JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Silva *et al.*, *Journal of General Virology* 2022;103:001806
DOI 10.1099/jgv.0.001806

MICROBIOLOGY SOCIETY

Virus classification based on in-depth sequence analyses and development of demarcation criteria using the *Betaflexiviridae* as a case study

João Marcos Fagundes Silva[1], Fernando Lucas Melo[2], Santiago F. Elena[3,4], Thierry Candresse[5], Sead Sabanadzovic[6], Ioannis E. Tzanetakis[7], Arnaud G. Blouin[8], Dan Edward Veloso Villamor[7], Dimitre Mollov[9], Fiona Constable[10], Mengji Cao[11], Pasquale Saldarelli[12], Won Kyong Cho[13] and Tatsuya Nagata[1,*]

Abstract

Currently, many viruses are classified based on their genome organization and nucleotide/amino acid sequence identities of their capsid and replication-associated proteins. Although biological traits such as vector specificities and host range are also considered, this later information is scarce for the majority of recently identified viruses, characterized only from genomic sequences. Accordingly, genomic sequences and derived information are being frequently used as the major, if not only, criteria for virus classification and this calls for a full review of the process. Herein, we critically addressed current issues concerning classification of viruses in the family *Betaflexiviridae* in the era of high-throughput sequencing and propose an updated set of demarcation criteria based on a process involving pairwise identity analyses and phylogenetics. The proposed framework has been designed to solve the majority of current conundrums in taxonomy and to facilitate future virus classification. Finally, the analyses performed herein, alongside the proposed approaches, could be used as a blueprint for virus classification at-large.

INTRODUCTION

Virologists have traditionally relied on comparisons of multiple virus properties for their classification. These include morphological, biological, serological, and epidemiological traits. However, with the advances of genome sequencing procedures and, in particular, high-throughput sequencing (HTS) technologies, virologists now rely primarily on genomic information to classify novel viruses [1]. A species is defined as 'a monophyletic group of mobile genetic elements (MGEs) whose properties can be distinguished from those of other species by multiple criteria' (https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code), and thus it may be defined by a combination of properties derived from genomic sequences.

Received 05 July 2022; Accepted 13 October 2022; Published 18 November 2022
Author affiliations: [1]Departamento de Biologia Celular, Instituto de Ciências Biológicas, University of Brasília, Brasília 70910-900, Brazil; [2]Departamento de Fitopatologia, Instituto de Biología Integrativa de Sistemas, University of Brasília, Brasília 70910-900, Brazil; [3]Instituto de Biología Integrativa de Sistemas (I2 13 SysBio), CSIC-Universitat de València, Paterna 14 46980 València, Spain; [4]The Santa Fe Institute, Santa Fe, NM 87501, USA; [5]Univ. Bordeaux, INRAE, UMR 1332 BFP, 33140 Villenave d'Ornon, France; [6]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA; [7]Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR 72701, USA; [8]Virology-Phytoplasmology Laboratory, Agroscope, 1260 Nyon, Switzerland; [9]USDA-ARS Horticultural Crops Disease and Pest Management Research Unit, Corvallis, OR, 97330, USA; [10]Department of Jobs Precincts and Regions, Agriculture Victoria Research, Agribio, Bundoora, VIC 3083, Australia; [11]National Citrus Engineering and Technology Research Center, Citrus Research Institute, Southwest University, Beibei, Chongqing 400712, PR China; [12]National Research Council of Italy (CNR), Institute for Sustainable Plant Protection (IPSP), Via Amendola 122/D, 70126 Bari, Italy; [13]College of Biotechnology and Bioengineering, Sungkyunkwan University, Seoburo 2066, Suwon 16419, Gyeonggi, Republic of Korea.
***Correspondence:** Tatsuya Nagata, tatsuya@unb.br
Keywords: betaflexivirus; demarcation criteria; phylogenetics; virus taxonomy.
Abbreviations: ACLSV, apple chlorotic leaf spot virus; AICc, corrected Akaike information criterion; ApLV, apricot latent virus; APV1, Asian prunus virus 1; APV2, Asian prunus virus 2; ASPV, apple stem-pitting virus; AU, approximately unbiased; ButMV, butterbur mosaic virus; CLBV, citrus leaf blotch virus; CMLV, cherry mottle leaf virus; CNRMV, cherry necrotic rusty mottle virus; CP, coat protein; CpMMV, cowpea mild mosaic virus; CRMaV, cherry rusty mottle-associated virus; CTLaV, cherry twisted leaf-associated virus; CVB, chrysanthemum virus B; FN, false negative; FP, false positive; FPR, false positive rate; GBINV, grapevine berry inner necrosis virus; GVA, grapevine virus A; GVD, grapevine virus D; GVE, grapevine virus E; GVH, grapevine virus H; GVJ, grapevine virus J; GVM, grapevine virus M; Hel, helicase; KDE, kernel density estimation; LVA, Ligustrum virus A; Met, methyltransferase; NLV, nerine latent virus; PMV, peach mosaic virus; PTI, pattern-triggered immunity; PVH, potato virus H; PVM, potato virus M; RdRP, RNA-dependent RNA polymerase; Rep, replication protein; SPCFV, sweet potato chlorotic fleck virus; TGB, triple gene block; TN, true negative; TP, true positive; TPR, true positive rate.
A supplementary figure is available with the online version of this article.

RNA viruses that use a RNA-dependent RNA polymerase (RdRP) for replication are classified in the realm *Riboviria* [2]. Although the RdRP is the defining feature of this broad and highly diverse taxon, species within the lower ranks are demarcated using pairwise comparisons of various genomic characteristics including whole genome/entire coding regions or pairwise comparisons of one or more genes or their products. However, many viruses lay at the borderline of species demarcation or are products of recombination; challenging the currently defined demarcation criteria and a streamlined taxonomic assignation process is needed.

One of the most diverse positive strand RNA virus ensembles is the family *Betaflexiviridae* (the order *Tymovirales*). Members have monopartite, polyadenylated RNA genomes and form flexuous filamentous particles. They encode an alpha-like replication protein (Rep) with methyltransferase (Met), helicase (Hel), and RdRP motifs [3]. The family is currently composed of 15 genera divided between two subfamilies; members of the subfamily *Quinvirinae* have a triple gene block (TGB) module that facilitates cell-to-cell movement [4] whereas viruses in the subfamily *Trivirinae* encode a 30K-like movement protein [5].

According to the current species demarcation criteria in the family *Betaflexiviridae*, viruses with sequences that have less than 72% nucleotide (nt) or 80% amino acid (aa) identity in the coat protein (CP) or Rep genes are considered members of distinct species. Betaflexviruses are classified in different genera if their CP and Rep sequences have less than 45% nucleotide identity [3]. This approach was effective at the time it was proposed. However, the increasing pace of virus discovery has led to many classification challenges [6–10]. For example, it is unclear which of the two genes/proteins should have precedence for taxonomic demarcation, so that it is difficult to reach an unambiguous conclusion in situations where one of the genes/proteins shows identity values below the species threshold whereas the reverse is observed for the other. The use of double thresholds, using both nt and aa sequences, for each of the two genes/proteins further complicates issues [6–10]. To avoid these problems, the criteria used in some families, like the family *Secoviridae,* take into consideration only the protein sequences and prioritize between the distinct protein domains considered, namely the Pro-Pol and CP [11].

In order to determine new viruses' taxonomy and streamline the process, we used the family *Betaflexiviridae* as a case study to assess and revise the sequence-based species discrimination criteria. Re-analysis of pairwise genetic distances of GenBank-available isolates in the family indicate that the Rep and CP genes/proteins do not diverge at the same rate, and therefore different thresholds should be used for the two proteins. We further propose to take into account these novel elements within a modified decision framework, aiming to limit the ambiguities emerging from the current system as a consequence of the absence of prioritization between proteins.

METHODS

Data collection

All GenBank accessions available in May 2021 that contained complete Rep and CP sequences from the family *Betaflexiviridae* members were retrieved and used for analyses. Taxonomic assignation for each accession was annotated based on currently accepted species [12]. For accessions that correspond to not yet recognized species, taxonomic assignation was annotated based on relevant publications [6, 10, 13–38], GenBank/EMBL/DDBJ description and BLAST [39] analyses. Complete Rep and CP aa sequences were extracted using Geneious R8.1.9. Prior to phylogenetic and pairwise identity analyses, both Rep and CP sequences were clustered using CD-HIT [40] such that only a representative sequence was included where sequences shared 100% aa identity. This resulted in two datasets, one for each protein, with 1230 sequences each.

Phylogenetic, recombination and pairwise identity analyses

Alignments of the aa sequences of the Rep and CP were performed with MAFFT v7.110 [41]. The N- and C-terminal portions of the alignments, often misaligned, as well as columns with more than 50% gap were removed with CIAlign [42], resulting in two alignments containing 1976 (Rep) and 435 (CP) sites, respectively. Phylogenetic inference was then conducted with FastTree v2.1.11 [43]. A tanglegram showing the position of the same virus on Rep and CP-based trees was constructed to investigate phylogenetic incongruences between the two proteins. Intragenus recombinant sequences were detected using the concatenated Rep and CP sequences with the RDP5 command line tool [44]. Available methods for recombination detection were used, with BootScan and SiScan in primary scan mode. For the genus *Carlavirus*, the step size of the BootScan method was increased to 70 to decrease run time. This change should not significantly impact the results as the step size was less than 50% of the window size of 200, which is recommended by the user's manual. All other parameters were run at the default settings. Recombination events supported by at least five methods (adjusted $P<0.05$) were considered as true positives. For the analyses of sequences for which a reassignment was sought, the relevant clade was extracted from the Rep tree with the R package treeio v1.15.7 [45].

The Rep and CP datasets of 1230 accessions were then used for protein identity analyses. Pairwise identities of complete Rep and CP sequences were obtained with SDT v1.2 [46] using MUSCLE [47]. The density of the data in the Rep and CP identity space was calculated by kernel density estimation (KDE) at each taxonomic level separately. To better visualize density estimates, they were divided by the maximum KDE value for each taxonomic level so that the highest density value at each level equals one. Species that exhibited a great disparity between the Rep and CP aa identities between non-recombinant isolates and selected species

composed solely of recombinants were subjected to further recombination analysis with GARD [48], where the relative likelihood of the corrected Akaike information criterion (AICc) scores of the single tree model was compared to the value of the tree with the most multiple trees (AICc min) with the formula $P=e^{(\text{AICc min - AICc single tree})/2}$. The AICc score estimates how well the model fits to the data. By comparing the AICc scores of the single tree and the most multiple trees models, it is possible to evaluate if the later model, where at least two partitions of the alignment have different tree topologies, fits significantly better to the data. For this analysis, in-frame Rep and CP alignments were uploaded to the datamonkey.org server (https://www.datamonkey.org/) and GARD was run with default settings. Additional analyses were also performed for species composed entirely of interspecies recombinants, i.e., species where all isolates are descendants of an interspecies recombinant virus (given that RDP5 was run at the genus level, some intergenus recombination events were not detected. If all members of a genus are descendants of an intergenus recombinant virus, it is not detectable). In these cases, an alignment of concatenated Rep and CP sequences was used as input to investigate intraspecific recombination events. Downstream analyses were performed using R 3.6.3 with the treeio, ggtree v2.0.4 [49], ape v5.5 [50] and phytools 0.7–70 [51] packages. Tree topology tests were performed with IQ-TREE [52].

Accuracy of taxonomic criteria analysis

The accuracy of the proposed criteria with varying CP thresholds was investigated for species chosen due to their high number of available isolates or because they would be affected by the choice of different CP thresholds. For each species, true positives (TP) and false negatives (FN) were determined based on the ability of the applied criteria to maintain the current species classification, whereas true negatives (TN) and false positives (FP) were determined based on the capability of the applied criteria to distinguish randomly selected isolates from other species. This analysis was also performed with R programming language.

RESULTS AND DISCUSSION

The Rep and CP exhibit different evolutionary histories

Phylogenetic trees were constructed for the Rep and CP aa sequences (Fig. 1). Two major groups corresponding to subfamilies *Trivirinae* and *Quinvirinae* are present in the Rep tree. However, the CP tree of the *Quinvirinae* also included viruses from the genera *Citrivirus* and *Wamavirus*, part of the *Trivirinae*, indicating that these genera may have arisen through recombination.

Recombination is known to be a major driving force in the evolution of RNA viruses [53], a phenomenon that is well documented within the family *Betaflexiviridae* [7, 17, 22, 54–59]. To visualize incongruences between the Rep- and CP-based phylogenies, both trees were represented as a tanglegram (Fig. 1). Some sequences did not maintain the same position in the Rep and CP trees, confirming that recombination likely occurred among family members. The positioning of the aforementioned genera as part of the subfamily *Quinvirinae* in the CP tree is well supported (Fig. 1). To test whether the CP tree topology is accurate, unconstrained and Rep-constrained CP trees were compared by performing the approximately unbiased (AU) tests [60] with IQ-TREE. Here, the Rep tree topology was used to construct a Rep-constrained tree using the CP alignment. Then, the AU test calculated the confidence of this tree as well as the confidence of an unconstrained tree. The Rep-constrained CP tree was rejected (P=0), thus confirming the Rep and CP have distinct evolutionary histories. Given that the CP-based phylogeny does not clearly separate the two subfamilies, this analysis provided an argument to support the notion that Rep should be chosen as the primary molecular demarcation criterion for the family.

Non-recombinant sequences exhibit a narrower pairwise distribution range for the CP

Given the discrepancies between the Rep and CP phylogenies, we hypothesized that recombination has a significant impact on pairwise identities distributions. Therefore, the Rep and CP aa identities between each sequence pair were determined for both datasets (Fig. 2a) and additionally for datasets composed only of sequences from non-recombinant isolates (Fig. 2b). At the species level, pairwise distributions between non-recombinant sequences were concentrated at identity levels above 80% for the Rep and above ~90% for the CP (Fig. 2b). The distances of each pairwise comparison to the centroid point of the Rep and CP aa identities space were significantly smaller at the species level (Wilcoxon rank sum test; P=0.003), confirming that pairwise identities between non-recombinant sequences are more tightly clustered. Additionally, the correlation between the Rep and CP aa identity was higher for non-recombinant sequences at both genus and species levels. These results indicate that recombination has a significant impact on pairwise identity distributions.

Pairwise identities of the Rep and CP do not follow the same distribution

In addition to incongruent phylogenetic topologies, the distributions of pairwise aa identities of the Rep and CP follow different patterns. At the lowest identity values (<50%), the Rep distribution showed two well-defined peaks at 33 and 41%, respectively; whereas for the CP, the lowest identity values were more extensive and had less defined peaks at 20 and 28% but with a lower number of pairwise comparisons falling at these percentages (Fig. 3a and Fig. 3b). On the other hand, at the higher identity values within the currently accepted species boundary (>80%), the Rep distribution showed two peaks at 85 and 98%; whereas the CP peaks were at 93 and 99% (Fig. 3a and Fig. 3b). These results indicate that at the species level the CP is more conserved than the Rep among currently

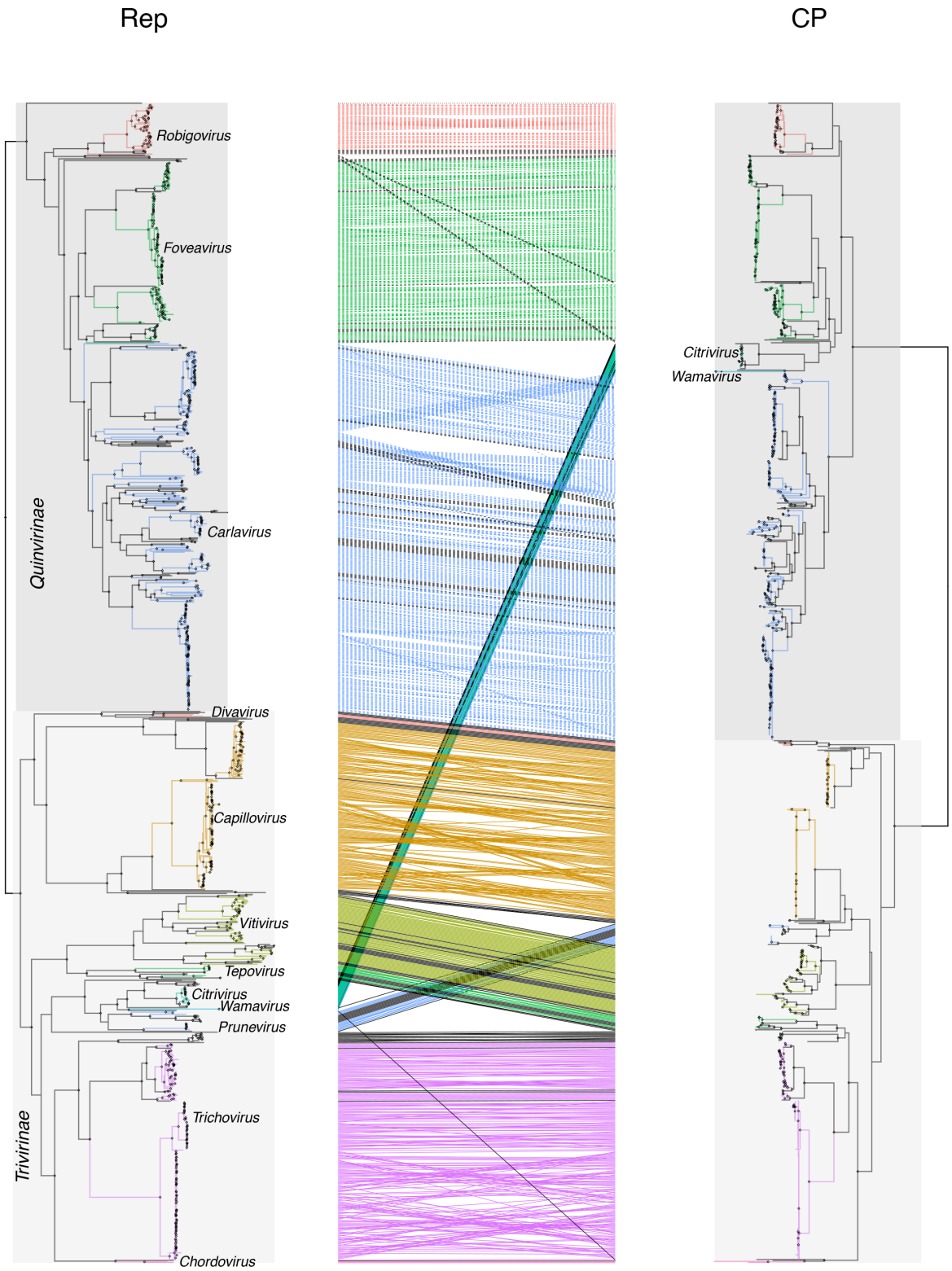

Fig. 1. Tanglegram of the Rep and CP phylogenies of the family *Betaflexiviridae*. Trees were constructed by approximately maximum-likelihood (AML) using FastTree with alignments of the aa sequences of the Rep and CP from 1230 sequences. The Rep tree was rooted using *Botrytis virus F* (BotVF; AF238884; *Deltaflexiviridae*) as an outgroup, whereas the CP tree was midpoint-rooted. Support values above 70% are represented by diamonds. The subfamilies *Quinvirinae* and *Trivirinae* are highlighted with dark and light grey rectangles, respectively.

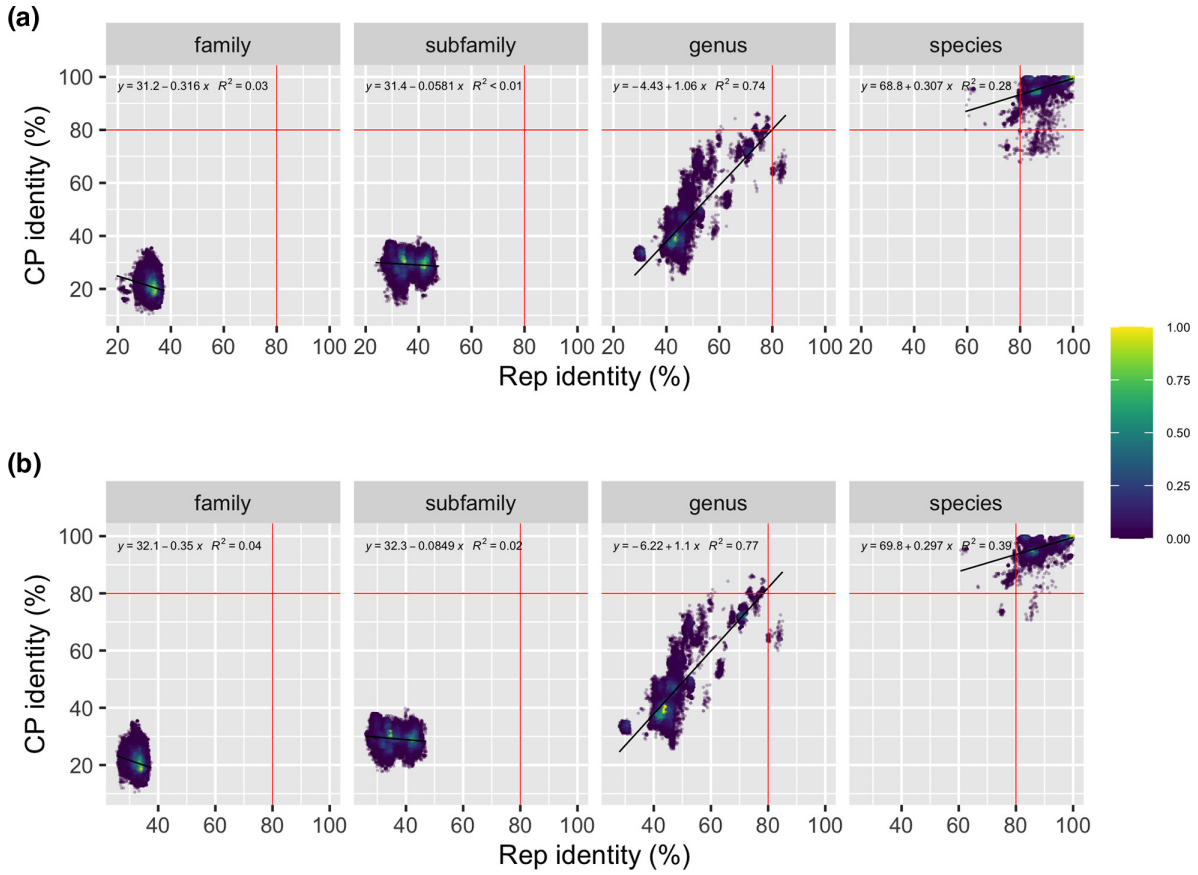

Fig. 2. Effects of recombination on pairwise identities distributions. Rep and CP aa identities among GenBank/EMBL/DDBJ accessions are represented as dots. Warm and cold colours represent high and low density of data, respectively. Linear regressions were conducted for each panel to investigate the relationship between the identities of the two proteins. Red lines represent the current identity thresholds (<80% amino acid identity for Rep and CP) for species demarcation in the family. (a) Dot plot generated using all available sequences from the family *Betaflexiviridae*; (b) pairwise comparisons between non-recombinant sequences.

known isolates (Fig. 3c). Linear regression and Pearson's correlation analyses of the Rep and CP pairwise identities indicate a higher conservation of the CP at the species level; whereas at the genus level, an almost linear relationship between the pairwise identities of the two proteins was found, leading to an R^2 of 0.77 (Fig. 2b). The high conservation of the CP at the species level may be related to evolutionary pressures including vector specificities, novel host adaptations or other possible functions, such as suppression of pattern-triggered immunity (PTI) [61]. These differences may also be at least partially attributed to undetected recombination events. The distribution of three types of comparisons (family, subfamily and genus) shows clear overlaps and are, therefore, not well separated.

Based on the different evolutionary rates and pairwise identities distributions of the Rep and CP, we conclude that recombination among members of the family *Betaflexiviridae* complicates the simultaneous use of Rep and CP parameters. Given that the evolution of RNA viruses is better understood based on the evolution of their polymerase, the only domain present in all members of the *Riboviria* [53], we propose that the Rep-based criterion should be given precedence for species demarcation. Secondly, at the species level, the CP is more conserved than the Rep (Fig. 3, insets in panels a vs b), and thus, the same threshold of 80% aa identity for both proteins is not justified. If the Rep aa identity criterion is used alone, it should be possible to resolve many of the controversies arising from application of currently valid criteria. However, species demarcation is not clear when there are pairwise comparisons with Rep aa identities near the 80% threshold (at the 78–82% range). In those cases, a second (auxiliary) criterion could be used to provide a more robust taxonomic separation. We therefore propose that the Rep aa identity should be given priority over the CP aa identity (primary criterion), and that a more stringent threshold for the CP should be used as an auxiliary tool when Rep aa identities are in the 78–82% borderline range. In synthesis, a threshold for Rep of 80% aa identity is maintained, whereas a new, 85% cut-off is proposed for the CP, as determined below.

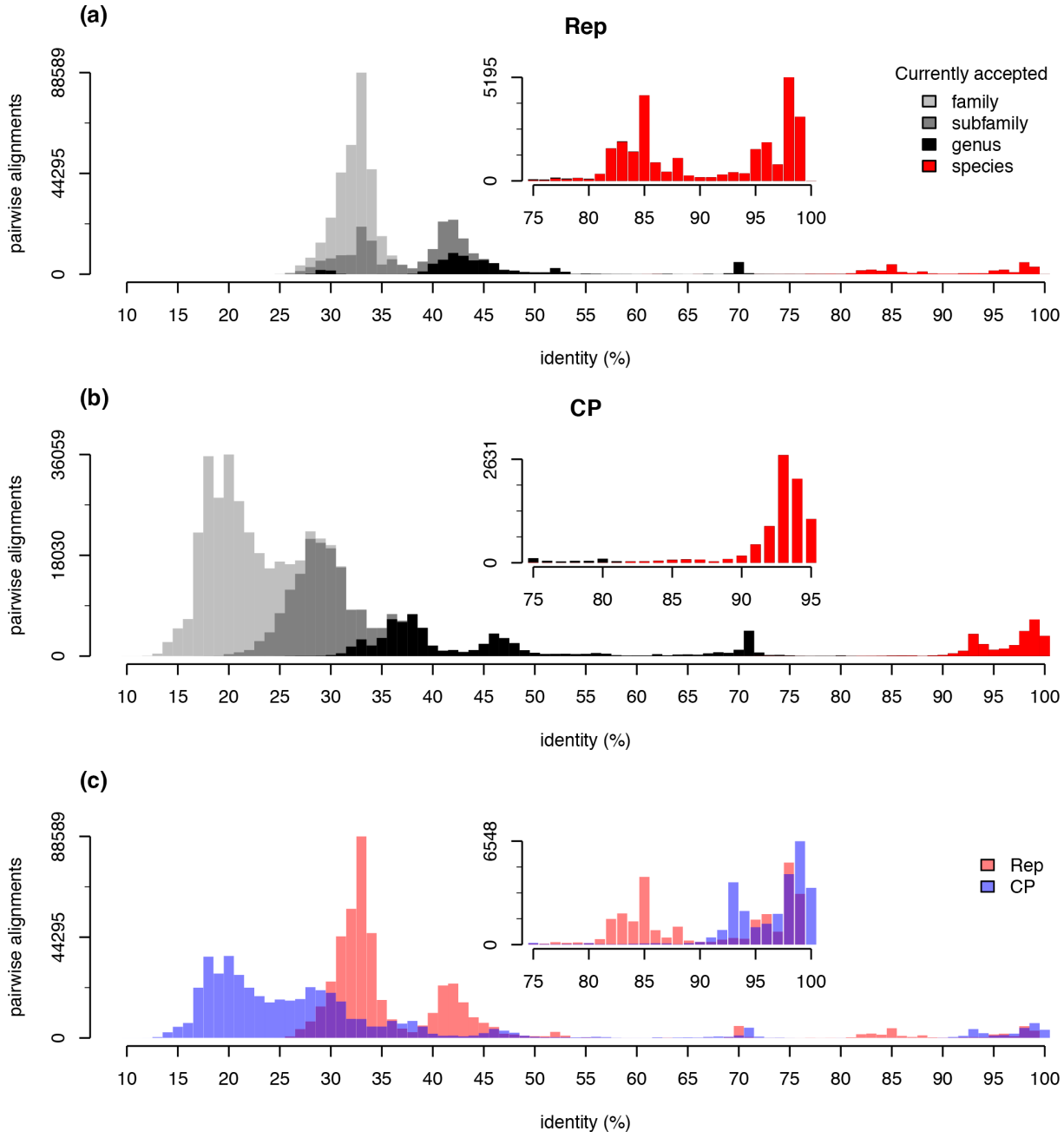

Fig. 3. Distributions of the pairwise amino acid sequence identities of the Rep (a), CP (b) and overlap of Rep and CP (c). Zoomed-in graphs of identities above 75% are shown for better visualization of within-species identities.

Impact of proposed demarcation on currently recognized species

We sought to determine the performance of the current identity criteria in recapturing accurate species demarcation. In other words, we were interested to learn whether there are conflicts between the identities criteria and biological properties for species demarcation purposes. We sought to determine a new threshold for the CP aa identities that can recapture adequate species demarcation while causing minimum changes to the currently accepted taxonomy. Therefore, the accuracy of the current (80% aa identity) and proposed criteria (varying CP aa identity thresholds) were investigated for members of selected species. These species were chosen due to their high number of available isolates in the databases, or because they would potentially be affected by a modification of the CP aa identity threshold. For each species, a positive (P) was counted when two members were classified in the same species, and a negative (N) was counted when two members were classified in different species. True positives (TP) and false negatives (FN) were determined based on the ability of the applied criteria to maintain isolates within their current

species, whereas true negatives (TN) and false positives (FP) were determined based on the capability of the applied criteria to separate isolates from distinct species that were randomly selected. TP and TN were used to calculate accuracy as (TP +TN)/(P+N). True positive rate (TPR=TP/P) and false positive rate (FPR=FN/N) were also calculated.

For the robigoviruses, i.e., cherry necrotic rusty mottle virus (CNRMV), cherry rusty mottle-associated virus (CRMaV) and cherry twisted leaf-associated virus (CTLaV), we investigated the ability of the CP-based criterion to accurately separate isolates between species, and as such, TN and FP were calculated based on randomly selected interspecies comparisons of these viruses only. CNRMV, CRMaV and CTLaV isolates can be distinguished based on their symptomatology [54, 62]; however, based on the current criteria, they have a controversial/ambiguous taxonomic status. These viruses could be considered as members of the same species (due to their CP aa identities being >80%), or could be classified in distinct species (based on the aa identities of their Rep <80%). Discrepancy between molecular identity criteria and distinct biological properties for these robigoviruses has been previously noticed [54].

The strategy to calculate TN and FP was also applied to cherry mottle leaf virus (CMLV) and peach mosaic virus (PMV), closely related species in the genus *Trichovirus* that infect different hosts and have Rep aa identities <80% and CP aa identities >80%. We found that the 85% CP aa identity threshold provided the best cost/benefit in maintaining the species demarcation while distinguishing between these various robigoviruses and trichoviruses (Fig. 4a). However, using an 85% threshold would also result in currently known CRMaV isolates being split and classified into two species, so that an 82% CP threshold was optimal for the distinction of CRMaV, CNRMV and CTLaV isolates and keeping their classification in distinct species. These results suggest that a flexible CP threshold to achieve optimal results would be needed to settle ambiguous cases using an accuracy statistics analysis. On the downside, this accuracy would be best calculated when bona fide species assignment can be determined with the aid of biological properties, which is currently not available for many viruses and is unlikely to become available in the near future.

We also compared the current CP threshold (80% identity) with the proposed one (85%) under different scenarios (Fig. 4b). A rise in accuracy and a drop in FPR was observed for CRMaV, CNRMV, CTLaV, CMLV and PMV. A drop in TPR was noted for potato virus M (PVM), grapevine berry inner necrosis virus (GBINV) and CRMaV, indicating that some isolates currently regarded as belonging to these species have CP sequence identities that are below the proposed 85% species demarcation threshold.

Implementation of the proposed taxonomic criteria in the family *Betaflexiviridae*

When analysing pairwise comparisons under the current and the proposed species demarcation thresholds, six scenarios arise (Fig. 5). Area I (Rep aa identity >80%; CP aa identity >85%) contains bona fide within-species comparisons. Area II (Rep aa identity <80%; CP aa identity >85%) contains both intrageneric and intraspecific comparisons, in which the identity of the Rep in some cases is much lower than the threshold. If they meet the Rep monophyly criterion, viruses with comparisons results falling within this area should be split into separate species under a proposed new rule giving priority to the Rep aa identity criterion. Area III (Rep aa identity >80%; 80%<CP aa identity <85%) contains within-species comparisons in which the identities of the CP are lower than the proposed new threshold (85%), but since the identities of the Rep are above the threshold, these species remain unaltered. Area IV (Rep aa identity <80%; 80%<CP aa identity <85%) contains both within-genus and within-species comparisons. Species with comparison results between current members falling within this area should be split into two separate species according to the proposed Rep criterion priority strategy. Area V (Rep aa identity >80%; CP aa identity <80%) also contains both within-genus and within-species comparisons. For such within-genus comparisons above the 80% Rep aa identity threshold, an analysis is needed to assess whether these species should be merged. Lastly, viruses with comparisons falling within area VI (Rep aa identity <80%; CP aa identity <80%), with comparisons below the Rep and CP cut-offs, should belong to different species.

We applied the new proposed taxonomic criteria to the family *Betaflexiviridae* and performed a thorough analysis to identify potential consequences of the proposed criteria changes. Some specific interspecies comparisons that have their CP or Rep aa identities above the current thresholds were further investigated (Fig. 6). By giving preference to the Rep-based criterion, a clear separation is already achieved between isolates of CRMaV, CTLaV, and CNRMV, as well as between those of grapevine virus D (GVD) and grapevine virus J (GVJ). However, the separation between currently known isolates of CMLV and PMV, or that between Asian prunus viruses 1 and 2 (APV1 and APV2, respectively) remain ambiguous. In such situations, CP aa identities can then be used as a secondary criterion.

The CP pairwise aa identity between characterized GVD and GVJ isolates is above 84.2% while the Rep aa identity between these viruses is well below 80%, implying these viruses could be merged into one species based on the current criteria but should remain as distinct species based on the new proposed thresholds (Fig. 6). Many of the pairwise identities among CNRMV, CRMaV, and CTLaV isolates are above the current CP aa identity criterion of 80% and below the 80% Rep aa identity criterion (Fig. 6). The pairwise Rep aa identities between APV1 and APV2 are at the ~80% threshold, whereas their CP aa identities are lower than 80% (Fig. 6). The Rep aa identities between apricot latent virus (ApLV) and apple stem-pitting virus (ASPV) isolates are above 80% whereas their CP aa identities are below 80% (Fig. 6). However, ApLV and ASPV can be differentiated based on natural host range. More specifically, it appears that ApLV is naturally restricted to hosts in the *Prunoideae* (being graft transmissible not only to apricot, but also to peach, plum, cherry, and Japanese plum) [63], whereas ASPV naturally infects hosts in the *Maloideae*,

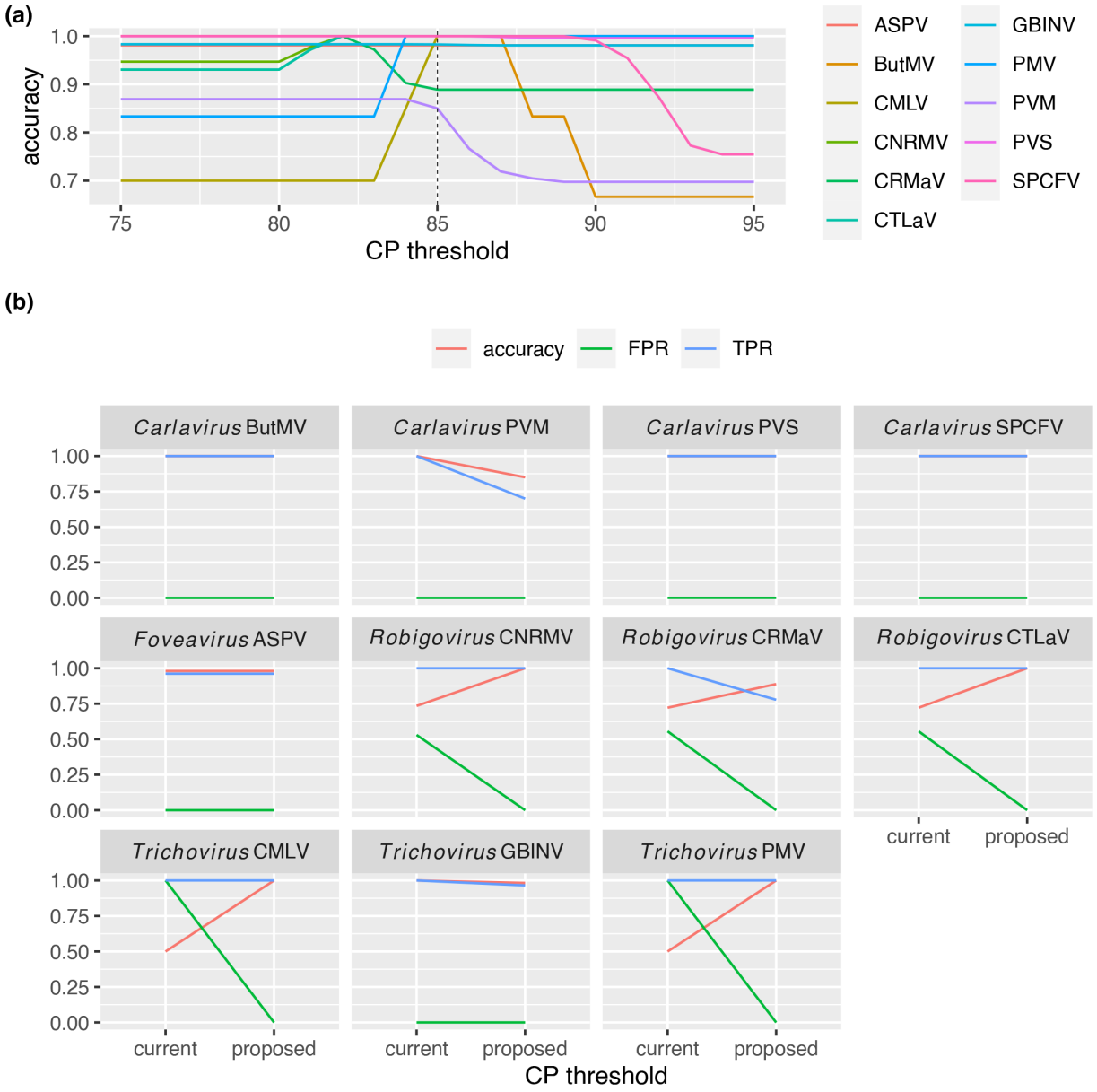

Fig. 4. Statistics describing the accuracy of the different criteria of CP aa identity for species validation. (a) Accuracy of the proposed criteria with varying CP thresholds for selected species. (b) Accuracy, false positive rate (FPR) and true positive rate (TPR) of the current (80%) and proposed (85%) criteria of CP threshold. Apple stem-pitting virus, ASPV; butterbur mosaic virus, ButMV; cherry mottle leaf virus, CMLV; cherry necrotic rusty mottle virus, CNRMV; cherry rusty mottle-associated virus, CRMaV; cherry twisted leaf-associated virus, CTLaV; grapevine berry inner necrosis virus, GBINV; peach mosaic virus, PMV; potato virus M, PVM; potato virus S. PVS; sweet potato chlorotic fleck virus, SPCFV.

including *Malus* spp., *Pyrus* spp. and *Cydonia* [64], plus a few other hosts in diverse families that do not include the *Prunoideae*. Thus, the 80% identity threshold of the Rep is not able to recapture their distinction. Further recombination analysis of sequences flagged as non-recombinants by the analyses using RDP5 were conducted using the GARD programme. Signals of recombination were found in the Rep and CP of APV1 and APV2 ($P<10^{-100}$ and $P<10^{-40}$, respectively) and of ASPV and ApLV ($P=0$ and $P<10^{-100}$, respectively). This indicates that some recombination events may have been missed by RDP5 (also by GARD), in cases of CP aa identities lower than expected. APV1 and APV2 should not be merged due to their lower-than-threshold CP aa identities, whereas ApLV and ASPV should not be merged due to different host range. The extent to which biological characteristics should be used to differentiate species may vary from case to case. For instance, different isolates of cucumber mosaic virus (CMV; family *Bromoviridae*) vary in their symptomatology and host range [65]. Here, we have two main considerations: viruses in the family *Betaflexiviridae* usually display quite narrow host ranges, and biological characteristics have already been used as species

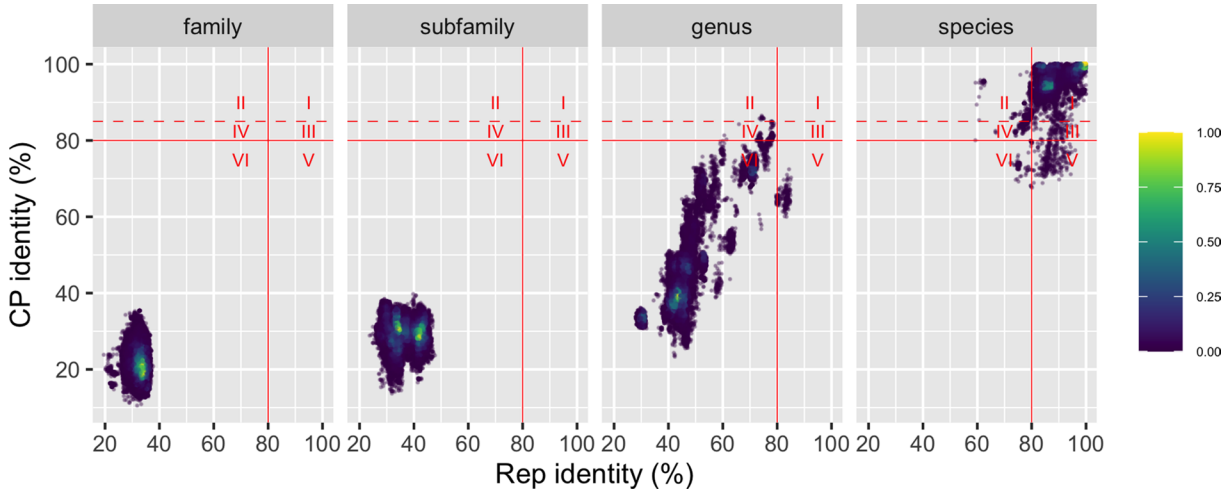

Fig. 5. Dot plot of Rep and CP aa identities of non-recombinant betaflexiviruses. Warm and cold colours represent high and low density of data, respectively. Red solid lines represent the current identity thresholds for the family and the red dashed line represents the proposed threshold for the CP of >85% identity.

demarcation criterion in the family *Betaflexiviridae* [3]. Thus, based on this criterion and to further avoid drastic changes to the current taxonomic classification, we propose that ApLV and ASPV should not be merged based on their distinct host ranges.

Next, we applied the proposed criteria to a subset of specific cases within areas II, IV and VI of Fig. 5 (Fig. 7). As a result, one isolate of apple chlorotic leaf spot virus (ACLSV; accession EU223295), with identities below the Rep and CP thresholds with other ACLSV isolates, should be reassigned as a member of the species peach chlorotic leaf spot virus (Fig. S1k). Similarly, one isolate of grapevine virus H (GVH) should be classified as an isolate of grapevine virus M (GVM) based on the priority given to the Rep-based criterion (Fig. S1m). The Rep pairwise identities between ASPV isolates span from below 80% to near 100%, whereas the corresponding CP aa identities are mostly below 85%. However, the Rep aa identities of isolates belonging to basal monophyletic groups in the ASPV phylogeny represented by blue and green colours (each colour representing a monophyletic group) also have identities well above the borderline range (>82%), thus, their current classification could remain unaltered (Fig. 7 and Fig. S1a, available in the online version of this article).

Two monophyletic groups of PVM isolates present pairwise identities at the lower borderline range for the Rep and around the proposed 85% CP aa identity threshold (Fig. 7 and Fig. S1b). One such group, represented by blue in the dot plot (Fig. 7) and phylogenetic tree (Fig. S1b), has Rep aa identities that extend below the lower borderline range. It could thus be argued that this group of isolates should be classified in a distinct species. However, the most external PVM monophyletic ramification represented solely by one isolate (accession: MH550835) (Fig. 7 and Fig. S1b) has its pairwise identities with the other isolates within the borderline range concerning the Rep, but above the proposed CP threshold of 85%. Thus, it cannot represent a new species. In this scenario, if the PVM isolates from a 'blue group' are assigned to a new species, the remaining PVM population would form a paraphyletic species, which is untenable. Therefore, reclassification of current single species embracing PVM isolates is ultimately not warranted.

Analyses of currently known data for CGRMV (Fig. S1j), cowpea mild mosaic virus (CpMMV) (Fig. S1q), CRMaV (Fig. S1l), chrysanthemum virus B (CVB) (Fig. S1i), grapevine virus A (GVA) (Fig. S1g), grapevine virus E (GVE) (Fig. S1h) and potato virus H (PVH) (Fig. S1n) isolates suggest that each of these species should be split, since they exhibit some Rep pairwise identities below 80% (in fact below 78%, Fig. 7) and form two well-supported monophyletic groups. CpMMV comparisons between sequences identified as non-recombinants by RDP5 that showed CP aa identities higher than expected (Fig. 7, sequences within rectangles) were subjected to further recombination analysis with GARD, and evidence of recombination within CpMMV was in fact detected for the Rep (P=0) and CP (P=3.42×10^{-12}).

Notably, CVB could be separated into distinct species that are composed entirely of recombinant isolates (Fig. 7), raising the question of the taxonomic status of recombinant isolates. Concatenated Rep and CP sequences of CVB were also subjected to further recombination analysis with GARD. Intraspecific recombination was also detected for CVB (P=0), which indicates that the lower-than-expected Rep aa identities of the green group is likely due to recombination (Fig. 7 and Fig. S1i).

Citrus leaf blotch virus (CLBV) (Fig. 7 and Fig. S1e) isolates could also be split into two species based on the identity criteria, however, assigning divergent isolates to a new species would again make the remaining CLBV isolates paraphyletic (Fig. S1e).

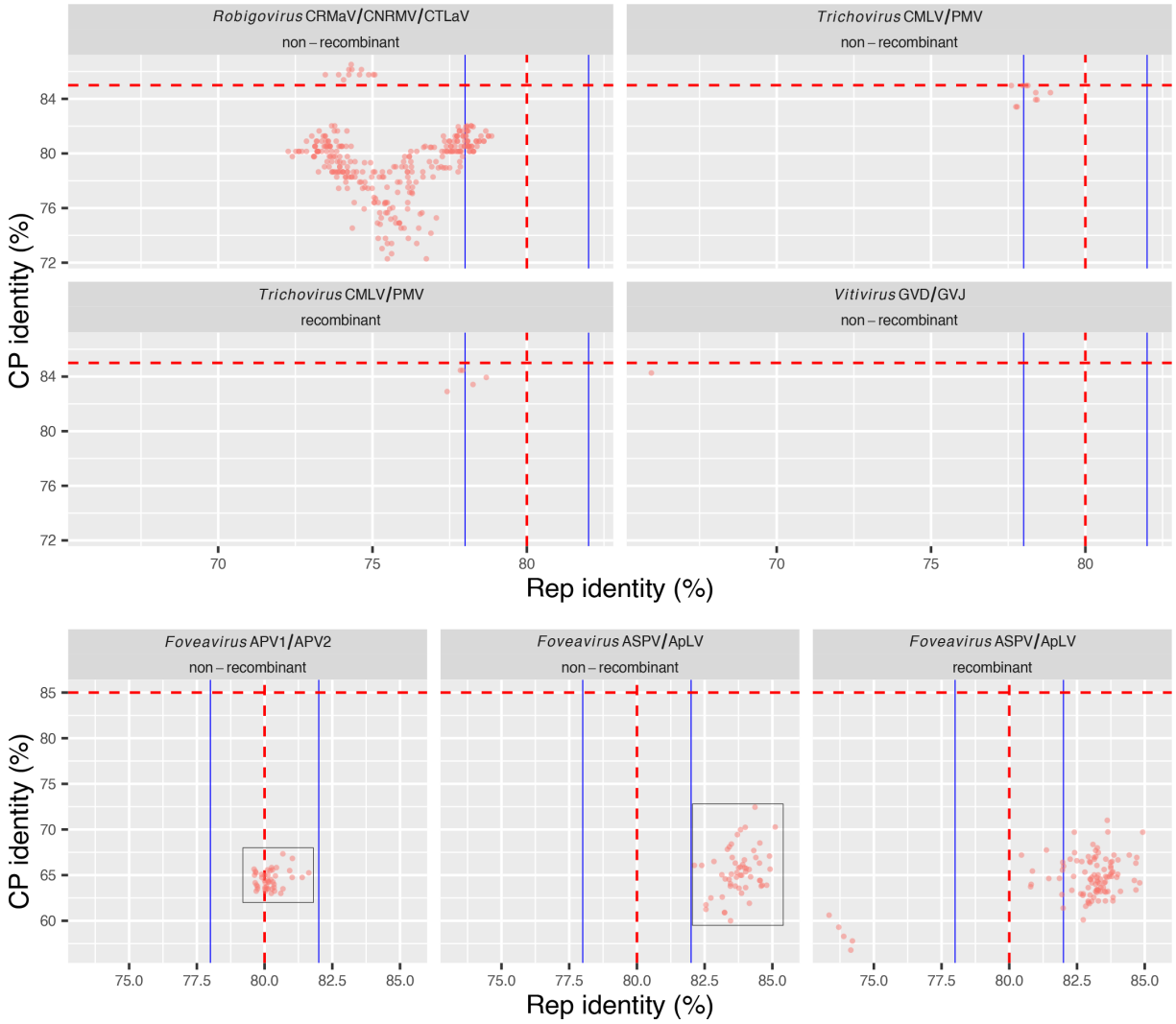

Fig. 6. Interspecies analyses of intragenus comparisons above the species threshold of the CP and Rep. Red dashed lines represent the proposed thresholds for the Rep and CP, and blue lines represent the borderline range of the Rep. Rectangles show cases that were further subjected to recombination analysis with GARD. Apricot latent virus, ApLV; Asian prunus virus 1, APV1; Asian prunus virus 2, APV2; apple stem-pitting virus, ASPV; cherry mottle leaf virus, CMLV; cherry necrotic rusty mottle virus, CNRMV; cherry rusty mottle-associated virus, CRMaV; cherry twisted leaf-associated virus, CTLaV; grapevine virus D, GVD; grapevine virus J, GVJ; peach mosaic virus, PMV.

GBINV isolates (Fig. 7) and (Fig. S1o) have some pairwise Rep aa identities slightly below 78% but all CP aa identities are above the 85% threshold. In this case a conservative approach would maintain this species to avoid changes to the current taxonomy. The remaining cases that were analysed, such as butterbur mosaic virus (ButMV), nerine latent virus (NLV) and sweet potato chlorotic fleck virus (SPCFV) would each remain as a single species (Fig. 7).

We analysed cases where the pairwise identities of the Rep are in the upper borderline range of 80–82%. CTLaV and Ligustrum virus A (LVA) sequences could be each classified into two species (Fig. 8) since they contain monophyletic groups with their Rep aa identities at the borderline range and their CP aa identity values mostly below the CP threshold. We argue that if the Rep aa identity distribution is at the borderline range but exclusively at the upper limit of 80–82%, these sequences should be classified as one species. CTLaV sequences identified as non-recombinant by RDP5 showed CP aa identities lower than expected (Fig. 8, sequences within rectangles), and were subjected to further GARD recombination analysis, which identified signals of recombination for the both Rep ($P<10^{-200}$) and CP ($P=7\times10^{-18}$).

Lastly, we compared the state of the current taxonomy with the one proposed after applying all above-mentioned changes (Fig. 9). Under the current criteria, Areas II and IV (Rep aa identity <80% and CP aa identity >80%) are composed both of intraspecific and intrageneric pairwise comparisons (Fig. 9a), whereas in the revised criteria, they are composed mainly of intrageneric

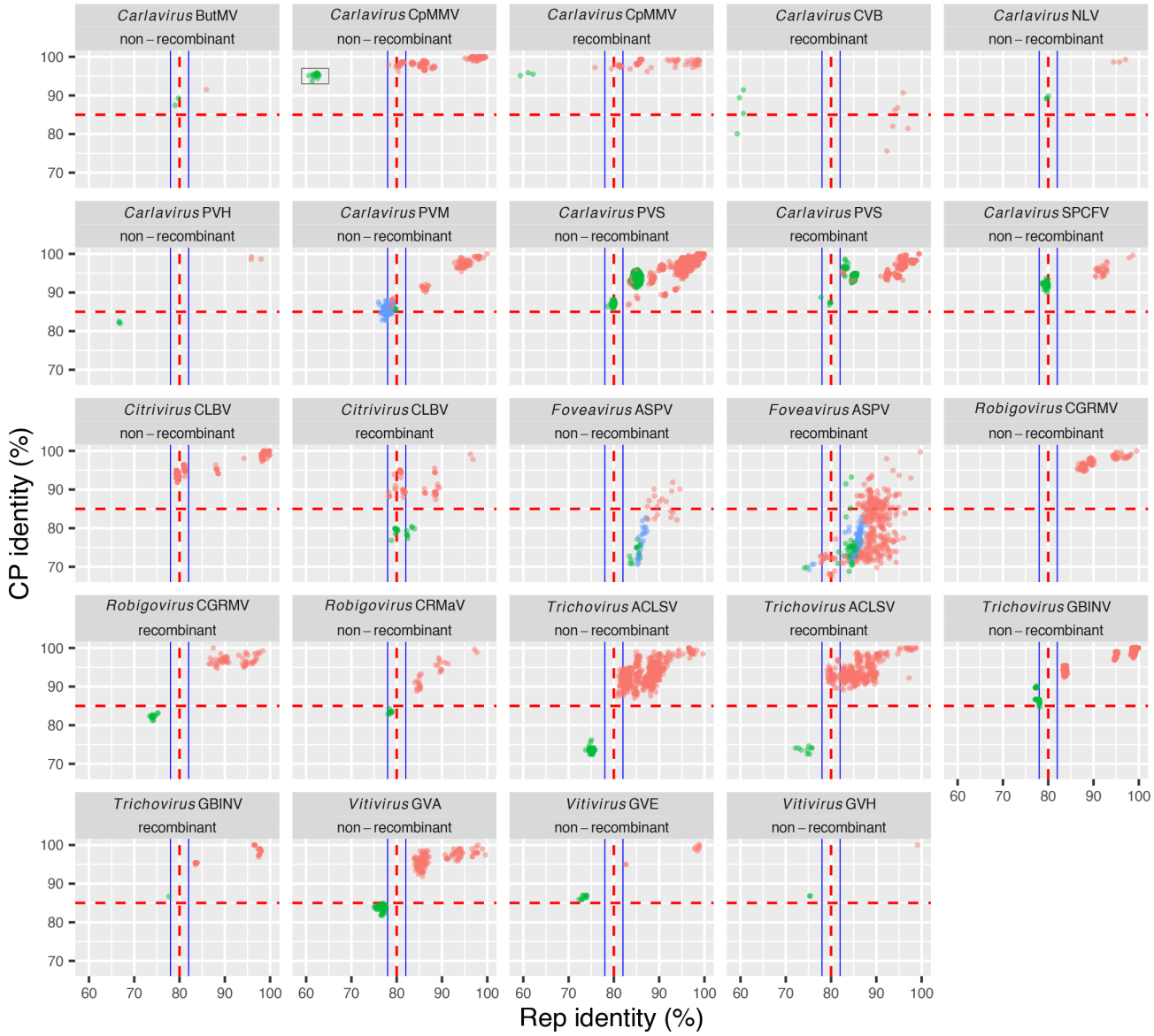

Fig. 7. Intraspecific analyses with pairwise member identities below the proposed thresholds. Species containing pairwise member identities within areas II, IV and VI of Fig. 5 were individually analysed. Green and blue dots represent monophyletic groups of viruses that could be assigned to a new species (see Fig. S1) and red dots represent pairwise comparisons within members of the same monophyletic group. Red dashed lines represent the proposed thresholds for the Rep and CP, and blue lines represent the borderline range of the Rep. Rectangles show cases that were further subjected to recombination analysis with GARD. Apple chlorotic leaf spot virus, ACSLV; apple stem-pitting virus, ASPV; butterbur mosaic virus, ButMV; cherry green ring mottle virus, CGRMV; citrus leaf blotch virus, CLBV; cowpea mild mosaic virus, CpMMV; cherry rusty mottle-associated virus, CRMaV; chrysanthemum virus B, CVB; grapevine berry inner necrosis virus, GBINV; grapevine virus A, GVA; grapevine virus E, GVE; grapevine virus H, GVH; nerine latent virus, NLV; potato virus H, PVH; potato virus M, PVM; potato virus S, PVS; sweet potato chlorotic fleck virus, SPCFV.

comparisons with a few intraspecific comparisons at the borderline range (Fig. 9b). Notably, a very few intraspecific comparisons remain in Area VI (Rep aa identity <80% and CP aa identity <80%) under the revised criteria (Fig. 9b). However, no intraspecific pairwise comparison is seen in this area when analysing only non-recombinant isolates (Fig. 9c and Fig. 9d), indicating that these ambiguous cases are very likely due to recombination. More specifically, these comparisons concern recombinant ASPV isolates (Fig. 7). Pairwise comparisons between ASPV and ApLV isolates are also influenced by recombination and represent the only interspecies comparisons in Area V with Rep aa identities significantly above the borderline range (Fig. 6). Even with these discrepancies, the proposed taxonomy can differentiate with high accuracy bona fide species in the family, and should aid the assignment of novel sequences, especially in cases where valuable biological information is missing.

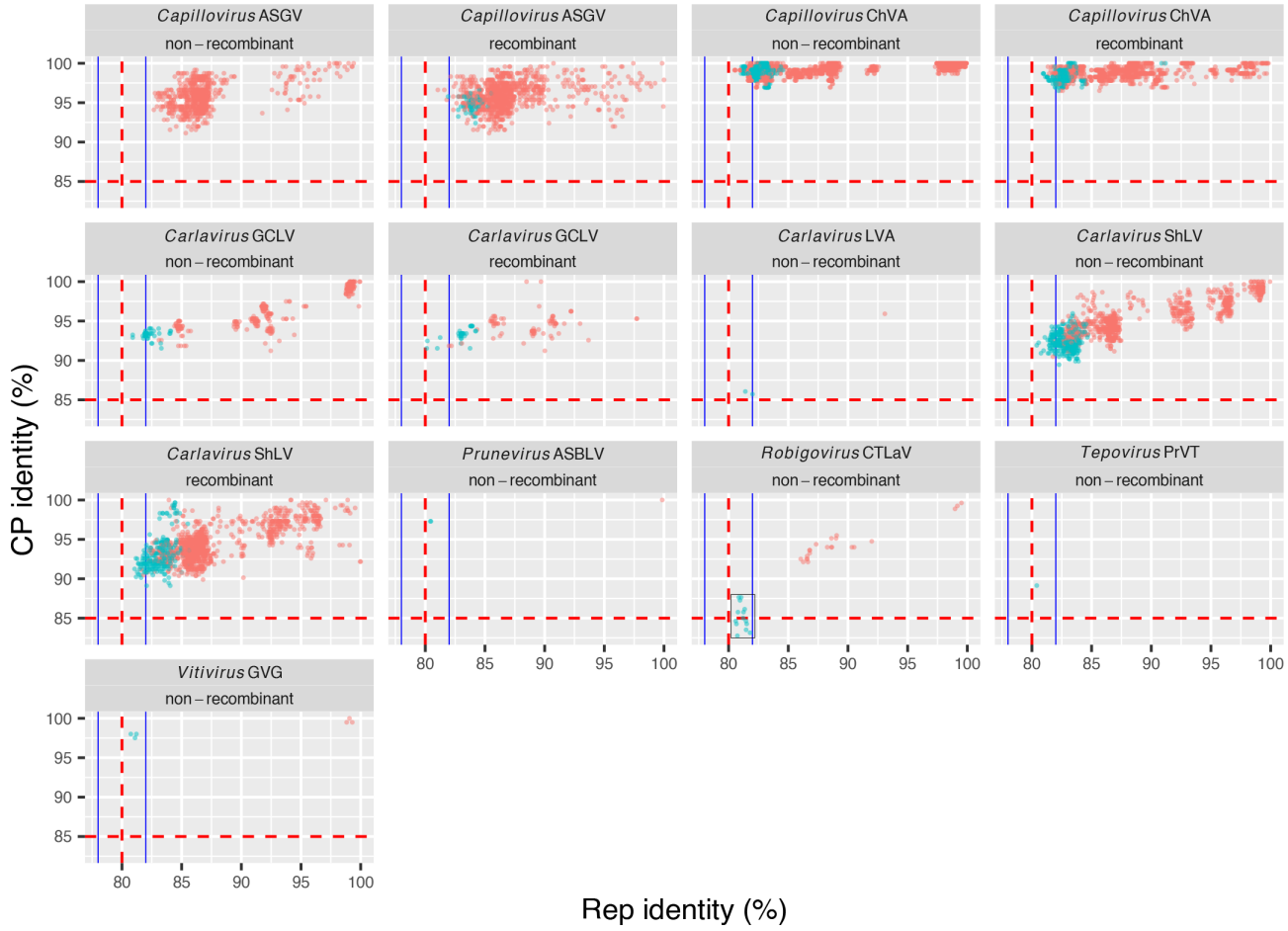

Fig. 8. Intraspecies analyses with pairwise percentage identities of the Rep at the upper borderline range. Red dashed lines represent the proposed thresholds for the Rep and CP, and blue lines represent the borderline range of the Rep. Blue dots represent monophyletic groups that could be assigned to a new species. Rectangles show cases that were further subjected to recombination analysis with GARD. ASBLV, Actinidia seed-borne latent virus; ASGV, apple stem grooving virus; ChVA, cherry virus A; CTLaV, cherry twisted leaf-associated virus; GCLV, garlic common latent virus; GVG, grapevine virus G; LVA, ligustrum virus A; PrVT, prunus virus T; SLV, shallot latent virus.

CONCLUSION

We propose an update on the current taxonomic criteria of the family *Betaflexiviridae* based on detailed phylogenetic and pairwise protein sequence identity analyses. The new criteria will resolve most of the current problems in family classification and should facilitate the assignment of newly discovered viruses to existing or novel species. Notably, these criteria are aimed at classifying viruses based on genomic sequences-derived data alone, in particular identities of aa sequences of the entire Rep polypeptide, complemented with those of the CP, when needed. However, biological properties such as host range, symptomatology, vector specificity and mode of transmission should be used as supplementary criteria whenever this information is available.

A schematic representation of the proposed criteria and workflow is shown in Fig. 10 and is summarized below:

(1) The primary demarcation criterion for the family *Betaflexiviridae* should be less than 80% aa identity of the Rep protein between members of distinct species. If the aa identity between certain isolates is in the borderline range (78–82%), the CP aa identity can be used as a secondary criterion to help/solve their final allocation. In this case, the threshold for the CP is set at 85% identity. If all pairwise comparisons of the Rep are at the borderline range but above the threshold (80–82%), we advise not to create a new species, but to classify those viruses in existing taxa. Alternatively, a flexible threshold can be applied to determine species-specific thresholds when accuracy statistics can be calculated, preferentially with the aid of biological properties.

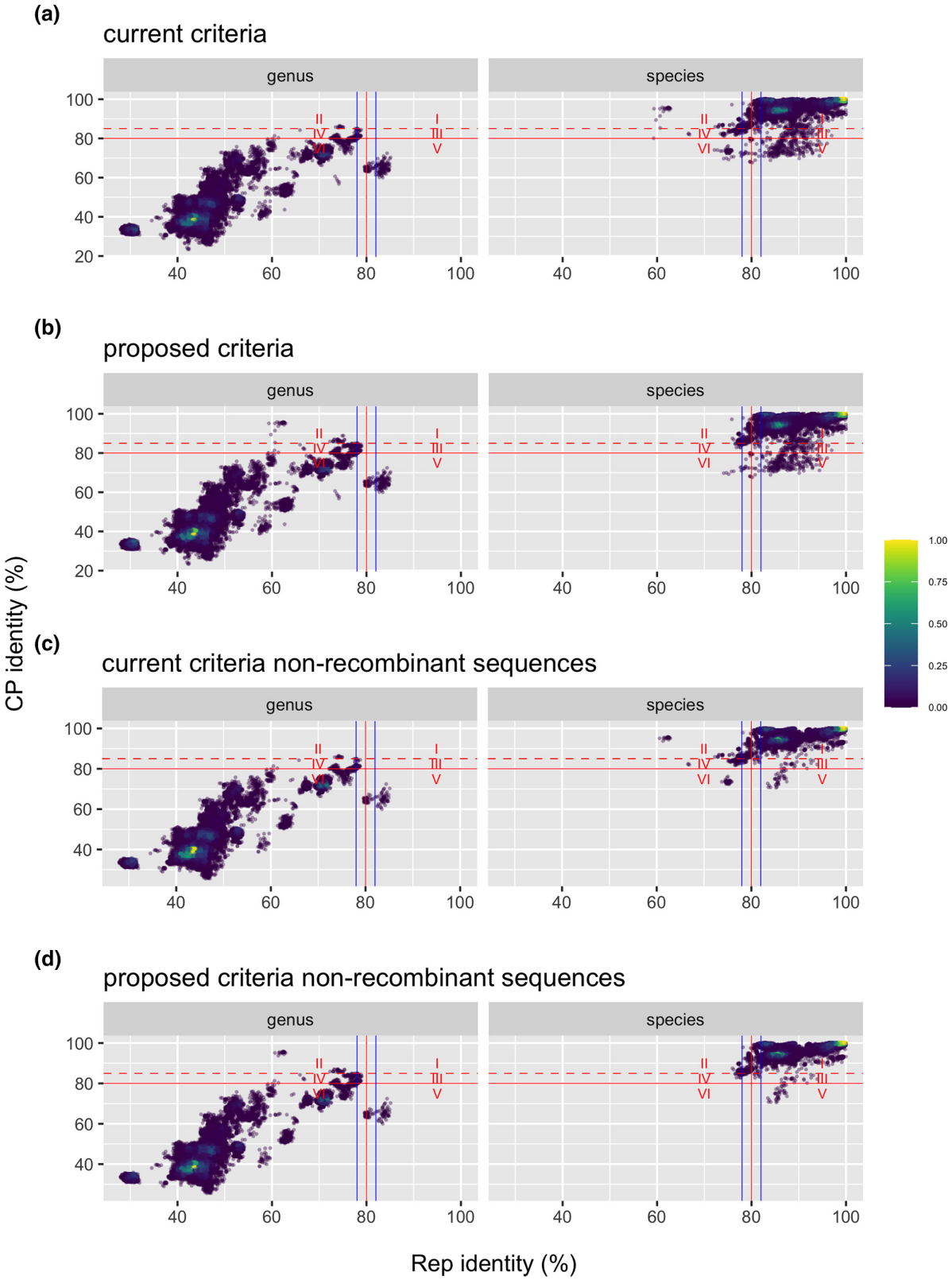

Fig. 9. Dot plots of Rep and CP pairwise identities of the current and proposed criteria at the genus and species levels. Warm and cold colours represent high and low density of data, respectively. Red solid lines represent the current identity thresholds for the family and the red dashed line represents the proposed threshold for the CP aa identity of 85%.

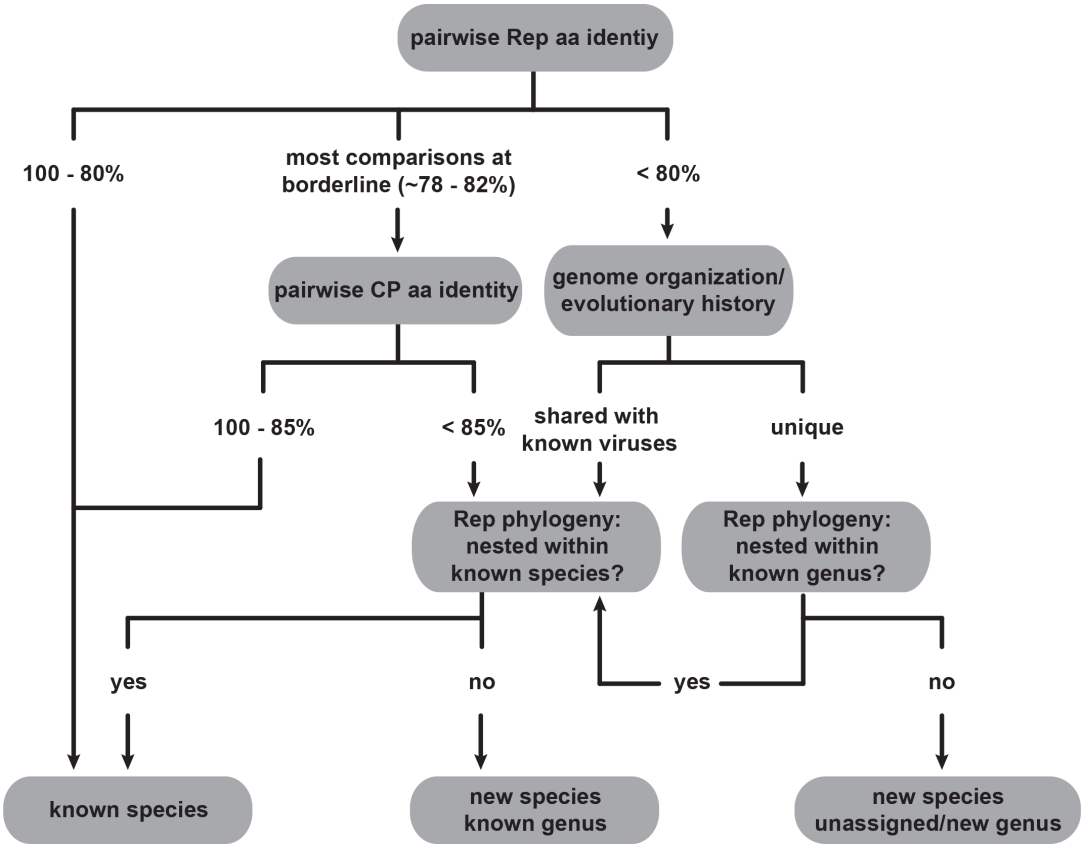

Fig. 10. Proposed taxonomic criteria for the family *Betaflexiviridae* presented as a decision tree to classify newly described viruses based on genomic information.

(2) Recombination among the members in the family *Betaflexiviridae* makes the evolutionary history of CP different from that of the Rep. Because of this characteristic and of the Rep direct link to viral evolution and lineages, the Rep protein identity demarcation criterion should be given priority.
(3) Members of the same species must be monophyletic based on the Rep phylogeny.
(4) Biological characteristics, if available, should be considered when appropriate to differentiate species, especially for borderline situations.

Although this study focused on addressing current problems in classifying viruses belonging to the family *Betaflexiviridae*, methods and approaches described here to revise demarcation criteria may be easily applied to classify viruses belonging to other families of the order *Tymovirales* (as they have similar genome organization) and to other taxa where principles of species distinction based upon sequence differences in taxonomically informative genes and/or their product have become blurred with the recent massive influx of data generated by HTS-based studies.

However, it is important to highlight those revised criteria described here are not an official part of the ICTV taxonomy yet. Indeed, an official taxonomic proposal for comprehensive taxonomic revision of viruses in the family *Betaflexiviridae*, based upon criteria proposed in this paper will be prepared and submitted to the ICTV for consideration in 2023.

Funding information
This work received no specific grant from any funding agency.

Author contribution
The design of the study was done by J.M.F.S., F.L.M., S.F.E, T.C., S.S., I.E.T. and T.N. The data analyses were done by J.M.F.S. The manuscript was drafted, edited and revised by all authors.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, *et al*. Consensus statement: virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 2017;15:161–168.
2. Gorbalenya AE, Krupovic M, Siddell SG, Varsani A, Kuhn JH. Proposal 2017.006G. Riboviria: a single taxon that comprises RNA viruses at the basal rank of virus taxonomy. (ICTV, 2018); (n.d.). https://ictv.global/filebrowser/download/4916
3. Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, *et al*. *Betafexiviridae*. In: King AMQ, Adams MJ, Carstens EB and Lefkowitz EJ (eds). *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*. Cambridge: Elsevier Academic Press; 2012. pp. 920–941.
4. Morozov SY, Solovyev AG. Triple gene block: modular design of a multifunctional machine for plant virus movement. *J Gen Virol* 2003;84:1351–1366.
5. Melcher U. The "30K" superfamily of viral movement proteins. *J Gen Virol* 2000;81:257–266.
6. Marais A, Faure C, Candresse T. New insights into Asian prunus viruses in the light of NGS-based full genome sequencing. *PLoS One* 2016;11:e0146420.
7. Cao M, Li P, Zhang S, Yang F, Zhou Y, *et al*. Molecular characterization of a novel citrivirus from citrus using next-generation sequencing. *Arch Virol* 2018;163:3479–3482.
8. Liu H, Wu L, Zheng L, Cao M, Li R. Characterization of three new viruses of the family Betaflexiviridae associated with camellia ringspot disease. *Virus Res* 2019;272:197668.
9. Maree HJ, Blouin AG, Diaz-Lara A, Mostert I, Al Rwahnih M, *et al*. Status of the current vitivirus taxonomy. *Arch Virol* 2020;165:451–458.
10. Reynard JS, Brodard J, Remoliff E, Lefebvre M, Schumpp O, *et al*. A novel foveavirus identified in wild grapevine (*Vitis vinifera* subsp. *sylvestris*). *Arch Virol* 2020;165:2999–3002.
11. Thompson JR, Dasgupta I, Fuchs M, Iwanami T, Karasev AV, *et al*. ICTV virus taxonomy profile: *Secoviridae*. *J Gen Virol* 2017;98:529–531.
12. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Arch Virol* 2021;166:2633–2648.
13. Goh CJ, Park D, Kim H, Sebastiani F, Hahn Y. Novel divavirus (the family *Betaflexiviridae*) and mitovirus (the family *Narnaviridae*) species identified in basil (*Ocimum basilicum*). *Acta Virol* 2018;62:304–309.
14. Wang R, Dong J, Wang Z, Zhou T, Li Y, *et al*. Complete nucleotide sequence of a new carlavirus in chrysanthemums in China. *Arch Virol* 2018;163:1973–1976.
15. Zhou J, Zhang Z, Lu M, Xiao H, Habili N, *et al*. Complete nucleotide sequence of a new virus, peach chlorotic leaf spot virus, isolated from flat peach in China. *Arch Virol* 2018;163:3459–3461.
16. Alabi OJ, McBride S, Appel DN, Al Rwahnih M, Pontasch FM. Grapevine virus M, a novel vitivirus discovered in the American hybrid bunch grape cultivar Blanc du Bois in Texas. *Arch Virol* 2019;164:1739–1741.
17. Goh CJ, Hahn Y. Identification of a novel member of the family *Betaflexiviridae* from the hallucinogenic plant *Salvia divinorum*. *Acta Virol* 2019;63:373–379.
18. Marais A, Faure C, Theil S, Candresse T. Characterization of the virome of shallots affected by the shallot mild yellow stripe disease in France. *PLoS One* 2019;14:e0219024.
19. Park D, Zhang M, Hahn Y. Novel foveavirus (the family *Betaflexiviridae*) species identified in ginseng (*Panax ginseng*). *Acta Virol* 2019;63:155–161.
20. Silva G, Bömer M, Rathnayake AI, Sewe SO, Visendi P, *et al*. Molecular characterization of a new virus species identified in yam (*Dioscorea* spp.) by high-throughput sequencing. *Plants* 2019;8:E167.
21. De La Torre-Almaráz R, Pallás V, Sánchez-Navarro JA. Molecular characterization of a new trichovirus from peach in Mexico. *Arch Virol* 2019;164:2617–2620.
22. Da Silva LA, Oliveira AS, Melo FL, Ardisson-Araújo DMP, Resende FV, *et al*. A new virus found in garlic virus complex is A member of possible novel genus of the family Betaflexiviridae (order Tymovirales). *PeerJ* 2019;7:e6285.
23. Thekke-Veetil T, Ho T. Molecular characterization of a new vitivirus discovered in a blueberry plant with green mosaic symptoms. *Arch Virol* 2019;164:2609–2611.
24. Wu L, Liu H, Bateman M, Komorowska B, Li R. First identification and molecular characterization of a novel cherry robigovirus. *Arch Virol* 2019;164:3103–3106.
25. Brewer E, Cao M, Gutierrez B, Bateman M, Li R. Discovery and molecular characterization of a novel trichovirus infecting sweet cherry. *Virus genes* 2020;56:380–385.
26. Diaz-Lara A, Mollov D, Golino D, Al Rwahnih M. Complete genome sequence of rose virus A, the first carlavirus identified in rose. *Arch Virol* 2020;165:241–244.
27. Gazel M, Roumi V, Ördek K, Maclot F, Massart S, *et al*. Identification and molecular characterization of a novel foveavirus from *Rubus* spp. in Turkey. *Virus Res* 2020;286:198078.
28. Li Z, Wang H, Zhao R, Zhang Z, Xia Z, *et al*. Complete genome sequence of a novel capillovirus infecting *Hevea brasiliensis* in China. *Arch Virol* 2020;165:249–252.
29. Liu Q, Yang L, Xuan Z, Wu J, Qiu Y, *et al*. Complete nucleotide sequence of loquat virus A, a member of the family *Betaflexiviridae* with a novel genome organization. *Arch Virol* 2020;165:223–226.
30. Maachi A, Nagata T, Silva JMF. Date palm virus A: first plant virus found in date palm trees. *Virus Genes* 2020;56:792–795.
31. Marais A, Šafářová D, Navrátil M, Faure C, Cornaggia D, *et al*. Complete genome sequence of cherry virus T, a novel cherry-infecting tepovirus. *Arch Virol* 2020;165:1711–1714.
32. Mumo NN, Mamati GE, Ateka EM, Rimberia FK, Asudi GO, *et al*. Metagenomic analysis of plant viruses associated with papaya ringspot disease in *Carica papaya* L. in Kenya. *Front Microbiol* 2020;11:205.
33. Peracchio C, Forgia M, Chiapello M, Vallino M, Turina M, *et al*. A complex virome including two distinct emaraviruses associated with virus-like symptoms in *Camellia japonica*. *Virus Res* 2020;286:197964.
34. Zhao L, Cao M, Huang Q, Jing M, Bao W, *et al*. Occurrence and molecular characterization of Actinidia virus C (AcVC), a novel vitivirus infecting kiwifruit (*Actinidia* spp.) in China . *Plant Pathol* 2020;69:775–782.
35. Diaz-Lara A, Mollov D, Golino D, Al Rwahnih M. Detection and characterization of a second carlavirus in *Rosa* sp. *Arch Virol* 2021;166:321–323.
36. Goh CJ, Park D, Hahn Y. A novel tepovirus, Agave virus T, identified by the analysis of the transcriptome data of blue agave (*Agave tequilana*). *Acta Virol* 2021;65:68–71.
37. Luo Q, Hu S, Lin Q, Xu F, Peng J, *et al*. Complete genome sequence of a novel foveavirus isolated from *Allium sativum* L. in China. *Arch Virol* 2021;166:983–986.
38. Thekke-Veetil T, McCoppin NK, Hobbs HA, Hartman GL, Lambert KN, *et al*. Discovery of a novel member of the Carlavirus genus from soybean (Glycine max L. Merr.). *Pathogens* 2021;10:223.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
40. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150–3152.
41. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.

42. **Tumescheit C**, **Firth AE**, **Brown K**. CIAlign - A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. *Bioinformatics* 2020. DOI: 10.1101/2020.09.14.291484.
43. **Price MN**, **Dehal PS**, **Arkin AP**, **Poon AFY**. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010;5:e9490.
44. **Martin DP**, **Varsani A**, **Roumagnac P**, **Botha G**, **Maslamoney S**, ***et al***. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol* 2021;7:veaa087.
45. **Wang L-G**, **Lam T-Y**, **Xu S**, **Dai Z**, **Zhou L**, ***et al***. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol Biol Evol* 2020;37:599–603.
46. **Muhire BM**, **Varsani A**, **Martin DP**. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 2014;9:e108277.
47. **Edgar RC**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
48. **Kosakovsky Pond SL**, **Posada D**, **Gravenor MB**, **Woelk CH**, **Frost SDW**. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006;22:3096–3098.
49. **Yu G**, **Smith D**, **Zhu H**, **Guan Y**, **Lam TT**. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.
50. **Paradis E**, **Schliep K**. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526–528.
51. **Revell LJ**. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 2012;3:217–223.
52. **Nguyen LT**, **Schmidt HA**, **von Haeseler A**, **Minh BQ**. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
53. **Wolf YI**, **Kazlauskas D**, **Iranzo J**, **Lucía-Sanz A**, **Kuhn JH**, ***et al***. Origins and evolution of the global RNA virome. *mBio* 2018;9:e02329-18.
54. **Singh L**, **Hallan V**, **Martin DP**, **Ram R**, **Zaidi AA**. Genomic sequence analysis of four new chrysanthemum virus B isolates: evidence of RNA recombination. *Arch Virol* 2012;157:531–537.
55. **Villamor DEV**, **Eastwell KC**. Viruses associated with rusty mottle and twisted leaf diseases of sweet cherry are distinct species. *Phytopathology* 2013;103:1287–1295.
56. **Alabi OJ**, **Al Rwahnih M**, **Mekuria TA**, **Naidu RA**. Genetic diversity of grapevine virus A in Washington and California vineyards. *Phytopathology* 2014;104:548–560.
57. **Yoon JY**, **Joa JH**, **Choi KS**, **Do KS**, **Lim HC**, ***et al***. Genetic diversity of a natural population of apple stem pitting virus isolated from apple in Korea. *Plant Pathol J* 2014;30:195–199.
58. **Zanardo LG**, **Silva FN**, **Lima ATM**, **Milanesi DF**, **Castilho-Urquiza GP**, ***et al***. Molecular variability of cowpea mild mottle virus infecting soybean in Brazil. *Arch Virol* 2014;159:727–737.
59. **Marais A**, **Faure C**, **Mustafayev E**, **Candresse T**. Characterization of new isolates of apricot vein clearing-associated virus and of a new *Prunus*-Infecting virus: evidence for recombination as a driving force in *Betaflexiviridae* evolution. *PLoS One* 2015;10:e0129469.
60. **Shimodaira H**. An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 2002;51:492–508.
61. **Nicaise V**, **Candresse T**. Plum pox virus capsid protein suppresses plant pathogen-associated molecular pattern (PAMP)-triggered immunity. *Mol Plant Pathol* 2017;18:878–886.
62. **Villamor DEV**, **Susaimuthu J**, **Eastwell KC**. Genomic analyses of cherry rusty mottle group and cherry twisted leaf-associated viruses reveal a possible new genus within the family *Betaflexiviridae*. *Phytopathology* 2015;105:399–408.
63. **Grimová L**, **Ryšánek P**. Apricot latent virus – review. *Hort Sci* 2012;39:144–148.
64. **Jelkmann W**, **Paunovic S**. Apple stem pitting virus. In: **Hadidi A**, **Barba M**, **Candresse T** and **Jelkmann W** (eds). *Virus and Virus-like Diseases of Pome and Stone Fruits. APS Press. Chapter 8*. 2018.
65. **Sulistyowati E**, **Mitter N**, **Bastiaan-Net S**, **Roossinck MJ**, **Dietzgen RG**. Host range, symptom expression and RNA 3 sequence analyses of six Australian strains of *Cucumber mosaic virus* . *Austral Plant Pathol* 2004;33:505.