCDV-M ($n = 2$), -N ($n = 1$), -O ($n = 1$) and -P ($n = 2$). For the purpose of this study, we further classified CpCDV strains into genotype groupings based on a threshold of 94–97% pairwise sequence identity (i.e., isolates with >97% pairwise sequence identity represent the same genotype, whereas those with >94% but <97% pairwise sequence identity represent different genotypes). The monophyly of these genotypes were also phylogenetically supported (Fig. 1 and Sup. Table 2). Additionally, we further classified these isolates into variant groupings (i.e., groups of isolates with pairwise sequence identities of $\geqslant 99\%$ were classified as belonging to the same group of variants). Different variant groupings were denoted with a numerical subscript after the strain letter designator: e.g., CpCDV-H variant group one and two were denoted CpCDV-H_1 and CpCDV-H_2.

We noted that strains CpCDV-J and -I include genomes that share more than 97% sequence identity and we therefore merged these strains into strain CpCDV-I. Similarly, two CpCDV-G isolates were found to share >95% sequence identity with the CpCDV-F isolates and we therefore merged these isolates into the strain CpCDV-F. We have followed the recommendation laid out in Muhire et al. (2013). Please note that strains are not officially recognised by the International Committee for the Taxonomy of viruses.

3.2. The CpCDV population in Sudan is highly diverse

The CpCDV genomes determined here were primarily obtained from the major chickpea producing areas of Sudan. These regions were located along the Nile in the proximity of Wad Medani (Gezira State), Shendi and Berber (Nile State) and Selaim (Northern State) (Fig. 2). Prior to this study only two publically available full genomes of CpCDV (AM933134 and AM933135) had been determined (isolated from chickpeas) from Sudan. Both of these genomes belonged to strain E and both were sampled in 1997 from Abu Haraz, near Wad Medani (Gezira State). Here we have recovered an additional Sudanese CpCDV-E genome from a

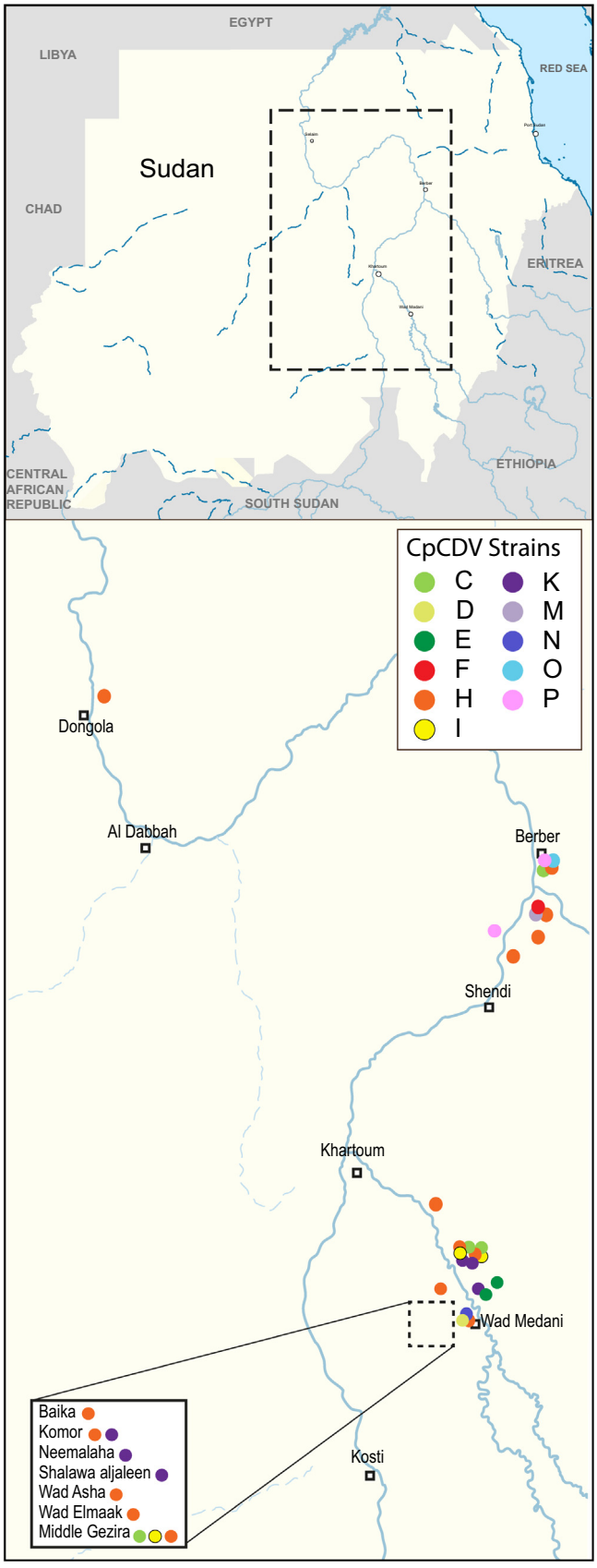

Fig. 2. Sampling sites and the distribution of CpCDV strains in Sudan.

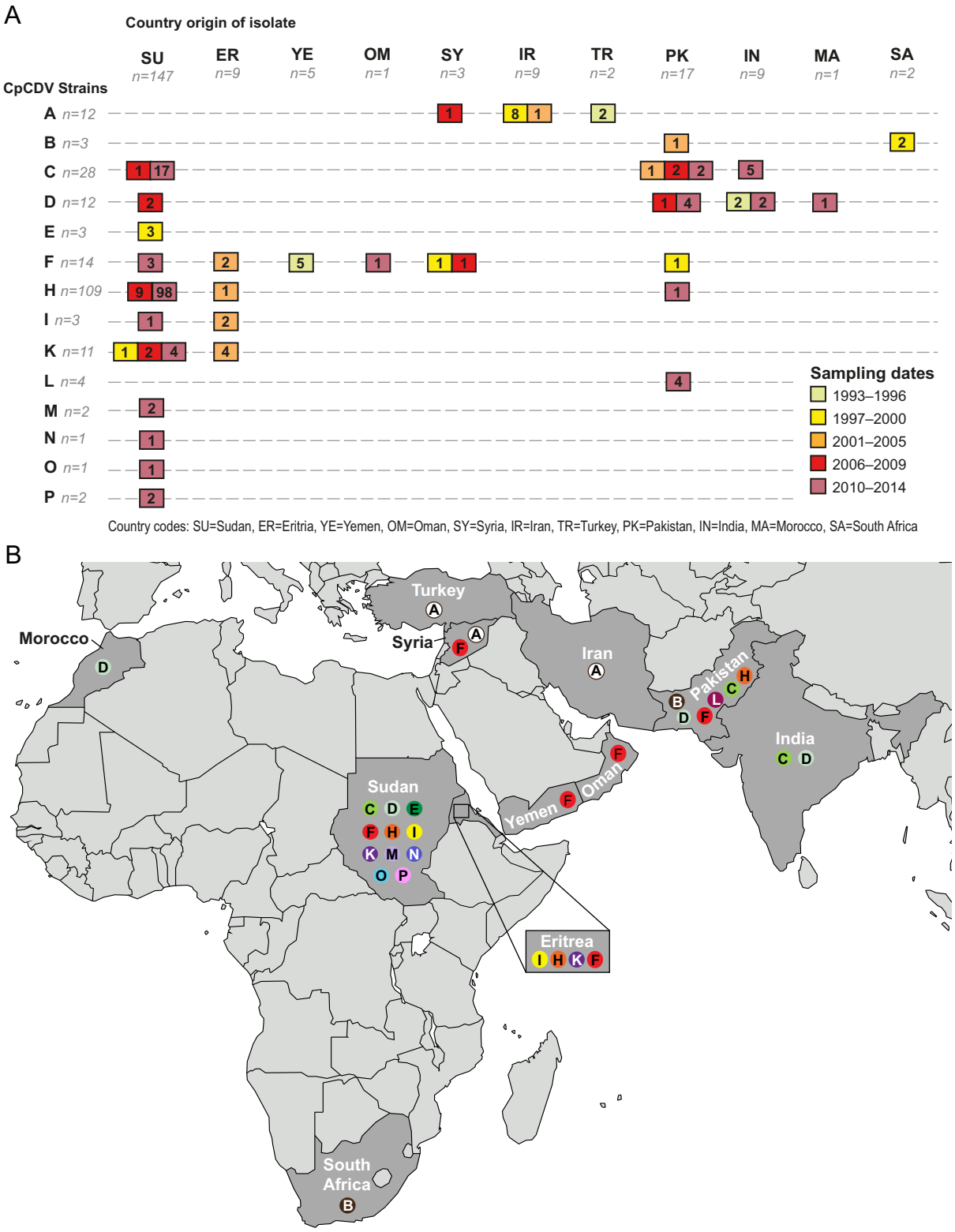

Fig. 3. (A) Summary of country of origin, strain, sampling year and total numbers of CpCDV genomes identified including those recovered in this study and all other genomes available in GenBank. Sampling dates have been clustered into four year intervals which are indicated by the gradient of colours shown in the key. Numbers of genome sequences determined for each strain and country are indicated within each corresponding box. (B) Partial World map highlighting those countries from which full-length CpCDV genomes have been sequenced (Dark grey). Coloured circles indicate strains that have been recovered from each country. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

plant sample collected in 1997 El Rayafa (Fig. 3A and Sup. Table 1). It is interesting to note that, despite the scale of our sampling, no other CpCDV-E isolates were recovered (Fig. 2). The absence of CpCDV-E variants in any samples collected here and elsewhere since 1997 suggests that this strain is a rare variant in the CpCDV population of Sudan.

Other CpCDV strains are apparently more persistent than CpCDV-E. CpCDV-K has so far only ever been found in the region

Table 1
Summary of intra-species recombination events. Major and Minor parent labels indicate the inferred parent(s) respectively donating the larger and smaller fraction of the recombinant's genome. The method with the most significant *p*-value is indicated in bold and the associated *p*-value is shown.

Event	Recombinant region	Potential major parent	Potential minor parent	Detection method	*p*-Value
Intra-species recombination					
1	234–1096	All CpCDV-C, CpCDV-$D_{1\text{-II}}$, CpCDV-$D_{1\text{-III}}$, CpCDV-$D_{1\text{-I}}$ (KM377671, KM229788, KC172664), CpCDV-N	All CpCDV-K, All CpCDV-P	RGBCS**T**	9.42×10^{-47}
2	1689–138	All CpCDV-C, CpCDV-O	All CpCDV-H (except KM229850, KM229885)	RGMCS**T**	2.70×10^{-57}
3	708–1668	All CpCDV-$K_{1\text{-IV}}$, All CpCDV-$K_{1\text{-I}}$, CpCDV-$K_{1\text{-III}}$	All CpCDV-H (except KM229850), CpCDV-$I_{2\text{-II}}$	RGBMCS**T**	5.60×10^{-62}
4	1700–310	All CpCDV-$K_{1\text{-I}}$, CpCDV-$K_{1\text{-III}}$, All CpCDV-$K_{1\text{-IV}}$, CpCDV-$K_{2\text{-I}}$	All CpCDV-H, All CpCDV-I_2	RGBMCS**T**	9.01×10^{-44}
5	1942–2543	All CpCDV-C_1, CpCDV-$D_{1\text{-I}}$ (KM377674, KM377671, FR687960, KF176553, KC172664, KC172665), CpCDV-$D_{1\text{-II}}$, CpCDV-$D_{1\text{-III}}$, CpCDV-O	All CpCDV-K, CpCDV-$I_{1\text{-I}}$	RGBMCS**T**	6.18×10^{-26}
6	119–1085	All CpCDV-F_3	CpCDV-$F_{1\text{-I}}$ (KC172667, KC172669, KC172670, KC172671)	RGBMCS**T**	6.86×10^{-38}
7	404–564	CpCDV-I_1, CpCDV-$I_{2\text{-I}}$	All CpCDV-H	RG**B**MCST	7.88×10^{-15}
8	1133–1583	CpCDV-I_1, CpCDV-$I_{2\text{-I}}$	All CpCDV-H, CpCDV-K_2	RGBMCS**T**	3.72×10^{-27}
9	1372–2485	All CpCDV-$H_{1\text{-III}}$, All CpCDV-$H_{1\text{-IV}}$	All CpCDV-I_2	RGBMCS**T**	1.93×10^{-15}
10	1982–187	CpCDV-$C_{1\text{-I}}$ (AM900416), CpCDV-$C_{1\text{-II}}$ (KM229774, KM229776, KM229778, KM229779), CpCDV-$D_{1\text{-II}}$ (KM229787)CpCDV-$H_{1\text{-II}}$ (KM229893)CpCDV-$H_{1\text{-III}}$, CpCDV-$H_{1\text{-VI}}$ (KM377669, KM229803, KM229806, KM229810, KM229819, KM229820, KM229824, KM229825, KM229827, KM229830, KM229836, KM229838, KM229840, KM229842,KM229844, KM229845, KM229847, KM229849, KM229851, KM229852, KM229860, KM229866, KM229869, KM229871, KM229872, KM229873,KM229878, KM229884, KM229888, KM229889, KM229890, KM229891, KM229895, KM229898, KM229793, KM229797)	All CpCDV-K	**R**GBMCST	1.47×10^{-12}
11	1709–2483	CpCDV-$H_{1\text{-II}}$ (KM229877), CpCDV-$H_{1\text{-III}}$, All CpCDV-$H_{1\text{-IV}}$	CpCDV-$I_{2\text{-I}}$, CpCDV-$I_{2\text{-II}}$	**R**GBMCST	4.49×10^{-13}
12	2493–698	CpCDV-$D_{1\text{-I}}$ (KC172664)	All CpCDV-C	RGBMCS**T**	2.02×10^{-21}
13	2029–2479	All CpCDV-P, All CpCDV-$H_{1\text{-II}}$ (except KM229897)	Ancestral CpCDV-I-like	**R**GBMCST	9.47×10^{-12}
14	2548–147	All CpCDV-H (except KM229850, KM229885) CpCDV-$P_{1\text{-I}}$, CpCDV-$P_{1\text{-II}}$	All CpCDV-K	**RG**BMCST	1.73×10^{-09}
15	1602–1713	All CpCDV-$K_{1\text{-I}}$, CpCDV-$K_{1\text{-III}}$, CpCDV-$K_{1\text{-IV}}$ (KM229904)	CpCDV-$H_{1\text{-I}}$ (KM229843), All CpCDV-$H_{1\text{-II}}$ (except KM229770), CpCDV-$H_{1\text{-III}}$, All CpCDV-$H_{1\text{-IV}}$ (except KM229895), All CpCDV-$H_{1\text{-V}}$	**G**BLT	1.44×10^{-09}
16	305*–425*	CpCDV-$K_{1\text{-I}}$ (KC172682)	CpCDV-$H_{1\text{-V}}$ (KM229850), All CpCDV-$H_{1\text{-II}}$ (except KM229815, KM229816, KM229817), All CpCDV-$H_{1\text{-IV}}$ (except KM229889, KM229890, KM229884, KM229848, KM229838, KM229810, KM377669)	**R**BT	1.59×10^{-03}
17	1123–1371	All CpCDV-K	All CpCDV-E	RGBMCS**T**	1.43×10^{-17}
18	1258–1547	All CpCDV-$D_{1\text{-I}}$, CpCDV-$D_{1\text{-II}}$	All CpCDV-L	**R**GBMCST	4.14×10^{-08}
19	433–524	All CpCDV-$K_{1\text{-I}}$, CpCDV-$K_{1\text{-III}}$, All CpCDV-$K_{1\text{-IV}}$, CpCDV-K_2, CpCDV-O, All CpCDV-P	All CpCDV-E	**R**GBM	2.52×10^{-08}
20	1949–2548	Ancestral CpCDV-A-like	All CpCDV-$F_{3\text{-I}}$	R**G**BMCST	9.66×10^{-07}
21	1579–1850	All CpCDV-$H_{1\text{-II}}$ (except KM229816, KM229876, KM229877), CpCDV-$H_{1\text{-III}}$, All CpCDV-$H_{1\text{-IV}}$ (except KM229797, KM377669, KM229812, KM229813, KM229824, KM229834, KM229836, KM229837, KM229844, KM229845, KM229849, KM229851, KM229852, KM229861, KM229867, KM229872, KM229873, KM229878, KM229879)	All CpCDV-M	**R**BMC	1.05×10^{-05}
22	1873–6*	All CpCDV-E	All CpCDV-F_3, All CpCDV-A_1 (except KC172654)	RGBMC**S**	6.62×10^{-08}
23	1957–2406	All CpCDV-$F_{2\text{-I}}$, All CpCDV-F_3, CpCDV-$F_{1\text{-II}}$	Ancestral CpCDV-P-like, CpCDV-H-like	GMC**S**	9.07×10^{-10}
24	330*–867	Ancestral CpCDV-I-like	All CpCDV-$H_{1\text{-II}}$, All CpCDV-$H_{1\text{-IV}}$	RBMC**S**T	1.01×10^{-07}
25	2549*–118*	CpCDV-$F_{1\text{-I}}$ (KC172667)	Ancestral CpCDV-$F_{2\text{-I}}$-like	RB**T**	7.00×10^{-05}
26	2510–23	All CpCDV-$C_{1\text{-I}}$, CpCDV-$C_{1\text{-II}}$ (except KM229771, KM229773, KM229772, KM229775), All CpCDV-$D_{1\text{-I}}$ (except KM229786, KM377672), CpCDV-$D_{1\text{-II}}$, CpCDV-O	All CpCDV-M	R**G**B	1.41×10^{-09}
27	2514*–2562	CpCDV-$H_{1\text{-IV}}$ (KM229886, KM229857, KM229850, KM229847, KM229811, KM229797, KM229800), All CpCDV-$H_{1\text{-II}}$ (except KM229804, KM229807, KM229814, KM229815, KM229817, KM229822, KM229862, KM229864, KM229870, KM229877, KM229893, KM229894, KM229897, KM229859) CpCDV-$H_{1\text{-III}}$, CpCDV-$H_{1\text{-V}}$ (KM229885)	CpCDV-$D_{1\text{-I}}$ (KC172664, FR687960, KC172665)	RGB**L**	1.68×10^{-28}
28	206–536	All CpCDV-E	Ancestral CpCDV-C_1-like and CpCDV-D_1-like	MC**S**	5.40×10^{-04}
29	1708*–2508*	All CpCDV-$H_{1\text{-II}}$, CpCDV-$H_{1\text{-III}}$, All CpCDV-$H_{1\text{-IV}}$ (except KM229879), CpCDV-$P_{1\text{-I}}$	All CpCDV-C_1, All CpCDV-$C_{2\text{-I}}$ (except KM229770), CpCDV-O	RBMCS**T**	1.35×10^{-40}

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

surrounding Wad Medani, it was sampled there in 1997, 2008 and 2013. Similarly CpCDV-C and CpCDV-H isolates were consistently sampled in Sudan between 2006 and 2014. CpCDV-C has only ever been found in the Berber region, whereas CpCDV-H was found at every sampling site and accounted for 73% of all CpCDV isolates collected in Sudan since 2006. Given the prevalence of CpCDV-H in Sudan, it is not surprising that this strain was also detected in neighbouring Eritrea in 2005 (Kraberger et al., 2013).

Due to the low numbers of CpCDV genomic sequences that have been sampled outside of Sudan it is not currently possible to accurately infer the world-wide diversity and prevalence of the various CpCDV strains. Nonetheless, based on the available data, CpCDV-F is apparently the most widely distributed CpCDV strain in that it has been detected in six of the eleven countries where CpCDV genomes have been recovered (Fig. 3B).

Other strains appear to be confined to specific regions. For example, CpCDV-A has only been identified in the region encompassing Iran, Syria and Turkey. It is, however, important to reiterate that there are only five countries (Sudan, India, Pakistan, Eritrea and Iran) from which nine or more CpCDV genomes have been

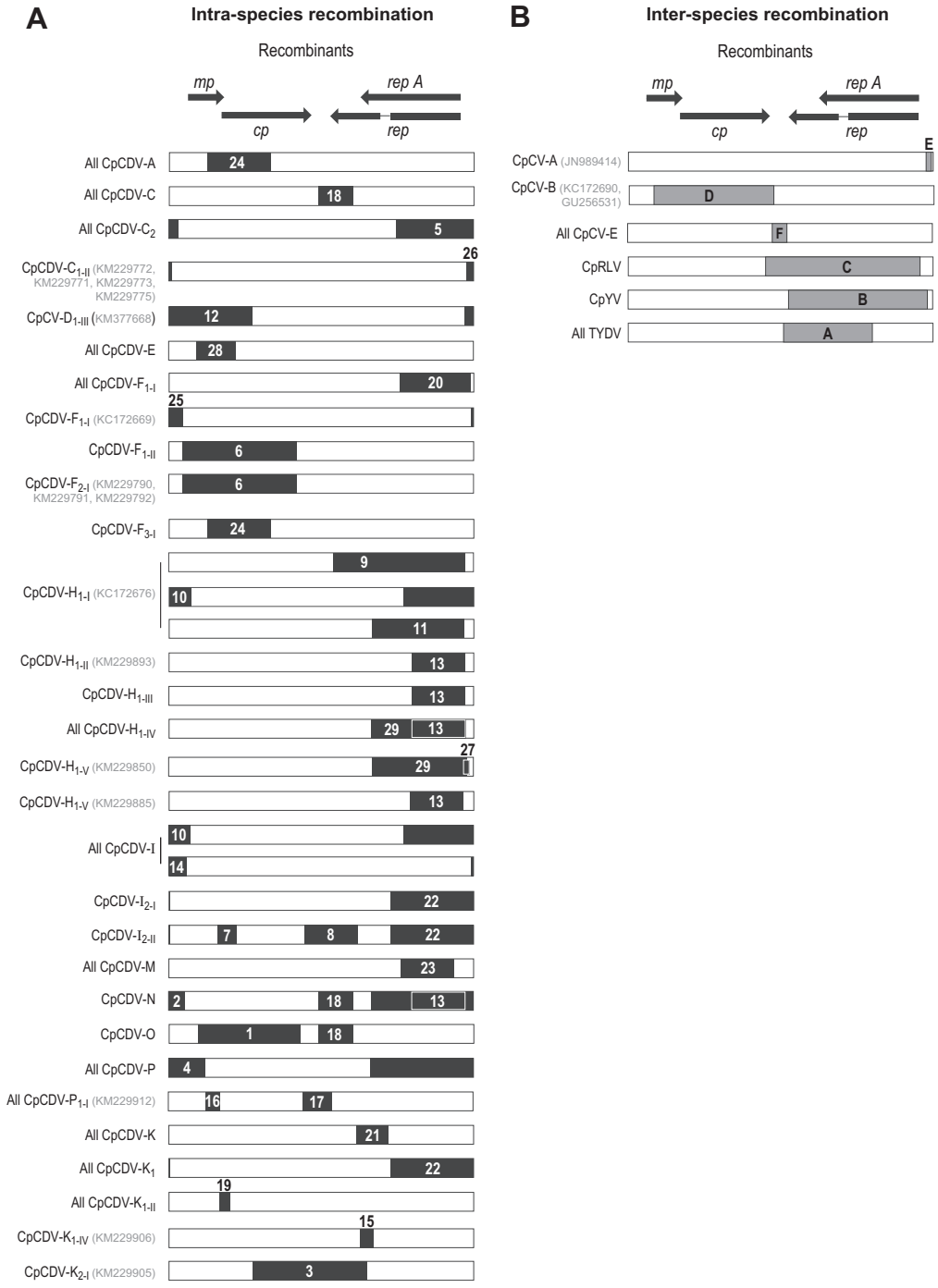

Fig. 4. Recombination amongst dicot-infecting mastreviruses involving CpCDV. The positions of the *mp* (movement protein), *cp* (capsid protein), *repA* (replication-associated A protein) and *rep* (replication-associated protein) ORFs in relation to the recombination breakpoints are shown by arrows above recombinants. Genotype and variant information for each strain is shown in Sup. Table 2. (A) Intra-species recombination events: each event is represented in black and by a number which points to the corresponding entry in Table 1. (B) Inter-species recombination events: each event is represented in grey and by a letter which points to a corresponding entry in Table 2.

sampled and it is possible that extra sampling will reveal that the less commonly detected strains such as -B and -K have a greater geographical range than is presently apparent from the available data.

3.3. Evidence of extensive inter-strain recombination

As with other geminiviruses, recombination is apparently a major feature of mastrevirus evolution. We analysed the 146 genomes from this study together with other dicot-infecting mastrevirus genomes available in GenBank for evidence of recombination. Previously many inter-strain recombination events (Kraberger et al., 2013) and inter-species recombination events (Hadfield et al., 2012; Kraberger et al., 2013; Martin et al., 2011b) involving CpCDV have been identified. We therefore attempted to both characterise novel recombination events evident within the genome sequences determined here and refine the characterisations of previously identified recombination events involving CpCDV.

We detected evidence of two unique intra-strain, 29 unique inter-strain (intra-strain and inter-strain events are collectively referred to as intra-species events) and six unique inter-species recombination events, all of which involved CpCDV isolates as sequence acceptors. Of these events 19 intra-species events and one inter-species recombination event have not previously been identified. For four of the ten previously detected intra-species recombination events (Hadfield et al., 2012; Kraberger et al., 2013; Martin et al., 2011b) we were able to for the first time, identify both of the likely parental sequences to at least the strain level (events 8, 12, 18 and 21 in Table 1 and Fig. 4A).

Recombination appears to have played a particularly predominant role in the genesis of strains CpCDV-N, -O and -P (three of the novel strains identified here for the first time). The ancestral progenitor of each of these strains was likely derived from several recombination events involving nine parental sequences belonging to other CpCDV strains found in Sudan. One event inferred to have occurred in the common ancestor of CpCDV-N and -O involves a parental sequence that is most similar to CpCDV-L (event 18; Fig. 4A and Table 1). Although CpCDV-L has so far only been found in Pakistan (infecting cotton), it is entirely plausible that the CpCDV-L-like parent of the ancestral recombinant that yielded the O and N strains could have existed almost anywhere within (or even outside of) the presently known geographical range of CpCDV.

Of the 146 sequences classified here as belonging to CpCDV-H (the predominant CpCDV strain found in Sudan), 107 are recombinants with parental viruses likely belonging to CpCDV strains -I, -O, -C, -D and a currently unsampled strain (events 9, 10, 11, 13, 27 and 29 in Fig. 4A and Table 1). Interestingly all but one of these recombination events (the exception being event 27) involved a CpCDV-H acquiring a Rep coding region fragment from one of the other CpCDV strains that are found in Sudan.

In fact, many of the other detected inter-strain recombination events have also involved the transfer of *rep* gene fragments. This pattern of sequence exchange mirrors that seen in other geminiviruses (specifically those in the *Curtovirus*, *Mastrevirus* and *Begomovirus* genera). Specifically, the recombination patterns evident here and elsewhere indicate that for geminiviruses in general, recombination frequencies may be higher in genome regions encoding complementary sense genes than they are in the regions encoding virion sense genes. Alternatively, if basal recombination frequencies are similar across the genome, it would imply that selection might generally disfavour the survival of recombinants with breakpoints within the virion-sense genes more than it disfavours recombinants with breakpoints in the complementary-sense genes (Kraberger et al., 2012; Lefeuvre et al., 2009; Martin et al., 2011a,b; Varsani et al., 2008).

Four of the six inter-species events that we detected involved the transfer of large genome fragments ranging in size from 753

Table 2
Summary of inter-species recombination events. Major and Minor parent labels indicate the inferred parent(s) respectively donating the larger and smaller fraction of the recombinant's genome. The method with the most significant *p*-value is indicated in bold and the associated *p*-value is shown.

Event	Recombinant region	Potential major parent	Potential minor parent	Detection method	*p*-Value
Inter-species recombination					
A	1311-2064	All CpCAV	All CpCDV	**R**GBMCST	3.05×10^{-14}
B	1348-2517	All CpCDV	All CpCAV	RBM**C**ST	6.82×10^{-11}
C	1168-2482	All CpCDV-B, All CpCDV-E, CpCDV-$D_{1\text{-}I}$ (KM229787, KM377671, KM229788, KC172664, KC172665), All CpCDV-$C_{1\text{-}I}$, All CpCDV-$F_{2\text{-}I}$, All CpCDV-$F_{1\text{-}II}$, All CpCDV-$F_{1\text{-}I}$, All CpCDV-L, All CpCDV-M, All CpCDV-A (except KC172657, KC172658, KC172661), All CpCDV-$H_{1\text{-}III}$, CpCDV-$H_{1\text{-}IV}$ (KM229793, KM229794KM229796, KM229797, KM229799, KM377669, KM229802, KM229803, KM229806, KM229809-11, KM229813, KM229819, KM229820, KM229825, KM229830, KM229836, KM229839-42, KM229844, KM229845, KM229847, KM229848, KM229851, KM229854, KM229856, KM229857, KM229860, KM229861, KM229866, KM229868KM229869, KM229872, KM229878, KM229879, KM229880, KM229882, KM229884, KM229886, KM229889, KM229890, KM229891, KM229895, KM229896, KM229898)	All CpCAV (except KC172687, KC172688)	RBMC**S**	6.62×10^{-10}
D	209-1234	Ancestral CpRV	All CpCV-E, CpCV-A (GU256530)	RBMC**S**	4.17×10^{-24}
E	2512-2556	CpCV-A (JN989415, GU256530, JN989413, KC172684), CpCV-E (JN989431, JN989433)	All CpCDV-$B_{2\text{-}I}$, All CpCDV-$_{C1\text{-}I}$, All CpCDV-$C_{1\text{-}II}$, All CpCDV-D (except KC172665, KC172664), All CpCDV-$H_{1\text{-}IV}$, All CpCDV-$_{I1\text{-}I}$, All CpCDV-$K_{1\text{-}III}$, All CpCDV-$K_{1\text{-}IV}$, CpCDV-N, CpCDV-O, CpCDV-E (KM229901, AM933135), CpCDV-$B_{1\text{-}I}$ (AM849096), CpCDV-$C_{2\text{-}I}$ (KM229800), CpCDV-$H_{1\text{-}II}$ (KM229893)	RG**B**M	2.09×10^{-09}
F	1230-1335*	CpCV-F (KC172700)	Ancestral CpCDV-H-like	MS**T**	4.60×10^{-06}

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

to 1314 nt. (events A, B, C and D in Fig. 4B and Table 2). Overall the inter-species recombination events involve the transfer of on average 21% of the genome. This percentage is substantially larger than those inferred in previous analyses of both dicot- and monocot-infecting mastreviruses (Hadfield et al., 2012; Kraberger et al., 2013, 2012; Martin et al., 2011b; Monjane et al., 2011; Varsani et al., 2009, 2008).

CpCDV, which is one of only two dicot-infecting mastrevirus species known to occur outside of Australia, was identified here as potentially being a parent (or at least being most closely related to the actual parent) in four of the six inter-species events; all of which involved sequence transfers to viruses currently found in Australia. This information, together with the discovery of the Australian-like mastrevirus, CpYDV, in Pakistan, implies that an ancestor of CpCDV may have moved out of Australia (or at least that region of the world) and into the Middle East, Africa and the Indian subcontinent where it subsequently became established; a hypothesis supported by the phylogeographic analysis undertaken by Kraberger et al. (2013).

3.4. Signals of natural selection within dicot-infecting mastrevirus species

The large amount of CpCDV sequence data that we generated provided a good opportunity to compare and contrast signals of natural selection acting on the coding sequences of the various well sampled species of dicot-infecting mastreviruses. We used two different codon-model based methods, FUBAR and MEME, to infer selective processes acting on individual codon sites within the *mp* (Fig. 5A), *cp* (Fig. 5B) and the *rep* (Fig. 6) coding regions of all available CpCDV ($n = 205$), CpCAV ($n = 13$), CpCV ($n = 28$) and TYDV ($n = 9$) full-length genome sequences.

Due to the low numbers of available CpCAV and TYDV sequences these analyses only had sufficient power to detect statistically significant signals of natural selection at a few individual codon sites in each of these species.

Whereas Non-synonymous substitution rates/Synonymous substitution rates (dN/dS) values significantly lower than one imply negative or purifying selection favouring the maintenance

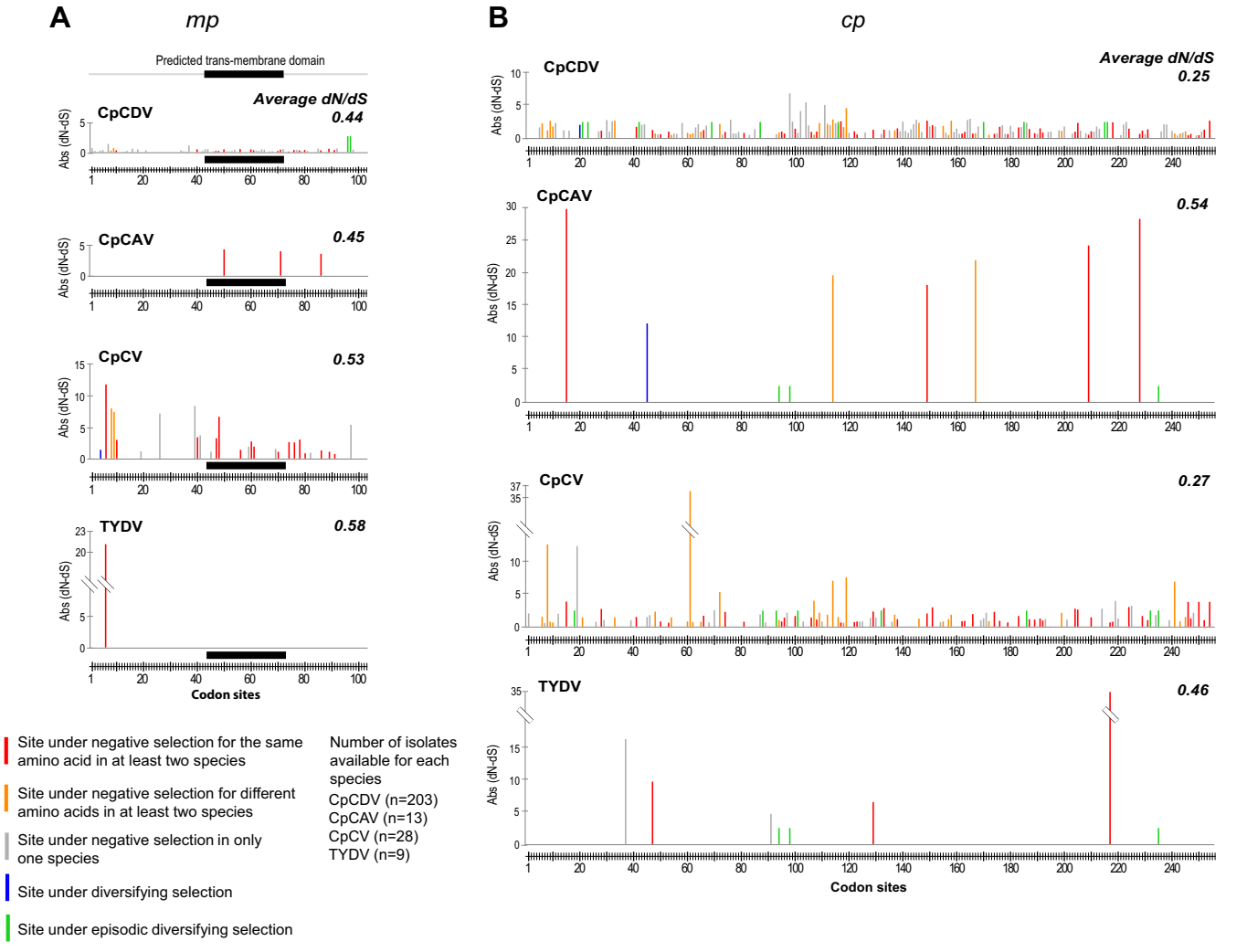

Fig. 5. Plot representing signals of natural selection acting on individual codon sites within (A) the *mp* and (B) the *cp* of CpCDV, CpCAV, CpCV and TYDV. Absolute (Abs) values of $dN–dS$ are plotted for positive selection (blue) and negative selection (orange, red and grey) signals with an associated FUBAR p-value <0.05. Abs values for episodic positive selection signals with an associated MEME p-value <0.05 are given in green. Bar heights for Abs ($dN–dS$) values correspond to the degree of positive or negative selection detected using FUBAR. Sites at which episodic diversifying selection was detected using MEME have been represented by green bars with uniform height across the genes since Abs ($dN–dS$) values averaged across the entire phylogeny do not accurately reflect degrees of episodic diversifying selection (which by definition occurs only on specific subsets of branches within the phylogeny). Overall, averages for the dN/dS ratios (all of which are significantly <1) are indicated for each gene and species. Codon sites are indicated based on a codon alignment of all species for each gene. The locations of the predicted trans-membrane domain (Boulton et al., 1993) in relation to their position in these alignments are shown. dN = Non-synonymous substitution rates and dS = Synonymous substitution rates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

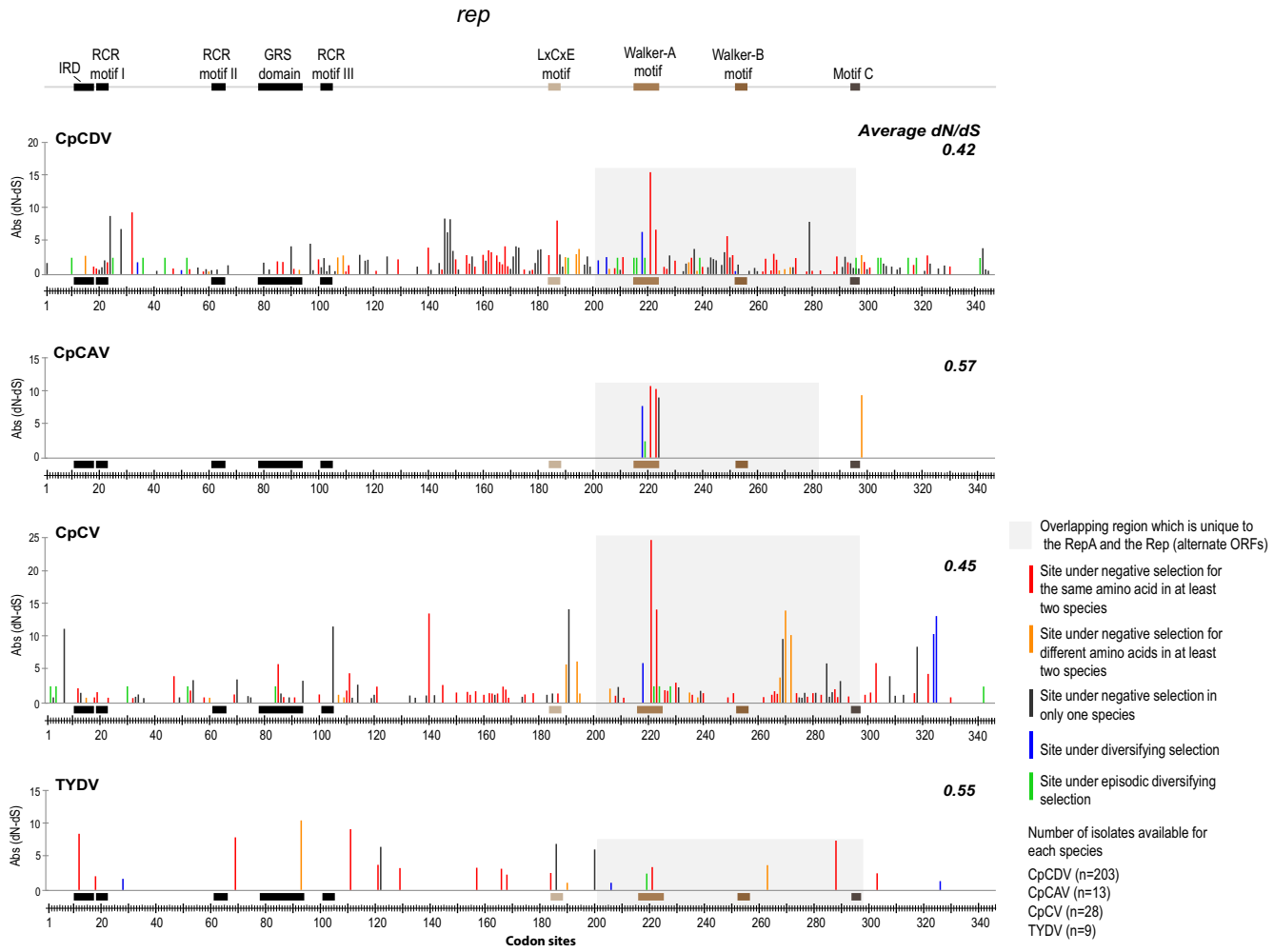

Fig. 6. Plot representing significant signals of natural selection acting on individual codon sites within the *rep* of CpCDV, CpCAV, CpCV and TYDV. Absolute (Abs) values of *dN–dS* are plotted for positive selection (blue) and negative selection (orange, red and grey) signals with an associated FUBAR *p*-value <0.05. Abs values for episodic positive selection signals with an associated MEME *p*-value <0.05 are given in green. Bar heights for Abs (*dN–dS*) values correspond to the degree of positive or negative selection detected using FUBAR. Sites at which episodic diversifying selection was detected using MEME have been represented by green bars with uniform height across the genes since Abs (*dN–dS*) values averaged across the entire phylogeny do not accurately reflect degrees of episodic diversifying selection (which by definition occurs only on specific subsets of branches within the phylogeny). Overall, averages for the *dN*/*dS* ratios (all of which are significantly <1) are indicated for each gene and species. Codon sites are indicated based on a codon alignment of all species for each gene. The locations of conserved domains and motifs in relation to their positions in these alignments are shown; iteron-related domain (IRD) (Argüello-Astorga and Ruiz-Medrano, 2001), rolling circle replication (RCR) motifs I, II and II (Ilyina and Koonin, 1992; Laufs et al., 1995; Rosario et al., 2012), the geminivirus Rep sequence (GRS) domain (Nash et al., 2011), and the helicase domain Walker-A, -B and motif C (Gorbalenya and Koonin, 1993; Gorbalenya et al., 1990). *dN* = Non-synonymous substitution rates and *dS* = Synonymous substitution rates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of amino acid sequences, *dN*/*dS* values significantly greater than one imply positive or diversifying selection favouring the modification of amino acid sequences. It is expected that most expressed viral proteins should have close to functionally optimal amino acid sequences and that their coding regions should therefore be evolving under predominantly negative selection (Duffy and Holmes, 2008; Kraberger et al., 2012; Shackelton et al., 2005; Stenzel et al., 2014). It is unsurprising then that all of the analysed dicot-infecting mastrevirus coding regions had *dN*/*dS* values that were significantly lower than one, with the values for the *mp* generally displaying the lowest degree of purifying selection (i.e., the highest *dN*/*dS*) and the *cp* the highest degree (i.e., the lowest *dN*/*dS*). It is important to note here that whereas it is valid to compare the magnitudes of the *dN*/*dS* values between different coding regions of the same species, unless the datasets for the different species being analysed have similar degrees of diversity, it is not valid to compare the magnitudes of *dN*/*dS* values for the same coding region in different species. In this regard, it is apparent that for all species the *cp* is evolving under stronger purifying selection than the *rep*. Also, with the exception of CpCAV, the *mp* is evolving under the weakest negative selection. Curiously, with CpCAV the *mp* is apparently evolving under the strongest negative selection. It should be noted, however, that the CpCAV dataset was both less diverse, and contained fewer sequences, than the other datasets analysed: factors which both strongly impact the power of the analyses we have performed.

The influence of dataset size and diversity is also clearly reflected in the differences between the datasets with respect to the numbers of codon sites detectably evolving under either positive or negative selection. We were nevertheless able to detect a number of individual codon sites that appear to be consistently evolving under negative selection in two or more of the analysed species (sites indicated in red and orange in Fig. 5 and 6). These sites reflect specific amino acid positions that are likely crucial to the functioning of the various expressed proteins. Codon sites indicated in red, (*mp* = 19 sites, *cp* = 48 sites and *rep* = 104 sites) reflect

particular residues within proteins that presently have amino acid states that are broadly adaptive in the context of multiple dicot-infecting mastrevirus hosts. Whereas sites in orange (*mp* = 2 sites, *cp* = 32 sites and *rep* = 27 sites) also reflect functionally important amino acid positions, the most adaptive amino acid at these positions differs from species to species. The amino acids at these sites are likely adaptive to features of niches that are specific to the different species. The large number of sites within the various coding regions that appear to be consistently evolving under negative selection is possibly due to the fact that these species all occupy similar ecological niches: something that is not surprising since they all have largely overlapping host ranges and similar vector species.

There are also a number of interesting patterns in the codon sites that are detectably evolving under either constant (in blue) or episodic (in green) positive selection (i.e., sites at which dN/dS is significantly higher than one in all or a significant fraction of lineages in the particular datasets analysed). For example, the first 52 codon sites of *rep* (the gene region encoding the portion of *rep* involved in recognition and binding to the virion strand origin of replication) contains an unusually high proportion of sites (7/52 in CpCDV and 5/52 in CpCV) that are evolving either under constant (indicated in blue) or episodic (indicated in green) positive selection (i.e. selection favouring change), with codon 52 evolving under episodic positive selection in both. This suggests that the optimal amino acid configuration in this part of the protein is in a state of flux, with, for example, different configurations perhaps being optimally suited to the different host species that these viruses infect. It is also noteworthy that this is the precise region of *rep* that is either most frequently exchanged during recombination amongst these viruses, or, when it is transferred, is frequently adaptive and therefore yields genomes that are favoured by natural selection.

In the portion of *rep* encoding the actual origin of replication recognition sequences (called the iteron related domain – labelled IRD in Fig. 6), three uniformly spaced codon sites at positions 12, 15 and 18 are detectably evolving under negative selection. This suggests that despite the apparently fluctuating selection pressures acting on the remainder of the DNA binding regions (rolling-circle replication motifs; RCR motifs I, II and III, and the superfamily 3 helicase motifs; walker-A, walker-B and Motif C) of the *rep*, the selective pressures on origin recognition are relatively constant across all the species examined here.

Another strikingly conserved pattern of positively and negatively selected codon sites occurs within the region of *rep* encoding the Walker-A motif (Fig. 6). This region of *rep* however also falls within the portion of the gene that is expressed in two different frames in Rep and RepA (indicated by a grey shaded box in Fig. 6). The apparently conserved positive selection signals detectable throughout this region of overlap between *rep* and *repA* are therefore possibly an artefact of negative selection acting simultaneously on the different proteins these genes encode. For example, the positive selection signals detected in the *rep* codons between positions 215 and 219 could simply be a consequence of negative selection acting to preserve the amino acid coding potential of overlapping codons in *repA*. Regardless of the causes of these positive selection signals, the highly conserved negative selection signals at positions 221 and 223 (both encoding a glutamic acid in three species) clearly indicate that selection is strongly favouring these two amino acids within the Rep Walker-A motif.

Other notably conserved negative selection signals occur in *mp* both between codons 6 and 10, and between codons 40 and 81 (Fig. 5). Whereas conserved negative selection signals are pervasive throughout *cp*, codons 94, 98, and 235 are all detectably evolving under episodic diversifying selection in both TYDV and CpCAV. Although the region encompassing codons 94 and 98 has not been identified as playing any role in transmission, the C-terminal region has been associated with vector specificity in the begomoviruses AbMV and TYLCV (Hohnle et al., 2001; Noris et al., 1998). Therefore it is possible that codon 235 maybe an important site for movement and possibly transmission efficiency.

4. Concluding remarks

In this study we analysed the diversity of CpCDV in Sudan by identifying, cloning and sequencing 145 CpCDV genomes from symptomatic pulse samples and included two full genome CpCDV sequences from Sudan which had been deposited in GenBank prior to this study. In addition an opportunistically sampled CpCDV isolate from Morocco was recovered. Amongst these isolates four new CpCDV strains have been identified, all with complex patterns of recombination. CpCDV-H, the predominant strain circulating in Sudan, is evidently also frequently recombining with several other CpCDV strains found within the country. The high frequencies of inter-strain recombination evident within these Sudanese viruses likely reflect high frequencies of infections containing multiple CpCDV strains.

Recently CpCDV has been found in the field infecting previously unsuspected hosts such as cotton and peppers (Akhtar et al., 2013; Manzoor et al., 2014). These and other recent CpCDV diversity studies have raised many questions with regard to the natural host range of this dicot-infecting mastrevirus species, and the role that genetic recombination might play in facilitating its emergence as a pathogen of important crops such as cotton (which is currently one of Sudan's principal cash crops). A survey in Sudan in 2002 using serological tests showed a high incidence of CpCDV-like mastreviruses in various wild plant species, one of which, pigeon pea (*Cajanus cajan*), is commonly planted by Sudanese farmers on the margins of their fields. This has prompted the suggestion that, in this country at least, pigeon pea may facilitate the circulation of CpCDV between other currently unknown uncultivated reservoir species and pulse crops (Ali et al., 2004). In Australia, TYDV-like mastreviruses were shown to have a wide host range (infecting species in seven dicot families) in areas and at a time when chickpeas were not widely grown (Thomas and Bowyer, 1984). It is clear that chickpea is particularly susceptible to, and visibly affected by, mastreviruses. In order to enable more informed CpCDV control strategies it might be worthwhile for future CpCDV sampling surveys to characterise dicot-infecting mastreviruses in uncultivated dicotyledon plant species that are commonly found growing in the same areas as important crop species such as cotton and chickpea.

GenBank accession numbers

KM229768–KM229913

Acknowledgements

We thank Georgina Limmer for her technical assistance in the lab. Simona Kraberger is supported by a scholarship from the School of Biological Sciences, University of Canterbury, New Zealand. Gordon W. Harkins, Darren P. Martin and Arvind Varsani are supported by the National Research Foundation of South Africa.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2014.11.024.

References

Abraham, A., Menzel, W., Lesemann, D.-E., Varrelmann, M., Vetten, H., 2006. *Chickpea chlorotic stunt virus*: a new polerovirus infecting cool-season food legumes in Ethiopia. Phytopathology 96, 437–446.

Akhtar, K.P., Ahmad, M., Shah, T.M., Atta, B.M., 2011. Transmission of *Chickpea chlorotic dwarf virus in chickpea* by the leafhopper *Orosius albicinctus* (Distant) in Pakistan – short communication. Plant Prot. Sci. 47, 1–4.

Akhtar, S., Khan, A.J., Briddon, R.W., 2013. A Distinct Strain of *Chickpea chlorotic dwarf virus* infecting pepper in Oman. Plant Dis. 98, 286–286.

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. Arab. J. Plant Protect. 22.

Argüello-Astorga, G.R., Ruiz-Medrano, R., 2001. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid–base pairs by a comparative approach. Arch. Virol. 146, 1465–1485.

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

Boulton, M.I., Pallaghy, C.K., Chatani, M., MacFarlane, S., Davies, J.W., 1993. Replication of *Maize streak virus* mutants in maize protoplasts: evidence for a movement protein. Virology 192, 85–93.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772–772.

Delport, W., Poon, A.F.Y., Frost, S.D.W., Kosakovsky Pond, S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26, 2455–2457.

Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. J. Virol. 82, 957–965.

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., Ahoonmanesh, A., 2008. PCR detection and partial molecular characterization of *Chickpea chlorotic dwarf virus* in naturally infected sugar beet plants in Iran. J. Plant Pathol. 90, 247–251.

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.

Gorbalenya, A.E., Koonin, E.V., 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. 3, 419–429.

Gorbalenya, A.E., Koonin, E.V., Wolf, Y.I., 1990. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. 262, 145–148.

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.

Hamed, A.A., Makkouk, K.M., 2002. Occurrence and management of *Chickpea chlorotic dwarf virus* in chickpea fields in northern Sudan. Phytopathol. Medit. 41, 193–198.

Hohnle, M., Hofer, P., Bedford, I.D., Briddon, R.W., Markham, P.G., Frischmuth, T., 2001. Exchange of three amino acids in the coat protein results in efficient whitefly transmission of a nontransmissible *Abutilon mosaic virus* isolate. Virology 290, 164–171.

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.

Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. 20, 3279–3285.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22, 3096–3098.

Kraberger, S., Harkins, G.W., Kumari, S.G., Thomas, J.E., Schwinghamer, M.W., Sharman, M., Collings, D.A., Briddon, R.W., Martin, D.P., Varsani, A., 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, 282–291.

Kraberger, S., Mumtaz, H., Claverie, S., Martin, D.P., Briddon, R.W., Varsani, A., submitted for publication. Identification of an Australian-like dicot-infecting mastrevirus in Pakistan. *Arch. Virol.* http://dx.doi.org/10.1007/s00705-014-2299-5.

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.

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. Medit. 47, 42–49.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., Gronenborn, B., 1995. Identification of the nicking tyrosine of geminivirus Rep protein. FEBS Lett. 377, 258–262.

Lefeuvre, P., Lett, J.-M., Varsani, A., Martin, D., 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J. Virol. 83, 2697–2707.

Liu, L., van Tonder, T., Pietersen, G., Davies, J.W., Stanley, J., 1997. Molecular characterization of a subgroup I geminivirus from a legume in South Africa. J. Gen. Virol. 78, 2113–2117.

Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary Analysis. Version 2.75 http://mesquiteproject.org.

Makkouk, K., Dafalla, G., Hussein, M., Kumari, S., 1995. The natural occurrence of chickpea chlorotic dwarf geminivirus in chickpea and faba bean in the Sudan. J. Phytopathol. 143, 465–466.

Makkouk, K.M., Fazlali, Y., Kumari, S.G., Farzadfar, S., 2002. First record of *Beet western yellows virus*, *Chickpea chlorotic dwarf virus*, *Faba bean necrotic yellows virus* and *Soybean dwarf virus* infecting chickpea and lentil crops in Iran. Plant Pathol. 51, 387–387.

Makkouk, K.M., Hamed, A.A., Hussein, M., Kumari, S.G., 2003. First report of *Faba bean necrotic yellows virus* (FBNYV) infecting chickpea (*Cicer arietinum*) and faba bean (*Vicia faba*) crops in Sudan. Plant Pathol. 52, 412–412.

Manzoor, M., Ilyas, M., Shafiq, M., Haider, M., Shahid, A., Briddon, R., 2014. A distinct strain of *Chickpea chlorotic dwarf virus* (genus *Mastrevirus*, family *Geminiviridae*) identified in cotton plants affected by leaf curl disease. Arch. Virol. 159 (5), 1217–1221.

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

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.

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.P., Brown, J.K., Navas-Castillo, J., Moriones, E., Zerbini, M.F., Rivera-Bustamante, R.F., Malathi, V.G., Briddon, R.W., 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.

Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9, e108277.

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

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S.L.K., Scheffler, K., 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol. Biol. Evol. 30, 1196–1205.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Pond, S.L.K., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764.

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.

Nash, T.E., Dallas, M.B., Reyes, M.I., Buhrman, G.K., Ascencio-Ibañez, J.T., Hanley-Bowdoin, L., 2011. Functional analysis of a novel motif conserved across geminivirus Rep proteins. J. Virol. 85, 1182–1192.

Noris, E., Vaira, A.M., Caciagli, P., Masenga, V., Gronenborn, B., Accotto, G.P., 1998. Amino acids in the capsid protein of *Tomato yellow leaf curl virus* that are crucial for systemic infection, particle formation, and insect transmission. J. Virol. 72, 10050–10057.

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® Classic Card technology and use of bacteriophage ϕ29 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 geminiviruses by frequent recombination. Virology 265, 218–225.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U.S.A. 98, 13757–13762.

Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol. 157, 1851–1871.

Shackelton, L.A., Parrish, C.R., Truyen, U., Holmes, E.C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. U.S.A. 102, 379–384.

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.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Stenzel, T., Piasecki, T., Chrząstek, K., Julian, L., Muhire, B.M., Golden, M., Martin, D.P., Varsani, A., 2014. Pigeon circoviruses display patterns of recombination, genomic secondary structure and selection similar to those of Beak and feather disease viruses. J. Gen. Virol. 95, 1338–1351.

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.

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

Thomas, J.E., Bowyer, J.W., 1984. Tobacco Yellow Dwarf Virus, CMI/AAB Descriptions of Plant Viruses, p. 4.

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., 2009. Comparative analysis of *Panicum streak virus* and *Maize streak virus* diversity, recombination patterns and phylogeography. Virol. J. 6, 194.

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., 2008. Recombination, decreased host specificity and increased mobility may have driven the emergence of *Maize streak virus* as an agricultural pathogen. J. Gen. Virol. 89, 2063–2074.