

Comparative analysis of common regions found in babuviruses and alphasatellite molecules

Daisy Stainton¹ · Darren P. Martin² · David A. Collings^{1,3} · Arvind Varsani^{1,2,4} 

Received: 18 July 2016 / Accepted: 24 October 2016 / Published online: 22 November 2016
© Springer-Verlag Wien 2016

Abstract Viruses in the genus *Babuvirus* have multi-component ssDNA genomes and often associate with alphasatellite molecules containing two common motifs, a common-region stem-loop (CR-SL) involved in initiation of rolling-circle replication and a common-region major (CR-M) motif involved in secondary-strand synthesis. We compared known babuvirus genome components and alphasatellite CR-SL and CR-M sequences, defining five divergent CR-SL sequence classes. We identified iterated sequence elements in babuvirus genome components that have particularly conserved sequences and spatial arrangements between known babuviruses.

Babuviruses (family *Nanoviridae*) are single-stranded DNA viruses that infect monocotyledons. Members of the three established species, *Banana bunchy top virus* (BBTV), *Abaca bunchy top virus* (ABTV) and *Cardamom bushy dwarf virus* (CBDV) [23] have multi-component genomes comprising at least six individually encapsidated

components that are all essential for infectivity. DNA-R encodes a replication initiation protein (Rep); DNA-S, a capsid protein (CP); DNA-M, a movement protein (MP); DNA-C, a cell-cycle link protein (Clink); and DNA-N, a nuclear shuttle protein (NSP). The sixth canonical component, DNA-U3, contains no obvious genes and has no known function [23]. CBDV has two additional components, DNA-Uf1 and DNA-Uf2, neither of which contains known genes [15]. All six canonical components contain two highly conserved motifs, the common region stem-loop (CR-SL) and the common region major (CR-M) motif [3, 15]. As babuviruses have only one Rep-encoding component, the other components must be trans-replicated. This is facilitated by conservation in CR-SLs of the origin of replication (*v-ori*), stem-loop sequences where rolling-circle replication [RCR] is initiated, and iterated Rep recognition sequences called iterons F1, F2 and R (Sup. Fig. 1) [3, 8, 9]. In BBTV, mutations in any of these iterons have some effect on replication, but those in F2 generally have the greatest impact [9].

BBTV DNA-U3 is unique among babuvirus genome components, having the smallest CR-SL and no identified iteron R, its CR-SL starting 14–21 nt closer to the *v-ori* hairpin than other BBTV components [3], although an iteron R sequence may be present in the ~90 bp 5' direction from the *v-ori* hairpin [9]. Consistent with BBTV DNA-U3 having a larger CR-SL than other components, Wang et al. [24] identified a sequence element in the 5' direction from the DNA-U3 CR-SL that is conserved in CR-SL sequences of all BBTV components

Babuviruses and other nanoviruses frequently associate with Rep-encoding DNA-R-like alphasatellite molecules, with babuvirus alphasatellites being ~20% larger than nanovirus alphasatellites [14]. These alphasatellites are most closely related to alphasatellites associated with

Electronic supplementary material The online version of this article (doi:10.1007/s00705-016-3168-1) contains supplementary material, which is available to authorized users.

✉ Arvind Varsani
arvind.varsani@asu.edu

- ¹ School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
- ² Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa
- ³ School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia
- ⁴ The Biodesign Center for Fundamental and Applied Microbiomics, and School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA

plant-infecting single-stranded DNA viruses in the family *Geminiviridae* [18, 25]. Alphasatellites are capable of autonomous replication but cannot trans-replicate the genome components of nanoviruses and geminiviruses [11, 22] and are probably encapsidated by their associated viruses [2, 19]. CR-SL sequences occur in BBTV alphasatellites but the only similarities found between these and canonical BBTV genome components were in *v-ori* hairpin structures that contained a TAGTATTAC loop sequence in alphasatellites and TATTATTAC in BBTV [12].

We used MUSCLE [5] to align separate component-specific datasets of BBTV, ABTV [21] and CBDV [15, 20] (Sup. Table 1), identifying and separating CR-M and C-SL sequences for each babuvirus species and component.

Given that CR-SL of BBTV DNA-U3 had previously been identified without an iteron R, a region in the 5' direction from the previously identified CR-SL boundary was included in the CR-SL datasets. The CR-SL sequences of all three babuviruses were aligned using MAFFT [13], the 5' region was trimmed to include only CR-SL sequences that were obviously similar between different components, and the remaining sequence fragments were realigned. The CR-SL sequences were then grouped into species- and component-specific datasets, and a consensus representation (sequence logo) was constructed using WebLogo v3.4 [4]. As the BBTV DNA-U3 CR-SL dataset had numerous sequences containing an insert in the CR-SL region, the dataset was split into two separate datasets, and separate CR-SL sequence logos were generated. Pairwise identities (PIs) of iterons R, F1 and F2 were determined using SDT v1.2 [17].

All analysed components, including BBTV DNA-U3 and CBDV DNA-N, had sequences resembling iteron R, F1 and F2. Overall, iteron R was most conserved, with 88% of sequences sharing 100% PI (Sup. Table 3). Also, 99.5% of the components of all three species had an identical nonanucleotide (TATTATTAC) at their presumed *v-ori*.

We noted various insertions in the BBTV DNA-U3 CR-SL sequences that affected component alignments. Two Taiwanese sequences in the DNA-U3 CR-SL group contributed an alignment gap, indicated by vertical lines in Fig. 1 between positions 94 and 115, due to 16- and 17-nt inserts at different sites in the 3' direction from the *v-ori* hairpin. The ~ 30-nt insert that split the DNA-U3 datasets was present in 111 sequences in the 5' direction from the *v-ori* hairpin and replaced a 9-nt region seen in most other sequences. Seventy-six of the 77 BBTV DNA-U3 sequences without this insertion had the expected GTCCC iteron R sequence. In contrast, none of the 111 sequences with the insertion had an iteron R sequence at the same position as in most other components. Instead, 95% of these sequences had a GCCTC sequence (Fig. 1), which

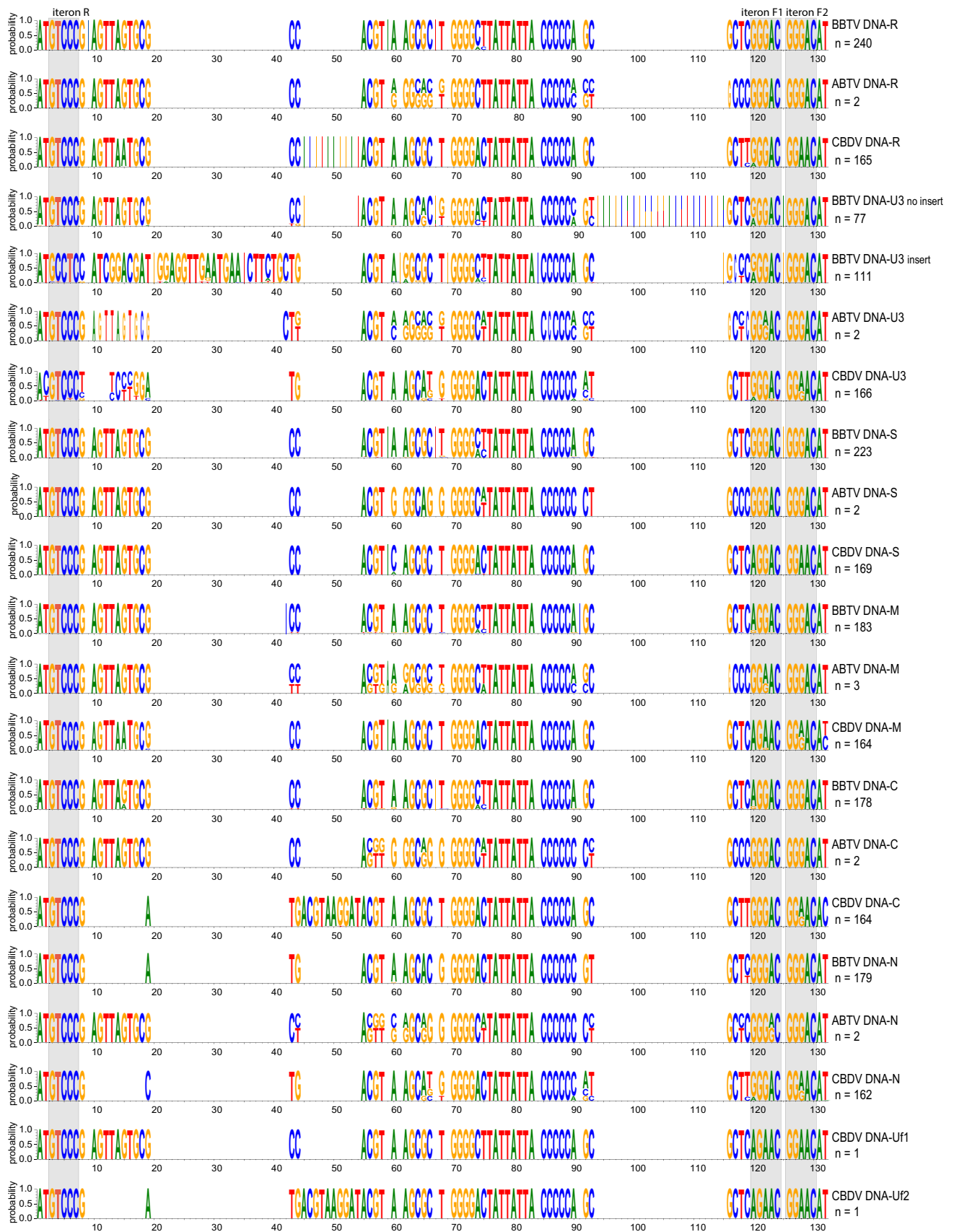
Fig. 1 Logo of aligned babuvirus CR-SL sequences. Iterons R, F1 and F2 are highlighted in grey. The letter heights indicate relative proportions of nucleotides at each site, while letter widths reflect the number of gaps at that site across all the sequences in the alignment (narrower implying more gaps)

may be functionally equivalent to iteron R. There was, however, an iteron-R-like GTCCC sequence ~ 50 nt in the 5' direction from the GCCTC sequence that previously had been identified as a possible DNA-U3 iteron R [9]. Based on the alignment of the CR-SL with all the components, it is more likely that GCCTC acts as iteron R for these 111 DNA-U3 sequences. Thus, we propose that we have identified the full CR-SL and all iterons of BBTV DNA-U3.

A consequence of using an alignment encompassing a larger section of the DNA-N sequence than was used in previous analyses [15] was that we identified probable iteron R sequence in the CR-SL of this component in CBDV (Fig. 1). The two additional, largely uncharacterised components of CBDV, CBDV-Uf1 and CBDV-Uf2 also contained CR-SL sequences more similar to those of the six canonical CBDV components than to known babuvirus alphasatellites, suggesting that these are genuine CBDV genome components.

Trans-replication has been shown experimentally for three nanoviruses, faba bean necrotic yellows virus (FBNYV), milk vetch dwarf virus (MDV) and subterranean clover stunt virus (SCSV), where each DNA-R was able to replicate the DNA-C of a member of another species [22]. Similarities between the CR-SL sequences of all three babuviruses suggest that similar trans-replication might also occur in the babuviruses. Although the consensus iteron R of all six babuvirus genome components is GTCCC, individual sequences had variations of this consensus sequence (GTCTC/GTCGC/GCCCC/TTCGC/CTC CC/CCCC/GTCCT/ATCCC/CTCTC/TGCTC/TCGCC/TCCTC). More variation was detected from the consensus iteron F1 (GGGAC) and iteron F2 (GGGAC), with only 48% and 44%, respectively, of the sequences sharing 100% identity (Sup. Table 3). Variability was found both between members of different species, with iteron F1 in DNA-M AGGAC/GGAAC/AGAAC in BBTV, ABTV, and CBDV respectively, and within species, with isolates of CBDV iteron F2 containing GG[A/G]AC (Fig. 1). Although mutagenesis of these conserved iterons results in varying degrees of replicative fitness loss [9], the extensive variations seen in these iteron sequences suggests a degree of flexibility in the exact sequence motifs that Rep can recognise.

Having identified conserved elements in babuvirus CR-SLs, we attempted to discover conserved sequences in CR-M. Due to the low sequence similarity between babuviruses, the CR-Ms were aligned separately using MAFFT,



with the percentage PI of the CR-M sequences determined in SDT v1.2. The three alignments were re-split according to genome component, and a consensus representation of the CR-M sequences for each component was generated for each species using Weblogo v3.4 (Sup. Fig. 2). A GC-rich region of CR-M displayed detectable sequence similarity across all species. This conserved portion of CR-M from BBTV, ABTC and CBDV was combined into a single dataset and realigned using MAFFT to reveal a partially conserved GGGCCGNAGGCC sequence in 98.9% of the analysed babuvirus genome components (Fig. 2). This GC-rich sequence has the potential to form a stable secondary structure in single-stranded BBTV genomic DNA [7, 16] and is similar to the GC-boxes in the transcription-promoting rightward promoter element (Rpe1) in geminiviruses [3, 6]. Other GC-rich regions occur in single-stranded bacterial plasmids, where they help to prime complementary-strand synthesis following RCR [7, 10]. However, the 5' region of the BBTV CR-M lacks the conserved GC-rich sequence and has been identified as the binding site of an oligonucleotide primer involved in secondary-strand synthesis [7]. It is therefore likely to have this function in ABTV and CBDV.

We attempted to identify the CR-SL and CR-M sequences of all available geminivirus and nanovirus-associated alphasatellites to compare them to the DNA-R, CR-SL and CR-M sequences of BBTV, CBDV and ABTV (Fig. 3). The non-coding regions of alphasatellites from babu-, nano- and geminiviruses (Sup. Table 2) were aligned using MUSCLE. Regions containing the characteristic *v-ori* nonanucleotide sequence were identified and further aligned to identify the boundaries of the CR-SL. The alphasatellite CR-SL sequences, together with those of the BBTV, ABTV and CBDV DNA-R components, were aligned using MAFFT. This alignment was then separated into five groups of CR-SL sequences consisting of (1) babuvirus DNA-R; (2) babuvirus alphasatellites (babu-alphas), (3) the coconut foliar decay alphasatellite sequence, (4) nanovirus alphasatellite (nano-alphas), and (5) geminivirus alphasatellite (gemi-alphas).

While babuvirus DNA-R CR-SLs shared the same *v-ori* TATTATTAC sequence (Fig. 1), the consensus *v-ori* nonanucleotide in BBTV, TAGTATTAC, [12] (Fig. 3A) was the same in all the other alphasatellites, and alphasatellites were more similar to each other across their entire CR-SL than to babuvirus DNA-R, suggesting that they were not derived from babuvirus DNA-R sequences.

No sequences homologous to babuvirus iterons R, F1 and F2 were detected in the CR-SLs of the babuvirus-associated alphasatellites. Although 5-nt-long sequences resembling these iterons were found in the 5' direction from alphasatellite CR-SL in some sequences (not shown), these were not conserved within or between the different

alphasatellite groups, and these potential iterons were statistically less common than expected for random 5-nt-long sequences (expected frequency, 1 in every 1024 nucleotides; observed frequency, 1 in 1292 nucleotides).

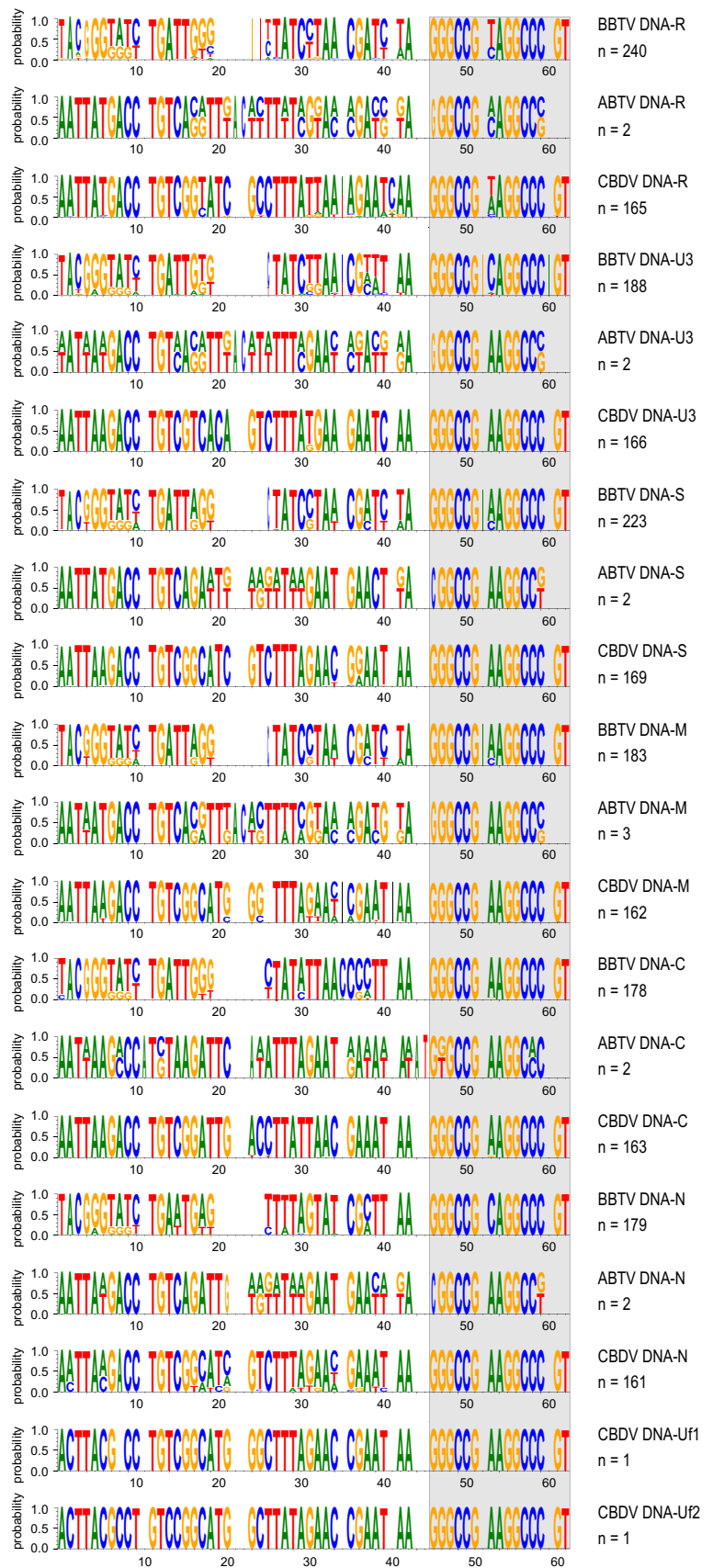
To identify elements in CR-M that are conserved between babuviruses and alphasatellites, the alphasatellite CR-M sequences were identified by realigning GC-rich CR-M regions of the various DNA-R sequences with complete alphasatellite sequences, and the probable CR-M sequences were aligned with the babuvirus DNA-R CR-M sequences using MAFFT. However, as alphasatellite CR-Ms were highly diverse, sequences from the five different groups were separately realigned with MAFFT, and some groups were subdivided. The gemini-alpha group was split into two groups, and the babu-alpha group was split into four groups (BBTV-1, BBTV-2, CBDV and a group containing all remaining babu-alphas). The percentage PIs for the CR-M of each group of alphasatellites was determined using SDT v1.2.

The alphasatellite CR-Ms, like those of babuviruses, displayed little conservation between the different groups outside of a specific GC-rich region in either the 5' or 3' portion of the CR-M (Fig. 3B). Like the canonical babuvirus components, the babuvirus alphasatellite GC-rich regions contained GGGCCGNAGGCC, whereas geminivirus alphasatellites had G/TGCCG/CCGCAG. Two distinct groups of BBTV alphasatellite CR-M sequences were identified, with group-2 sequences, which all share >80% PI, being more similar to the CBDV alphasatellite CR-M sequences (>71% PI) than to the group-1 sequences (>67% PI; Fig. 3B).

Two distinct groups of geminivirus alphasatellites were also evident. Whereas group 2 included most of the available geminivirus-associated alphasatellite sequences (155 sequences) and had a GC-rich region at the 3' end of the CR-M, members of group 1 contained a GC-rich region that was located further toward the 5' end as well as sequences that were more conserved in the 3' direction from the GC-rich region (Fig. 3B). Although the group-1 and -2 geminivirus-associated alphasatellites have similarities in their CR-M sequences in the 5' direction from the GC-rich region, this is highly diverse both across and within the groups.

Interestingly, a 5' region of the group-2 sequences is similar to a 3' region of the group-1 sequences (underlined, Fig. 3B), with both regions conserved in each group, suggesting biological functionality in geminivirus alphasatellites. This adenine-rich (A-rich) region was previously identified in both geminivirus-associated alphasatellites and geminivirus-associated betasatellites and may be involved in complementary-strand synthesis [1] or in increasing the size of satellites to approximately half that of full-length geminivirus genomes to ensure efficient

Fig. 2 Logo of aligned babuvirus CR-M sequences (only the conserved 3' region). Sequences of members of all species were aligned together. The GC-rich region is shown in grey. The letter widths reflect the number of gaps at that site across all the sequences in the alignment (narrower implying more gaps)



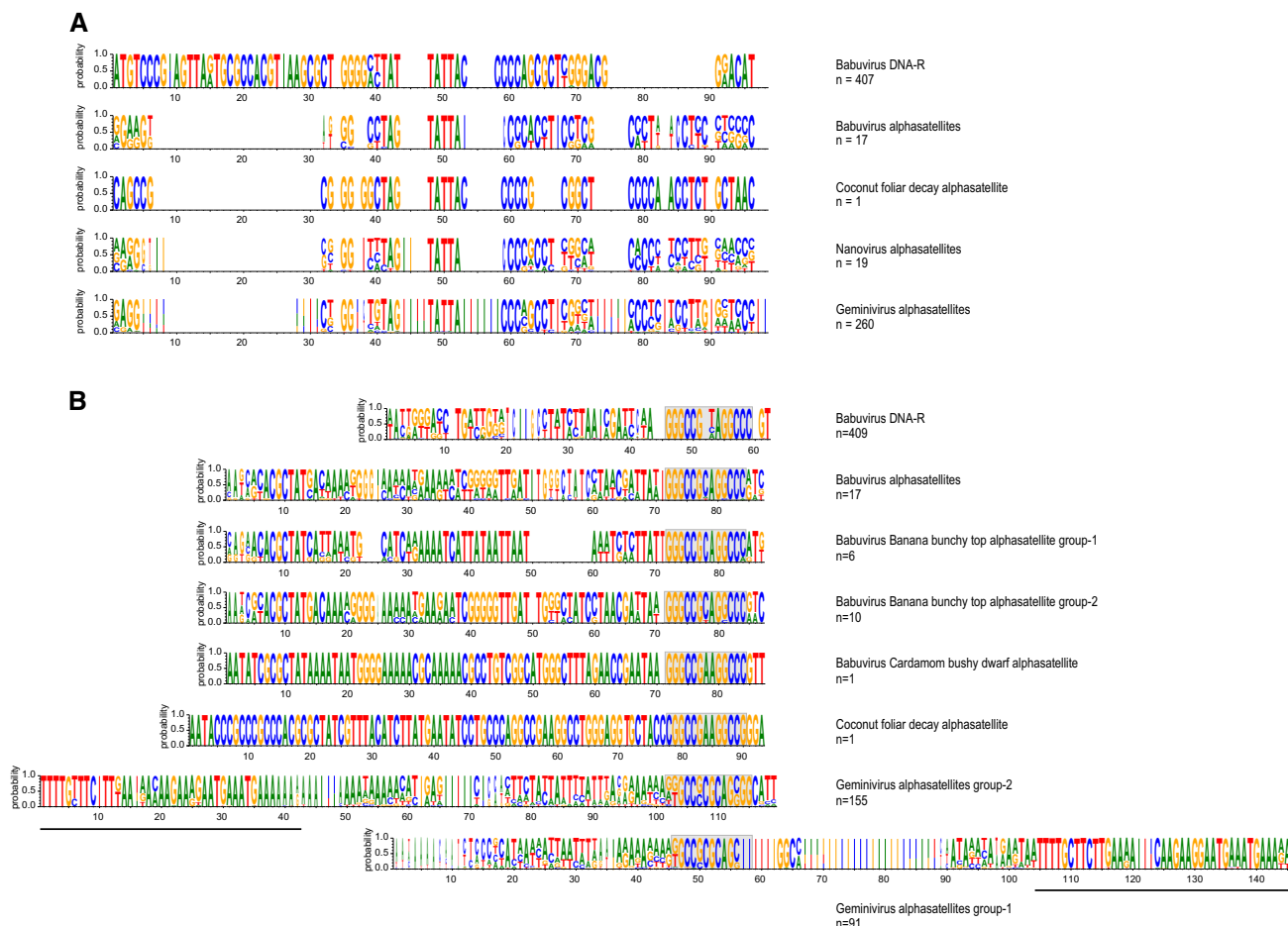


Fig. 3 The common regions of alphasatellites. **A)** The CR-SL of the alphasatellites along with the CR-SL sequences of babuvirus DNA-R components for comparison. **B)** The CR-M of the alphasatellites with the CR-M of babuvirus DNA-R components for comparison. BBTV and geminivirus-associated alphasatellites were split into two groups based on the similarities between their CR-M sequences. The black

horizontal line indicates a region of sequence conservation between the geminivirus-associated alphasatellite groups. The GC-rich region is shown in grey. In both A and B, the letter widths reflect the number of gaps at that site across all the sequences in the alignment (narrower implying more gaps)

trans-encapsidation [18]. As the 5' end of the CR-M in BBTV is a primer-binding site for complementary-strand synthesis [7], it is plausible that this conserved A-rich region might also act as a primer-binding site in geminivirus-associated alphasatellites. However, the reason that this conserved region is found on different sides of the CR-M GC-rich region in the group 1 and 2 alphasatellites remains unclear.

In summary, the sequences and structural arrangements of elements in babuvirus CR-SLs are strongly conserved between BBTV, ABTV and CBDV, suggesting that the Repls encoded by the DNA-R components of each of these viruses may trans-replicate the genome components of the others. Conversely, the CR-SL sequences of the gemini-, nano- and babuvirus-associated alphasatellite molecules are highly diverse but also detectably more similar to one another than to any known babuvirus DNA-R CR-SL sequences. This suggests that these alphasatellites are

unlikely to have been derived separately from their associated viruses and that they are probably incapable of trans-replicating either one another or the genome components of their helper viruses. Also, unlike the babuvirus CR-SL sequences, the babuvirus CR-M sequences were highly divergent, with high similarities being restricted to a GC-rich region at the 3' end of the CR-M. Like the babuvirus components, the CR-M of the alphasatellites was only conserved in a well-defined GC-rich region.

Acknowledgements This work was supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand (Grant UOC0903 to AV).

Compliance with ethical standards

There are no conflicts of interest; the research did not involve human participants or animals. The data used for these analyses are publicly available in GenBank.

References

1. Briddon RW, Bull SE, Amin I, Idris AM, Mansoor S, Bedford ID, Dhawan P, Rishi N, Siwatch SS, Abdel-Salam AM, Brown JK, Zafar Y, Markham PG (2003) Diversity of DNA β , a satellite molecule associated with some monopartite begomoviruses. *Virology* 312:106–121
2. Briddon RW, Stanley J (2006) Subviral agents associated with plant single-stranded DNA viruses. *Virology* 344:198–210
3. Burns TM, Harding RM, Dale JL (1995) The genome organisation of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76:1471–1482
4. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190
5. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
6. Fenoll C, Schwarz JJ, Black DM, Schneider M, Howell SH (1990) The Intergenic Region of Maize Streak Virus contains a Gc-Rich element that activates rightward transcription and Binds Maize nuclear factors. *Plant Mol Biol* 15:865–877
7. Hafner GJ, Harding RM, Dale JL (1997) A DNA primer associated with banana bunchy top virus. *J Gen Virol* 78:479–486
8. Hafner GJ, Stafford MR, Wolter LC, Harding RM, Dale JL (1997) Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J Gen Virol* 78:1795–1799
9. Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL (2006) An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. *J Gen Virol* 87:3409–3412
10. Honda Y, Sakai H, Hiasa H, Tanaka K, Komano T, Bagdasarian M (1991) Functional division and reconstruction of a plasmid replication origin: molecular dissection of the oriV of the broad-host-range plasmid RSF1010. *Proc Natl Acad Sci USA* 88:179–183
11. Horser CL, Harding RM, Dale JL (2001) Banana bunchy top nanovirus DNA-1 encodes the ‘master’ replication initiation protein. *J Gen Virol* 82:459–464
12. Horser CL, Karan M, Harding RM, Dale JL (2001) Additional rep-encoding DNAs associated with banana bunchy top virus. *Arch Virol* 146:71–86
13. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
14. Katul L, Timchenko T, Gronenborn B, Vetten HJ (1998) Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *J Gen Virol* 79:3101–3109
15. Mandal B, Shilpi S, Barman AR, Mandal S, Varma A (2013) Nine novel DNA components associated with the foorkey disease of large cardamom: evidence of a distinct babuvirus species in Nanoviridae. *Virus Res* 178:297–305
16. Muhire BM, Golden M, Murrell B, Lefevre P, Lett JM, Gray A, Poon AY, Ngandu NK, Semegni Y, Tanov EP, Monjane AL, Harkins GW, Varsani A, Shepherd DN, Martin DP (2014) Evidence of pervasive biologically functional secondary structures within the genomes of eukaryotic single-stranded DNA viruses. *J Virol* 88:1972–1989
17. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277
18. Saunders K, Stanley J (1999) A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: Evidence for interfamilial recombination between plant DNA viruses. *Virology* 264:142–152
19. Saunders K, Bedford ID, Stanley J (2002) Adaptation from whitefly to leafhopper transmission of an autonomously replicating nanovirus-like DNA component associated with ageratum yellow vein disease. *J Gen Virol* 83:907–913
20. Savory FR, Ramakrishnan U (2014) Asymmetric patterns of reassortment and concerted evolution in Cardamom bushy dwarf virus. *Infect Genet Evol* 24:15–24
21. Stainton D, Martin DP, Muhire BM, Lolohea S, Halafih M, Lepoint P, Blomme G, Crew KS, Sharman M, Kraberger S, Dayaram A, Walters M, Collings DA, Lemey P, Harkins GW, Batsirai M, Thomas JE, Varsani A (2015) The global distribution of Banana bunchy top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events. *Virus Evol* 1:vev009
22. Timchenko T, Katul L, Sano Y, de Kouchkovsky F, Vetten HJ, Gronenborn B (2000) The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. *Virology* 274:189–195
23. Vetten HJ, Dale JL, Grigoras I, Gronenborn B, Harding R, Randles JW, Sano Y, Thomas JE, Timchenko T, Yeh HH (2012) Nanoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruse*. Elsevier, Academic Press, Amsterdam, pp 395–404
24. Wang HI, Chang CH, Lin PH, Fu HC, Tang C, Yeh HH (2013) Application of motif-based tools on evolutionary analysis of multipartite single-stranded DNA viruses. *PLoS One* 8:e71565
25. Zhou X (2013) Advances in understanding begomovirus satellites. *Ann Rev Phytopathol* 51:357–381