

# SCIENTIFIC REPORTS



OPEN

## Maize chlorotic mottle virus exhibits low divergence between differentiated regional sub-populations

Luke Braidwood<sup>1</sup>, Diego F. Quito-Avila<sup>2</sup>, Darlene Cabanas<sup>3</sup>, Alberto Bressan<sup>4,5</sup>, Anne Wangai<sup>6</sup> & David C. Baulcombe<sup>1</sup>

Maize chlorotic mottle virus has been rapidly spreading around the globe over the past decade. The interactions of maize chlorotic mottle virus with Potyviridae viruses causes an aggressive synergistic viral condition - maize lethal necrosis, which can cause total yield loss. Maize production in sub-Saharan Africa, where it is the most important cereal, is threatened by the arrival of maize lethal necrosis. We obtained maize chlorotic mottle virus genome sequences from across East Africa and for the first time from Ecuador and Hawaii, and constructed a phylogeny which highlights the similarity of Chinese to African isolates, and Ecuadorian to Hawaiian isolates. We used a measure of clustering, the adjusted Rand index, to extract region-specific SNPs and coding variation that can be used for diagnostics. The population genetics analysis we performed shows that the majority of sequence diversity is partitioned between populations, with diversity extremely low within China and East Africa.

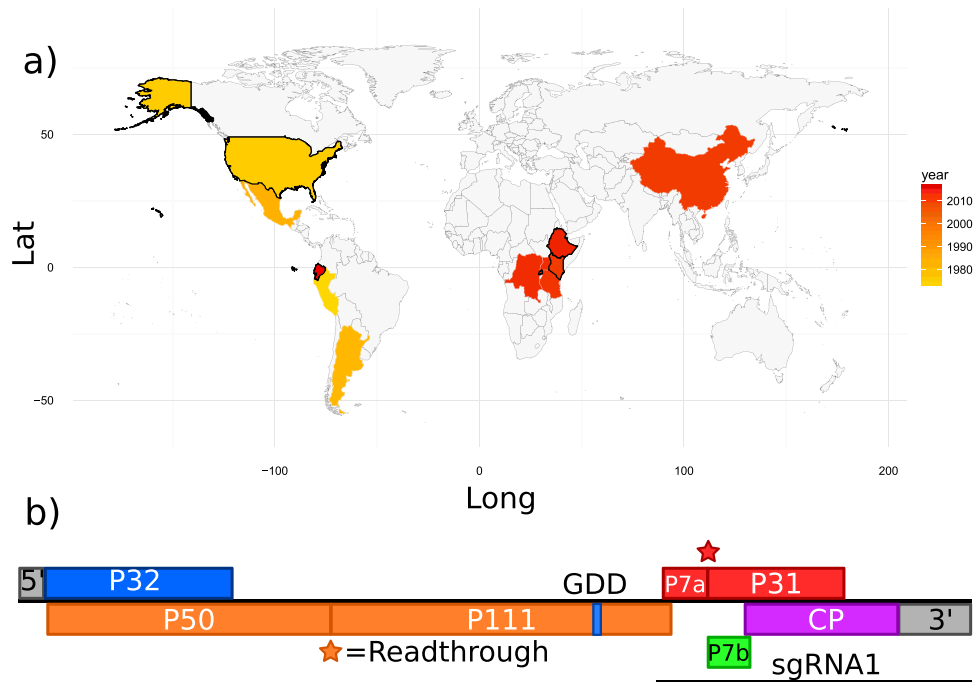
Maize chlorotic mottle virus (MCMV) is a positive-sense single-stranded RNA virus, and the sole member of the *Machlomovirus* genus in the *Tombusviridae* family. The virus was described in Peru, then reported shortly afterwards in Brazil, Argentina and the USA in the 1970s<sup>1-3</sup>. Within the past decade, MCMV has spread globally, with first reports in China, Taiwan, Ecuador, Spain, and widely in East Africa (Fig. 1a)<sup>4-9</sup>. MCMV can spread via soil water, seed, mechanical transmission, and is semi-persistently vectored by chrysomelid beetles and thrips (*Frankliniella*)<sup>3,7,10-14</sup>.

MCMV interacts synergistically with members of the *Potyviridae* family: the potyviruses sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV), and the tritimovirus wheat streak mosaic virus (WSMV). The interaction causes a more aggressive condition known as maize lethal necrosis (MLN). MLN causes around 80% yield loss in heavily affected areas, and destroyed an estimated 22% of Kenya's 2013 maize crop<sup>7,15</sup>. SCMV is the reported co-infecting virus in recent reports of MLN outbreaks in China, East Africa, and Ecuador<sup>7,8,16</sup>. The recent spread of MLN reflects the spread of MCMV since SCMV has been present in East Africa, China and South America for decades<sup>17-19</sup>.

In the absence of additional control measures or resistant varieties, it is likely that MLN will spread further. Ecological niche modelling suggests that large areas of Eastern sub-Saharan Africa are at high risk of MLN outbreaks<sup>20</sup>. Due to its rapid spread and interaction with local viruses, and the absence of resistant commercial maize lines, MCMV represents a significant threat to maize production in Sub-Saharan Africa, where it is the most important cereal crop. To investigate the spread of MCMV globally, and inform future control strategies, we decided to study genomic variation in MCMV isolates.

MCMV has a monopartite 4.4 kb genome (Fig. 1b), which generates 1.4 kb and 0.37 kb sub-genomic RNAs during infection<sup>21</sup>. The genome encodes conserved tombusviridae proteins - the RNA-dependent RNA

<sup>1</sup>University of Cambridge, Plant Sciences, Cambridge, CB2 3EA, United Kingdom. <sup>2</sup>Centro de Investigaciones Biotecnológicas del Ecuador, Facultad de Ciencias de la Vida, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador. <sup>3</sup>Department of Molecular Biosciences and Bioengineering, University of Hawaii, 3050 Maile Way, Gilmore Hall 310, Honolulu, Hawaii, USA. <sup>4</sup>Department of Plant and Environmental Protection Sciences, University of Hawaii, 3050 Maile Way, Gilmore Hall 310, Honolulu, Hawaii, USA. <sup>5</sup>Bayer CropScience LP, R&D Trait Research, 3500 Paramount Parkway, Morrisville, USA. <sup>6</sup>Kenya Agricultural and Livestock Research Organization (KALRO), Nairobi, Kenya. Correspondence and requests for materials should be addressed to L.B. (email: [lab70@cam.ac.uk](mailto:lab70@cam.ac.uk))



**Figure 1.** (a) Global distribution of maize chlorotic mottle virus (MCMV), coloured by year of first report<sup>55</sup>. Countries sampled in this study are outlined in black. Map generated using data from the maps package using ggplot2 in R v3.4.1. (b) MCMV genome structure.

polymerase and associated protein (P111 and P50) and two movement proteins (P7a and P7b). MCMV expresses two unique proteins, P31 and P32, which have unknown function but are linked to systemic spread and viral accumulation respectively. A viral suppressor of silencing (VSR) is yet to be identified<sup>22</sup>. P31 and P111 are expressed as readthrough proteins of P7a and P50, respectively, with P111 expressed through a conserved tom-busviridae amber stop codon readthrough motif (UAGGGR)<sup>23</sup> (Fig. 1b).

Full length MCMV genome sequences are available for North American, Chinese, and East African isolates, but until now there have been no data for South American and Hawaiian isolates. Here we use a combination of next-generation sequencing (NGS) and Sanger sequencing to obtain 37 novel MCMV genome sequences from East-Africa, Ecuador, and Hawaii, construct a first global phylogeny of MCMV, and identify genomic regions of interest for functional analysis and strain-typing. We also report evidence for recombination, and a natural non-sense mutation of P31 present in Hawaiian and Ecuadorian isolates.

## Results and Discussion

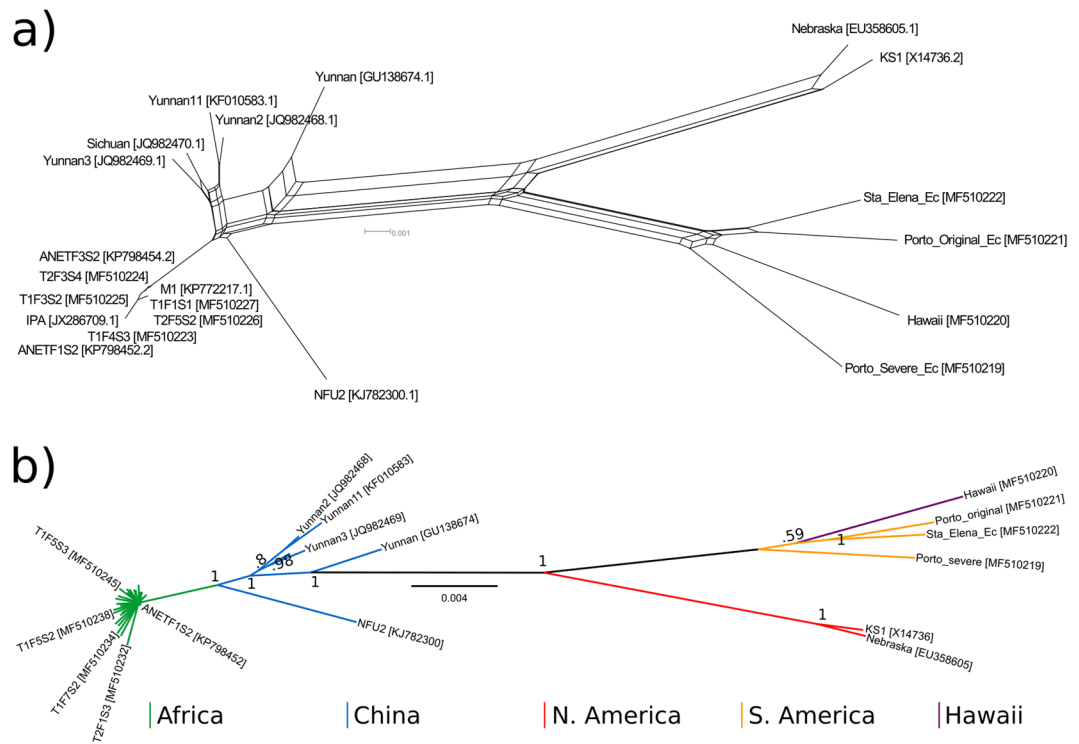
**Sequencing of MCMV isolates.** Between 2012 and 2016, we collected MCMV-infected maize leaves and generated 37 novel MCMV genome sequences from Kenya ( $n = 24$ ), Ethiopia ( $n = 5$ ), Rwanda ( $n = 4$ ), USA (Hawaii,  $n = 1$ ), and Ecuador ( $n = 3$ ) (Supplemental Table S1). African sequences were obtained via NGS, and the remainder through Sanger sequencing (see methods). GenBank accession numbers and isolate details are in Table S1. All NGS samples also contained the known MCMV partner virus SCMV, and a number contained additional viruses (Supplemental Table S1). Sequence lengths ranged from 4399–4440 bp; this is 99–100% coverage compared to the longest previously reported MCMV genome sequence. Including all GenBank isolates, nucleotide identity between genomes ranged from 100% to 96.55%. Nucleotide diversity between all sequences was 0.01, which is very low for viruses generally and similar to the lowest recorded diversity values in the tom-busviridae (Table 1)<sup>24–26</sup>.

**Phylogenetic analysis.** Phylogenetic analyses assume a single evolutionary history of each genome, an assumption which may be violated in the case of extensive recombination. We therefore performed a splits network analysis, which detects conflicting phylogenetic signals that can be caused by recombination. There is clear separation of isolates from different regions, indicating that there has been no recombination between MCMV genomes in geographically isolated regions (Fig. 2a). Conventional phylogenetic analysis is suitable, therefore, for investigating the relationships between regions. To construct an unrooted phylogenetic tree, we used a nucleotide alignment containing all novel MCMV isolates, including ambiguous bases at the 10% threshold, and genome sequences available in GenBank. We chose nucleotide alignment to enhance resolution due to the low divergence between MCMV isolates, and used Bayesian inference (MrBayes 3.2) to generate the tree (Fig. 2b). Clearly resolved clades contain North American isolates, Hawaiian with South American isolates, and Chinese with African isolates.

The phylogenetic tree shows clear separation between geographic regions, with the exception of Hawaii and Ecuador. Phylogenetic confidence estimates (Fig. 2b) and population genetics (below) illustrate that the majority

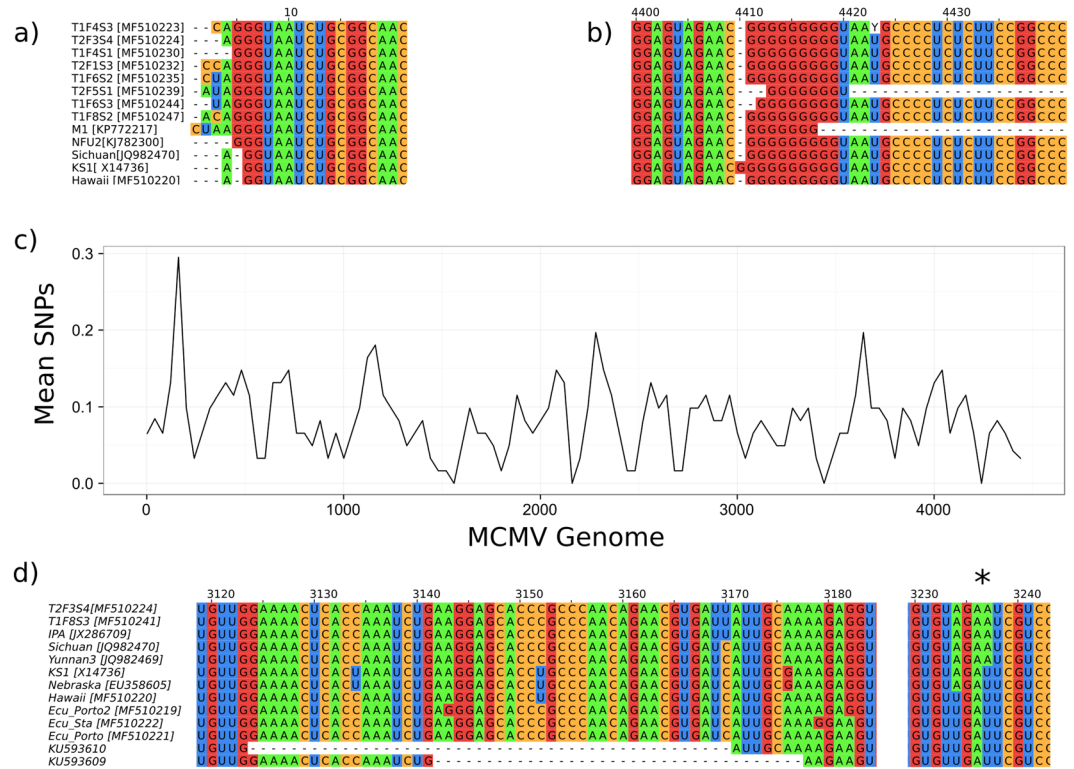
Virus	Family	Isolate number	Sequence type	Number of continents	Average nucleotide diversity	Ref
MCMV	<i>Tombusviridae</i>	49	Genome	4	0.01	This study
MCMV	<i>Tombusviridae</i>	49	CP	4	0.01	This study
OLV1	<i>Tombusviridae</i>	21	CP	3	0.02	26
OMMV	<i>Tombusviridae</i>	21	CP, in 81nt winows	1	0–0.02	25
TNV-D	<i>Tombusviridae</i>	14	CP, in 81nt winows	1	0–0.02	25
TMGMV	<i>Virgaviridae</i>	53	183 K ORF	3	0.057	56
CTV	<i>Closteroviridae</i>	30	1a, CP, 20k ORF	2	0.068	57
CYSDV	<i>Closteroviridae</i>	71	CP	3	0.002	58
RYMV	<i>Sobemovirus</i>	40	CP	1	0.0194	59
RTSV	<i>Sequiviridae</i>	38	Two CP genes	1	0.09	60
CMV-sat	<i>Bromoviridae</i>	25	Genome (satellite)	5	0.13	61

**Table 1.** Comparison of average nucleotide diversities calculated for positive-sense ssRNA viruses. CMV-sat = *Cucumber mosaic virus satellite RNA*, CTV = *Citrus tristeza virus*, CYSDV = *Cucurbit yellow stunting disorder virus*, MCMV = *Maize chlorotic mottle virus*, OLV1 = *Olive latent virus-1*, OMMV = *Olive mild mosaic virus*, RTSV = *Rice tungro spherical virus*, RYMV = *Rice yellow mottle virus*, TNV-D = *Tobacco necrosis virus D*, TMGMV = *Tobacco mild green mosaic virus*.



**Figure 2.** (a) Splits network of maize chlorotic mottle virus (MCMV) genomes, distances calculated with uncorrected P, and network generated by neighbour-net in Splitstree V4.6. (b) MCMV phylogeny generated using Bayesian inference in Mr Bayes 3.2. Scale bars are nucleotide substitutions per site.

of MCMV genome diversity occurs between sub-populations. For China, East Africa, and Ecuador, intra-group diversity is minimal, suggesting either (separate) single introduction events or repeated introductions from the same sub-populations. The most parsimonious method to infer the source of a viral outbreak is to identify the isolates most closely related to that outbreak<sup>27</sup>. Phylogenetic proximity suggests links between China and Kenya, and Hawaii and Ecuador. Due to the earlier presence of MCMV in China and Hawaii compared to Kenya and Ecuador respectively, the most likely explanation for these data is that China is the source of the East African outbreak, and Hawaii the source of the Ecuadorian. However corroborating evidence and more complete sampling of global MCMV genomes would be required to confirm this. For example, sequence data from other Central and South American regions with MCMV presence could determine whether the MCMV outbreak in Ecuador is linked most closely to neighbouring countries or Hawaii. It would also determine whether the observed limited sequence divergence amongst MCMV isolates is universal, suggesting either recent evolution or intense purifying



**Figure 3.** Genomic termini of the maize chlorotic mottle virus (MCMV) genome, 5' (a) and 3' (b). Sequences with identical termini have been removed for clarity. (c) SNP distribution across the MCMV genome, in 60 bp windows calculated every 40 bp<sup>55</sup>. (d) Variation in P31, with early stop codon indicated by asterisk.

selection, or whether currently undocumented diversity is present in its presumed ancestral range - the ancestral ranges of *Teosinte* and maize.

The similarity of Ecuadorian and Hawaiian isolates could be explained by seed transmission. Seed transmission of MCMV has been reported at a rate of 0–0.33%<sup>12</sup>. This rate may be an underestimate as parent populations were naturally infected (so likely had <100% infection). Twelve of 26 ten-seed samples bought at Kenyan markets and tested by RT-PCR were positive for MCMV, and 72% of kernels from a single plant, which demonstrates MCMV presence in commercial maize seed, but not transmission<sup>7</sup>. Seeds from two commercial hybrids planted in the region of the 2015 Ecuadorian outbreak were grown in sterile soil inside insect-proof growth chambers, and 8 and 12% of seedlings were MCMV-infected<sup>8</sup>. These isolates, and the Ecuadorian genomes collected in this study, were most closely related to the Hawaiian genome. Hawaii is an important region for maize seed production for a number of major agricultural companies, although it is unknown whether the infected seeds from Ecuador were produced in Hawaii. Regardless, it is vital that the ability of MCMV to be transmitted via maize seed is investigated further, especially given Hawaii's central role in maize seed production and the rapid spread of MCMV across East Africa. However, given the low rate of seed transmission, and the high populations of thrip vectors in the *Frankliniella* genus, we would suggest although seed transmission may be important for long distance dispersal, once arriving in an area vector and soil transmission will likely be more important for local spread<sup>28,29</sup>.

**SNP and coding analysis.** Structural variation within whole genomes was limited to the genomic termini. We found small 5' extensions of the genome - African isolates and the Taiwanese sweetcorn isolate (KJ782300.1) have a G insertion, giving a 5' terminal AGGG, as opposed to the AGG found in isolates from other regions. The G9 sequence at the 3' terminus of the genome has a G insertion in North American isolate X14736.2, while the first reported Kenyan isolate, JX286709.1, has a G insertion immediately downstream of the TAAT sequence in the 3' terminus (Fig. 3a and b).

There were 419 mutations at 388 polymorphic sites across the MCMV genome in our dataset (Fig. 3c). To identify geographic region-specific genomic variation we used the Adjusted Rand Index (ARI). The ARI is a chance-adjusted measure of the similarity between two data clusterings, varying between -1 (dissimilar) and 1 (similar). We used the phylogenetic clades (Africa, China, North America, and South America with Hawaii) as one data cluster, and a site in the nucleotide alignment for the other. Therefore, ARI values above zero were produced when SNPs segregated according to the phylogenetic clades. To identify sites which clustered significantly better than chance, we randomised the members of each geographic clade and recorded the ARI across the genome 1000 times. Sites above the 95% level of the randomised ARI scores were taken to be significant. Significant sites were extracted, identifying candidate SNPs for (uncharacterised) phenotypic variation between

Population	Genomes	Segregating sites	No. of haplotypes	Haplotype diversity	Average no. of SNPs	Nucleotide diversity
Global	49	388	42	0.986	45.5	0.0104
South America and Hawaii	4	123	4	1	64.7	0.0148
North America	2	22	2	1	22	0.00505
China	6	98	6	1	36.1	0.00829
Africa	37	114	30	0.974	7.51	0.00172

**Table 2.** *Maize chlorotic mottle virus* genome diversity globally and within phylogenetic clades.

clades. This information can be used to design diagnostics appropriate for separating isolates from different regions, which we illustrated by matching randomly selected and labelled MCMV sequences to their region using SNPs in a 500 bp interval (Supplemental Table S3).

To identify protein coding changes within the clades we repeated our ARI analysis on translated alignments for each MCMV ORF (Supplemental Tables S4–S9). The most significant variation was observed in the coat protein gene and the P7a/P31 region in which the systemic movement protein P31 is expressed by readthrough of the P7a ORF UGA stop codon. We observed unreported variation in an exposed loop of the MCMV capsid - Phe76Leu in Chinese sugarcane (KF010583.1). Exposed capsid loops can function in vector transmission, and Wang *et al.* (2015) previously identified two variable residues in the exposed loops of the MCMV capsid, Pro81Ser in Nebraska (NC\_003627) and Kansas (EU358605.1) isolates, and Ala62Asp in the 2012 Kenyan isolate (JX286709.1)<sup>5,30,31</sup>.

P31 enhances systemic MCMV movement via an unknown mechanism<sup>22</sup>. However, in Ecuadorian and Hawaiian isolates there is an early stop codon 18 bp downstream of the P7a UGA stop codon, truncating the majority of P31 (Fig. 3d). This unexpected stop codon is also present in the two previously reported partially sequenced Ecuadorian MCMV isolates, in which there are also 47 and 36 bp deletions in the P7a and P31 coding sequences upstream of the stop codon<sup>8</sup>. MCMV P31 has unknown function, a unique carboxy-terminal extension, and mutagenesis experiments suggest that it promotes systemic movement<sup>22</sup>. Both the Hawaiian isolate and Ecuadorian isolates contain an early stop codon, truncating P31 six amino acids after the read-through stop codon of the movement protein P7a. This is unlikely to be a sequencing error, as >5 independent clones were sequenced in Hawaii and Ecuador for each isolate. The stop codon leaves 162 nt of non-coding RNA in the Hawaiian and Ecuadorian isolates, before the capsid gene initiates. Interestingly, the partial Ecuadorian sequences reported previously also contain this stop codon, as well as deletions upstream within the P7a coding region<sup>8</sup>, which could represent selection pressure for smaller genomes once the function of P31 was lost, which is commonly observed in RNA viruses<sup>32</sup>.

Scheets (2016) used a mutant cDNA clone with an early stop codon mutation (p7bQ12N) to investigate P31 function, similar to the natural mutants we report from Hawaii and Ecuador<sup>22</sup>. Inoculation with p7bQ12N resulted in slower systemic spread of MCMV, and sequencing showed that the systemically infected leaves contained either a mixture of p7bQ12N with pseudorevertant (the stop codon removed), or exclusively pseudorevertant genomes. Mixtures were predominantly p7bQ12N, raising the possibility that the wild isolates from this study contained low-frequency genomes which complement the predominant genomes with mutated P31. Complementation of this form has been observed in tomato aspermy cucumovirus – 76% of genomes had a mutated movement protein, which were complemented by a minority of wild-type viral genomes within the host<sup>33</sup>. Experimentally mixing mutant and wild-type MCMV transcripts could establish whether this is a possibility for MCMV.

**Population genetics analysis.** We examined diversity and population structure with all novel and publicly available MCMV sequences. Overall sequence diversity was low, with 45.5 nucleotide differences between sequences on average (Table 2). Haplotype diversity was high, 0.99, with 42/49 haplotypes unique, although these are separated by a low number of SNPs. For subpopulation analysis, China and Africa were considered separately to allow estimation of gene flow between the populations. In terms of nucleotide diversity, the most diverse clade was South American and Hawaiian ( $\pi = 0.015$ ), while the least was Africa ( $\pi = 0.0017$ ), with only 7.51 nucleotide differences between African sequences on average. Variation in nucleotide identity was extremely limited for both Chinese and African sequences, with >99% sequence identity within each group. This high level of similarity suggests single introduction events for both the Chinese and East African outbreaks of MCMV.

To test genetic differentiation between subpopulations we used Hudson's test statistics ( $K_{st}$ ,  $K_{st}^*$ ,  $Z$ ,  $Z^*$ ,  $S_{nn}$ ), with all statistics indicating significant differentiation (Table 3)<sup>34,35</sup>.  $S_{nn}$  is powerful at all sample sizes and diversities, and so is most appropriate in this case, due to uneven sample sizes<sup>35</sup>. Hudson's Hst tests differentiation based on haplotype diversity, and was not significant, presumably due to the high proportion of unique haplotypes across all populations. Likewise,  $F_{st}$  and  $N_{st}$  values of 0.74 indicate that a high proportion of the genetic variation in MCMV is explained by population structure<sup>36,37</sup>.  $F_{st}$  and  $N_{st}$  values generated from pairwise comparisons between populations (Table 4) illustrate that the sub-populations with least genetic variation explained by population structure are China and Africa, which is expected given their phylogenetic proximity.

We used the Hypothesis testing using Phylogenies (HyPhy) server ([www.datamonkey.org](http://www.datamonkey.org)) to test for selection using algorithms based on dN/dS ratios, and alignments corresponding to each ORF in the MCMV genome. Mixed-Effects Model of Evolution (MEME) was used to detect episodic positive selection, with significant ( $p < 0.01$ ) codons found in P50, P32, P31, and P111 (Supplemental Table S2)<sup>38</sup>. Internal branches Fixed Effects

Population statistic	Data type	Value	P-value
$H_s$	Haplotype	0.978	0.072
$K_s$	Nucleotide	16.3	0
$K_s^*$	Nucleotide	2.24	0
Z	Nucleotide	400	0
$Z^*$	Nucleotide	5.69	0
$S_{nn}$	Nucleotide	1	0

**Table 3.** *Maize chlorotic mottle virus* population differentiation.

Population 1	Population 2	$H_s$	$K_s$	$N_{st}$	$F_{st}$
South America and Hawaii	North America	1	50.4	0.684	0.680
South America and Hawaii	China	1	47.5	0.638	0.633
South America and Hawaii	Africa	0.976	13.1	0.747	0.744
North America	China	1	32.6	0.790	0.786
North America	Africa	0.974	8.25	0.900	0.898
China	Africa	0.977	11.5	0.511	0.511

**Table 4.** *Maize chlorotic mottle virus* subpopulation differentiation.

Likelihood (iFEL) was used for detection of positive and negative selection in each MCMV ORF (Supplemental Table S2). Significant ( $p < 0.01$ ) positive selection on codons was detected in P50, P32, and P31, with significant negative (purifying) selection found at sites in all ORFs (Supplemental Table S2). Examining the density of negatively selected sites across the genome showed that the regions under most intense purifying selection corresponded to the readthrough region of P111, and the C-terminus of the coat protein.

Unlike previous dN/dS-based analyses of MCMV, we detected significant positive selection in the genome: at sites in P50, P32, and P31, three proteins of unknown function<sup>39</sup>. This can be attributed to the greatly increased volume of sequence data included in the analysis. The positive selection may represent adaptation to specific host interactions across different maize lines, or to different environments. We detected purifying selection across the MCMV genome, especially in the post-readthrough region of the RdRP, which is highly conserved amongst tombusviridae members, and the C-terminus of the CP. This is the location of an asymmetric unit important for interaction between capsid monomers in the assembled viral coat<sup>40</sup>. dN/dS analyses of selection in viral genomes can be constrained by overlapping ORFs and selection at the RNA sequence level to preserve functional RNA structures, such as the tombusviridae subgenomic RNA promoters or 3' Cap Independent Translational Enhancers<sup>41</sup>.

## Methods

**NGS of East African MCMV isolates.** During August 2014 maize leaf samples were collected from Kenya and Ethiopia, stored on dry ice in RNA-later (Ambion), then RNA was extracted using Trizol (Ambion) according to manufacturer's instructions. An additional Rwandan maize leaf RNA sample was received from FERA (UK). Ribosomal RNA (rRNA) depletion was performed with Ribo-Zero Magnetic Kit (Plant Leaf-Epicentre). Indexed stranded libraries were constructed using Scriptseq V2 RNA-Seq Library Preparation kits and Scriptseq Index PCR primers (Epicentre). Purification steps were performed using Agencourt AMPure XP beads. Library quantity and quality were checked using Qubit (Life Technologies) and on a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies). Libraries were sent to Beijing Genomics Institute for 100 bp paired-end sequencing on one lane of a HiSeq 2000 (Illumina).

**NGS sequence assembly.** Libraries were demultiplexed allowing one error within the index sequence using a custom python script, then adaptors trimmed using Trim galore! (parameters: -phred64 -fastQC -illumina-length 30 -paired -retain\_unpaired input\_1.fq input\_2.fq)<sup>42,43</sup>. Deduplication was performed by string-matching using scripts from the Quality Assessment of Short Read (QUASR) pipeline<sup>44</sup>. A bowtie reference was constructed from a fasta containing all 35 publicly available MCMV sequences using the bowtie2-build command, producing an index with each MCMV sequence as a separate chromosome. Reads were aligned to this index in order to capture all MCMV reads using bowtie2 (parameters: -D 20 -R 2 -N 1 -L 20 -i S,0,2.50 -phred64 -maxins 1000 -fr)<sup>45</sup>. Reads aligning to MCMV were assembled using Trinity, MCMV contigs of 1 kb or more were extracted by blast, and contigs were manually checked and assembled if necessary<sup>46</sup>. Libraries were then realigned to their respective trinity contigs by bowtie2, pileups generated by samtools, and consensus sequences called using QUASR script pileup\_consensus.py, with a threshold of zero or ten % of reads to use ambiguity codes (parameters: -ambiguity 2|10-dependent-cutoff 25 -lowcoverage 20)<sup>47</sup>.

**Sanger sequencing of Hawaiian and Ecuadorian samples.** Infected Hawaiian maize leaves were screened by ELISA for the presence of maize mosaic virus, wheat streak mosaic virus, maize dwarf mosaic virus, MCMV and SCMV. The leaf used for RNA extraction was negative for all viruses except SCMV. RNA was extracted using Trizol, and RT-PCR performed. Primers based on those used to sequence the Nebraska isolate (EU358605.1) were used to amplify six amplicons covering the genome<sup>48</sup>. Amplicons were cloned into

pCR4-TOPO TA plasmids and Sanger sequenced using M13 and internal primers to obtain total coverage. A consensus sequence was determined by sequencing six full-length clones.

A total of three Ecuadorean isolates were fully sequenced for this study. Primers used to amplify the genomes were designed from conserved regions using an alignment of available sequences including a local isolate obtained previously by degenerate-oligonucleotide-primed RT-PCR using double-stranded RNA as template<sup>8</sup>. Overlapping PCR amplification products (three independent reactions for each primer set) were sequenced in both directions by the Sanger method (Macrogen, South Korea).

**Sequence alignment.** MCMV genome sequences were aligned using MUSCLE in MEGA6, with a gap extension cost of  $-1000$ . The alignment was checked and refined manually in JALview<sup>49,50</sup>.

**Recombination Analysis.** SplitsTree4 was used to generate splits networks, using the default settings - distances were calculated by uncorrected P (match option for ambiguous bases), and network generated by neighbour-net<sup>51</sup>. RDP4 was used to examine the evidence for recombination, utilising the algorithms RDP, GENECONV, Chimeara, MaxChi, BOOTSCAN and SISCAN<sup>52</sup>. A number of highly similar (>99.5% identity) African sequences were excluded from the analysis to make figures clearer and decrease computational time in RDP4.

**Phylogenetic Tree Construction.** Alignment sites were split into three partitions: A) non-coding B) codon positions one and two and overlapping ORFs C) codon position three. To generate phylogenetic trees two runs of four Monte Carlo Markov Chain (MCMC) computations were run for 1000,000 generations under a general-time-reversible (GTR) model with a gamma distribution of rate variation between sites in MrBayes 3.2<sup>53</sup>. The first 10% of generations were discarded as burn-in (default). Convergence and effective sample size were examined using Tracer to confirm that estimated sample sizes for each parameter exceeded 200, as recommended by the MrBayes manual. We pooled together 1800 trees sampled every 500 generations (default) and constructed consensus trees.

**Population genetics and SNP analysis.** To extract alignment sites at which data clustered similarly to phylogeny, we used the Adjusted Rand Index to measure the clustering between phylogenetic groupings and alignment sites, extracting those with a score above zero. This was performed using custom R scripts. Population genetics indices and statistics were produced using an alignment of MCMV sequences without ambiguity codes (i.e. 0% threshold) in DnaSP v5, as DnaSP v5 does not accept ambiguity codes<sup>54</sup>. Significance for test statistics was assessed by a permutation test with 1000 replications. To extract alignment sites with data clustering similarly to phylogeny, we used the Adjusted Rand Index to compare the phylogenetic groupings with alignment sites, extracting those with a score above zero.

## References

- Castillo, J. & Hebert, T. Nueva enfermedad virosa afectando al maiz en el peru. *Fitopatologia* **9**, 79–84 (1974).
- Teyssandier, S. F. N. E. E. & Bo, E. Maize virus diseases in Argentina. In *Proceedings International Maize Virus Disease Colloquium and Workshop*, 87–92 (1983).
- Uyemoto, J. K. Biology and control of maize chlorotic mottle virus. *Plant disease* **67**, 7–10 (1983).
- Xie, L. *et al.* Characterization of Maize Chlorotic Mottle Virus Associated with Maize Lethal Necrosis Disease in China. *Journal of Phytopathology* **159**, 191–193, <https://doi.org/10.1111/j.1439-0434.2010.01745.x> (2011).
- Adams, I. P. *et al.* Use of next-generation sequencing for the identification and characterization of Maize chlorotic mottle virus and Sugarcane mosaic virus causing maize lethal necrosis in Kenya. *Plant Pathology* **62**, 741–749, <https://doi.org/10.1111/j.1365-3059.2012.02690.x> (2013).
- Lukanda, M. *et al.* First Report of Maize chlorotic mottle virus Infecting Maize in the Democratic Republic of the Congo. *Plant Disease* **98**, 1448–1448, <https://doi.org/10.1094/PDIS-05-14-0484-PDN> (2014).
- Mahuku, G. *et al.* Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-Saharan Africa. *Phytopathology* <https://doi.org/10.1094/PHYTO-12-14-0367-FI> (2015).
- Quito-Avila, D. F., Alvarez, R. A. & Mendoza, A. A. Occurrence of maize lethal necrosis in Ecuador: a disease without boundaries? *European Journal of Plant Pathology* **1914** <https://doi.org/10.1007/s10658-016-0943-5> (2016).
- Achon, M., Serrano, L., Clemente-Orta, G. & Sossai, S. First Report of Maize chlorotic mottle virus on a Perennial Host, Sorghum halepense, and Maize in Spain. *Plant Disease* **101**, 393 (2017).
- Nault, L. R. *et al.* Transmission of maize chlorotic mottle virus by chrysomelid beetles. *Phytopathology* **68**, 1071–1074 (1978).
- Jensen, S. G. Laboratory transmission of maize chlorotic mottle virus by three species of corn rootworms. *Plant Disease* **69**, 864–868, <https://doi.org/10.1094/PD-69-864> (1985).
- Jensen, S. G. Seed Transmission of Maize Chlorotic Mottle Virus. *Plant Disease* **75**, 497, <https://doi.org/10.1094/PD-75-0497> (1991).
- Cabanas, D., Watanabe, S., Higashi, C. & Bressan, A. Dissecting the Mode of Maize Chlorotic Mottle Virus Transmission (Tombsviridae: Machlomovirus) by Frankliniella williamsi (Thysanoptera: Thripidae). *Journal of Economic Entomology* **106**, 16–24, <https://doi.org/10.1603/EC12056> (2013).
- Zhao, M., Ho, H., Wu, Y., He, Y. & Li, M. Western Flower Thrips (Frankliniella occidentalis) Transmits Maize Chlorotic Mottle Virus. *Journal of Phytopathology* **162**, 532–536, <https://doi.org/10.1111/jph.12217> (2014).
- De Groote, H., Oloo, F., Tongruksawattana, S. & Das, B. Community-survey based assessment of the geographic distribution and impact of maize lethal necrosis (MLN) disease in Kenya. *Crop Protection* **82**, 30–35, <https://doi.org/10.1016/j.cropro.2015.12.003> (2016).
- Wang, Q. *et al.* Further characterization of Maize chlorotic mottle virus and its synergistic interaction with Sugarcane mosaic virus in maize. *Scientific Reports* **7**, 39960, <https://doi.org/10.1038/srep39960> (2017).
- Louie, R. Sugarcane Mosaic Virus in Kenya. *Plant Disease* **64**, 944, <https://doi.org/10.1094/PD-64-944> (1980).
- Chen, J., Chen, J. & Adams, M. J. Characterisation of potyviruses from sugarcane and maize in China. *Archives of Virology* **147**, 1237–1246, <https://doi.org/10.1007/s00705-001-0799-6> (2002).
- Perera, M. F. *et al.* Genetic Diversity Among Viruses Associated with Sugarcane Mosaic Disease in Tucumán, Argentina. *Phytopathology* **99**, 38–49 (2009).

20. Isabirye, B. E. & Rwomushana, I. Current and future potential distribution of maize chlorotic mottle virus and risk of maize lethal necrosis disease in Africa. *Journal of crop protection* **5**, 215–228 (2016).
21. Scheets, K. Maize Chlorotic Mottle Machlomovirus Expresses Its Coat Protein from a 1.47-kb Subgenomic RNA and Makes a 0.34-kb Subgenomic RNA. *Virology* **267**, 90–101, <https://doi.org/10.1006/viro.1999.0107> (2000).
22. Scheets, K. Analysis of gene functions in Maize chlorotic mottle virus. *Virus Research*, <https://doi.org/10.1016/j.virusres.2016.04.024> (2016).
23. Firth, A. E., Wills, N. M., Gesteland, R. F. & Atkins, J. F. Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic acids research* **39**, 6679–91, <https://doi.org/10.1093/nar/gkr224> (2011).
24. García-Arenal, F., Fraile, A. & Malpica, J. M. Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology* **39**, 157–186 (2001).
25. Varanda, C. M. R., Nolasco, G., Clara, M. I. & Félix, M. R. Genetic diversity of the coat protein of olive latent virus 1 isolates. *Archives of Virology* **159**, 1351–1357, <https://doi.org/10.1007/s00705-013-1953-7> (2014).
26. Varanda, C. M. R. *et al.* Genetic diversity of the coat protein of olive mild mosaic virus (OMMV) and tobacco necrosis virus D (TNV-D) isolates and its structural implications. *PLoS ONE* **9** <https://doi.org/10.1371/journal.pone.0110941> (2014).
27. Pybus, O. G., Fraser, C. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics* **10**, 540–550, <https://doi.org/10.1038/nrg2583> (2009).
28. Wangai, A. W. *et al.* First Report of Maize chlorotic mottle virus and Maize Lethal Necrosis in Kenya. *Plant Disease* **96**, 1582–1582, <https://doi.org/10.1094/PDIS-06-12-0576-PDN> (2012).
29. Deng, T.-C., Chou, C.-M., Chen, C.-T., Tsai, C.-H. & Lin, F.-C. First Report of Maize chlorotic mottle virus on Sweet Corn in Taiwan. *Plant Disease* **98**, 1748–1748, <https://doi.org/10.1094/PDIS-06-14-0568-PDN> (2014).
30. Liu, S., He, X., Park, G., Josefsson, C. & Perry, K. A Conserved Capsid Protein Surface Domain of Cucumber Mosaic Virus Is Essential for Efficient Aphid Vector Transmission A Conserved Capsid Protein Surface Domain of Cucumber Mosaic Virus Is Essential for Efficient Aphid Vector Transmission. *Virology* **76**, 9756–9762, <https://doi.org/10.1128/JVI.76.19.9756> (2002).
31. Schellenberger, P. *et al.* A stretch of 11 amino acids in the betaB-betaC loop of the coat protein of grapevine fanleaf virus is essential for transmission by the nematode *Xiphinema* index. *Journal of virology* **84**, 7924–7933, <https://doi.org/10.1128/JVI.00757-10> (2010).
32. Zwart, M. P., Willemsen, A., Daròs, J. A. & Elena, S. F. Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Molecular Biology and Evolution* **31**, 121–134, <https://doi.org/10.1093/molbev/mst175> (2014).
33. Moreno, I. M., Malpica, J. M., Rodríguez-Cerezo, E. & García-Arenal, F. A mutation in tomato aspermy cucumovirus that abolishes cell-to-cell movement is maintained to high levels in the viral RNA population by complementation. *Journal of virology* **71**, 9157–9162 (1997).
34. Hudson, R. R., Boos, D. D. & Kaplan, N. L. A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138–151 (1992).
35. Hudson, R. R. A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014 (2000).
36. Lynch, M. & Crease, T. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7**, 377–394 (1990).
37. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 DOI PMC1205159 (1992).
38. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* **8** <https://doi.org/10.1371/journal.pgen.1002764> (2012).
39. Boulila, M. Positive selection, molecular recombination structure and phylogenetic reconstruction of members of the family Tombusviridae: Implication in virus taxonomy. *Genetics and Molecular Biology* **34**, 647–660, <https://doi.org/10.1590/S1415-47572011005000046> (2011).
40. Wang, C.-Y. *et al.* Uncoating Mechanism of Carnation Mottle Virus Revealed by Cryo-EM Single Particle Analysis. *Nature Publishing Group* **5**, 3–9, <https://doi.org/10.1038/srep14825> (2015).
41. Newburn, L. R. & White, K. A. Cis-acting RNA elements in positive-strand RNA plant virus genomes. *Virology* **479–480**, 434–43, <https://doi.org/10.1016/j.virol.2015.02.032> (2015).
42. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads (2011).
43. Krueger, F. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files (2015).
44. Watson, S. J. *et al.* Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120205, <https://doi.org/10.1098/rstb.2012.0205> (2013).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, <https://doi.org/10.1038/nmeth.1923> (2012).
46. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644–52, <https://doi.org/10.1038/nbt.1883> (2011).
47. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)* **27**, 2987–93, <https://doi.org/10.1093/bioinformatics/btr509> (2011).
48. Stenger, D. C. & French, R. Complete nucleotide sequence of a maize chlorotic mottle virus isolate from Nebraska. *Archives of Virology* **153**, 995–997, <https://doi.org/10.1007/s00705-008-0069-y> (2008).
49. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 <https://doi.org/10.1093/bioinformatics/btg430> (2004).
50. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**, 2725–9, <https://doi.org/10.1093/molbev/mst197> (2013).
51. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254–267, <https://doi.org/10.1093/molbev/msj030> (2006).
52. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**, 1–5, <https://doi.org/10.1093/ve/vev003> (2015).
53. Ronquist, F. *et al.* MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539–542, <https://doi.org/10.1093/sysbio/sys029> (2012).
54. Librado, P. & Rozas, J. DnaSPv5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452, <https://doi.org/10.1093/bioinformatics/btp187> (2009).
55. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (2009).
56. Fraile, A., Malpica, J. M., Aranda, M. A., Rodríguez-Cerezo, E. & García-Arenal, F. Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant *Nicotiana glauca*. *Virology* **223**, 148–55, <https://doi.org/10.1006/viro.1996.0463> (1996).
57. Rubio, L. *et al.* Genetic variation of Citrus tristeza virus isolates from California and Spain: evidence for mixed infections and recombination. *Journal of virology* **75**, 8054–62, <https://doi.org/10.1128/JVI.75.17.8054> (2001).
58. Rubio, L., Abou-Jawdah, Y., Lin, H. X. & Falk, B. W. Geographically distant isolates of the crinivirus Cucurbit yellow stunting disorder virus show very low genetic diversity in the coat protein gene. *Journal of General Virology* **82**, 929–933 (2001).



59. Pinel, A., N'Guessan, P., Bousalem, M. & Fargette, D. Molecular variability of geographically distinct isolates of Rice yellow mottle virus in Africa. *Archives of Virology* **145**, 1621–1638, <https://doi.org/10.1007/s007050070080> (2000).
60. Azzam, O., Yambao, M. L. M., Muhsin, M., McNally, K. L. & Umadhay, K. M. L. Genetic diversity of rice tungro spherical virus in tungro-endemic provinces of the Philippines and Indonesia. *Archives of Virology* **145**, 1183–1197, <https://doi.org/10.1007/s007050070118> (2000).
61. Fraile, A. & García-Arenal, F. Secondary structure as a constraint on the evolution of a plant viral satellite RNA. *Journal of Molecular Biology* **221**, 1065–1069, [https://doi.org/10.1016/0022-2836\(91\)90916-T](https://doi.org/10.1016/0022-2836(91)90916-T) (1991).

## Acknowledgements

The authors would like to acknowledge colleagues at KALRO for their assistance in maize sampling, to farmers in all study areas for providing maize samples, to John Welch for guidance on phylogenetic analyses, and to Shizu Watanabe for her work on purifying MCMV in Hawaii. L.B. is supported by the BBSRC DTP, and D.C.B. is supported by the Royal Society Edward Penley Abraham Research Professorship.

## Author Contributions

L.B. designed and performed experiments, analysed the results and wrote the manuscript, D.Q.A., D.C., and A.B. performed experiments, A.W. and D.C.B. designed experiments. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-19607-4>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018