nature

Vol 299 No 5881 23–29 September 1982 £1.60 $4.00

CADANG-CADANG DISEASE VIROID

Reprinted from Nature, Vol. 299, No. 5881, pp. 316-321, 23 September 1982

Viroid RNAs of cadang-cadang disease of coconuts

James Haseloff, Nizar A. Mohamed*† & Robert H. Symons

Adelaide University Centre for Gene Technology, Department of Biochemistry, University of Adelaide, Adelaide, South Australia 5001
* FAO/UNDP Coconut Research Project, Philippine Coconut Authority, Albay Research Center, Guinobatan, Albay, Philippines 4908

Cadang-cadang is a serious lethal disease of coconut palms in the Philippines. Infectivity is associated with four viroid RNAs (ccRNAs) which are from 246 to 574 residues in size but do not vary in sequence complexity. They share sequence and structural homology with other viroids. Variant ccRNAs found in nine isolates sequenced may have arisen as by-products or intermediates of viroid replication.

CADANG-CADANG is a serious and economically important disease of coconut palms which was first reported in 1927 on San Miguel Island in the Philippines[1]. By 1962, all but 100 of the 250,000 palms on this island had died from the disease[2]. Cadang-cadang disease has spread rapidly and now, 56 years after its first known occurrence, it is widespread over the south-east part of Luzon and many neighbouring islands (Fig. 1). It is estimated that cadang-cadang is responsible for the death of 500,000 palms each year in the Philippines (B. Zelazny, personal communication). Recent work has indicated that tinangaja disease of coconuts on Guam, an island in the Pacific 1,500 miles east of the Philippines, has the same aetiology as cadang-cadang[3]. However, cadang-cadang disease has not been found in any other coconut growing area.

Four low-molecular weight RNA species (ccRNAs) are found associated with cadang-cadang disease[4]. Two of these RNAs, called ccRNA 1 fast and ccRNA 2 fast, of ~ 250 and 500 nucleotides[5] respectively, are present in infected palms at early stages of the disease[4]. As infection progresses over a period of years, two additional RNAs, ccRNA 1 slow and ccRNA 2 slow (of ~300 and 600 nucleotides, respectively), characterized by their slower electrophoretic mobilities on polyacrylamide gels, appear and then predominate[5]. Ribonuclease fingerprinting showed that the ccRNAs, although differing in size, contain common repeated sequences, suggesting that the ccRNA 2 fast and slow species may be dimers of the respective ccRNA 1 species[5] and that the three larger RNAs are closely related to ccRNA 1 fast.

All four ccRNA species have been shown recently to be infectious[6]. They are single-strand covalently closed circular RNAs with high degrees of secondary structure[7,8] and so possess properties similar to those of a small group of infectious plant pathogens, the viroids[9]. Thus, the ccRNAs can now be classed definitely as the coconut cadang-cadang viroid (CCCV) in view of the recent infectivity data. The other viroids characterized each consist of single infectious RNA species which do not seem to code for functional polypeptide products but are capable of autonomous replication[9,10].

We report here the primary sequences and secondary structures of the four RNAs of several CCCV isolates. The results are compared with the structures of other viroids and are discussed in terms of the possible origin of cadang-cadang disease and mechanisms for the replication of viroids.

RNA sequence and structure determination

The approaches we have used successfully for the determination of the sequences and structures of chrysanthemum stunt viroid (CSV)[10], avocado sunblotch viroid (ASBV)[11], citrus exocortis viroid (CEV)[12], and the viroid-like RNAs (virusoids) of velvet tobacco mottle virus (VTMoV)[13] and solanum nodiflorum mottle virus (SNMV)[13] were applied to the four ccRNAs. Native circular ccRNAs were subjected to limited digestion either by ribonuclease T_1, which catalysed cleavage of ccRNA 1 species at single sites and ccRNA 2 species at either or both of two sites to produce specific full-length linear ccRNAs, or by ribonucleases A or U_2, which produced smaller overlapping RNA fragments. These linear RNA molecules were 5'- or 3'-radiolabelled and then purified by polyacrylamide gel electrophoresis[10,14]. The sequences of these 5'- or 3'-labelled fragments were determined by partial enzymatic digestion[10]. The use of fragments labelled separately at both the 5'- and 3'-ends allowed the sequence determination of long RNA fragments up to 574 residues long and, with shorter fragments, resolved gel compression artefacts[15] when the relevant nucleotide sequences were determined from both directions. The sequences of overlapping fragments were assembled to construct the complete primary structures of the circular RNA molecules.

Secondary structure models for the native ccRNAs were constructed using the base pairing matrix procedure of Tinoco *et al.*[16], values for the thermodynamic stability of predicted RNA structures (G. Steger, H. J. Gross, J. W. Randles, H. L. Sänger and D. Riesner, personal communication), and experimental evidence for the location of ribonuclease-sensitive single-stranded regions on the native molecules. In addition,

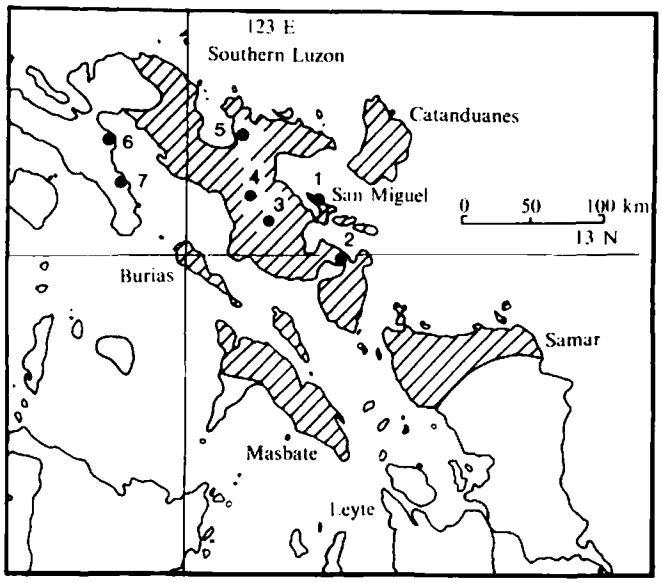

Fig. 1 Incidence of cadang-cadang disease of coconuts. Regions of the Philippine islands where the disease is widespread are shown cross-hatched, while isolated incidences of the disease occur in surrounding areas[32]. The different ccRNA isolates used in this work were each obtained from separate diseased coconut palms in one of the following locations: 1, San Miguel Island; 2, Sorsogon; 3, Ligao; 4, Lake Baao; 5, Tinambac; 6, Guinayangan; and 7, San Nasciso.

† Present address: National Health Institute, PO Box 50348, Porirua, New Zealand.

Table 1 Properties of RNAs of various ccRNA isolates

ccRNA isolate*	Relative proportions of sequence variants† C	CC	Total length of ccRNA (residues)	Length of sequence duplication (residues)
ccRNA 1 fast				
Baao 54	1.0	0	246	
Tinambac	1.0	0	246	
Ligao 14B	0.8	0.2	246/247‡	
Ligao 620C	0.6	0.4	246/247	
Ligao 191D	0.4	0.6	246/247	
Ligao T1	0.2	0.8	246/247	
ccRNA 2 fast				
Baao 54	1.0	0	492	
Ligao T1	0.2	0.8	492–494§	
ccRNA 1 slow				
Baao 54	1.0	0	287	41
Ligao 14B	1.0	0	296	50
Ligao 620C	1.0	0	296	50
Ligao 191D	1.0	0	296	50
Ligao T1	1.0	0	301	55
Ligao 5	1.0	0	296	50
Guinayangan	1.0	0	296	50
San Miguel	1.0	0	296	50
San Nasciso	0	1.0	297	50
ccRNA 2 slow				
Baao 54	1.0	0	574	41

* ccRNA isolates were purified[4] from nucleic acid extracts[33] of single, infected coconut palms.

† Relative proportions of sequence variants were determined by sequence analysis of RNase T_1-digested, 5′-^{32}P-radiolabelled full-length linear ccRNAs. If a ccRNA consisted of a mixture of variants, band doubling was observed on sequencing gels[10] after ccRNA 1 fast residue 197. Relative proportions of the two sets of band doublets were taken as estimates of the molar proportions of the two variant ccRNA species.

‡ These ccRNA 1 fast species consisted of a mixture of two species, one of 246 and the other of 247 residues.

§ Ligao T_1 ccRNA 2 fast species contained sequence heterogeneity at the positions corresponding to ccRNA 1 fast residue 198. Due to limitations of sequencing technique, we did not determine whether dimeric ccRNA 2 fast consisted of a 492(2×246)-residue species together with a 494(2×247)-residue species or only a 493(246+247)-residue species.

specific full-length linear ccRNAs, produced by ribonuclease T_1 cleavage, were either 5′- or 3′-labelled with ^{32}P and the single-strand regions in the native structures located by the nuclease S_1 mapping procedure[17]. The sites of cleavage were determined by co-electrophoresis of the radiolabelled fragments of the nuclease S_1 digest with products of sequencing reactions using the partial enzymatic digestion procedure[10].

ccRNAs differ in size but not sequence complexity

The sequences and predicted structures of the ccRNA 1 fast and slow forms isolated from a single infected coconut palm (isolate Baao 54) are shown in Fig. 1 together with the known structures of four viroids, potato spindle tuber viroid (PSTV)[18], CSV[10], CEV[12], and ASBV[11]. The two native ccRNAs 1 possess extensive regions of intramolecular base pairing and can form rod-like native structures similar to other viroids. The ccRNA 1 fast and ccRNA 1 slow possess 246 and 287 residues, respectively, and have calculated thermodynamic stabilities of -320 and -360 kJ mol^{-1}, respectively. ccRNA 1 slow contains the entire sequence and structure of the smaller ccRNA 1 fast but differs by an additional duplicated sequence and structure of 41 residues (residues 103–143 in ccRNA 1 fast) which is added at the right-hand end of the native molecule between residues 123 and 124 of ccRNA 1 fast (Fig. 2). Thus, the rod-like, base-paired native structure is maintained in the larger molecule.

The nucleotide sequences of ccRNA 2 fast and ccRNA 2 slow, consisting of 492 and 574 residues, respectively, are perfect dimers of the respective ccRNA 1 forms. A schematic summary of the relationships between the primary structures of all four ccRNAs and their predicted secondary structures is given in Fig. 3. While each of the monomeric ccRNA 1 forms can base pair intramolecularly to form a single rod-like conformer, the ccRNA 2 forms, due to their dimeric nature, can form either of two rod-like conformers (*A* or *B*, Fig. 3) and a large number of intermediate cruciform-shaped structures, one of which is given in Fig. 3.

The ccRNA 1 fast and ccRNA 1 slow molecules each possess, in experimental conditions, a highly accessible site for cleavage by ribonuclease T_1 at the right-hand terminal hairpin loop of the predicted native structure (between residues 124 and 125 in Baao 54 ccRNA 1 fast and between residues 145 and 146 in Baao 54 ccRNA 1 slow). Limited ribonuclease T_1 digestion of each ccRNA 2 species produced specific linear RNA fragments corresponding to both the respective full-length ccRNA 2 and ccRNA 1 molecules. Sequence determination of these fragments showed that cleavage of the ccRNA 2 molecules occurred at two sites located at the same sequences as for the two ccRNA 1 molecules. This suggests that either the predicted conformer A of the two ccRNA 2 molecules or possible cruciform intermediates exist in solution whereby the terminal hairpin loops are exposed. However, the existence of type B conformers cannot be precluded.

Variation in sequence between different ccRNA isolates

Different isolates of cadang-cadang RNAs were each obtained from single infected coconut palms from different localities in the Philippines (Fig. 1). Sequence differences between the different isolates consist of two types. First, the sequences of the ccRNA 1 slow forms can differ. While all ccRNA 1 fast forms are essentially identical (see below), ccRNA 1 slow forms can differ in the length of the repeated sequence inserted between residues 123 and 124 of ccRNA 1 fast. Three different repeated sequences found in nine sequenced isolates of ccRNA 1 slow are given in Fig. 4: these vary in length from 41 to 55 residues but they are internally base-paired to produce duplicated structures as well as sequences at the right-hand ends of the molecules. Interestingly, the right-hand ends of the native molecules of PSTV, CSV, CEV and ASBV (Fig. 2) are at a similar distance from the central conserved regions of these molecules. In contrast, the right-hand side of the ccRNA 1 fast molecule is shorter while those of the elongated ccRNA 1 slow molecules are closer in size to those of PSTV, CSV, CEV and ASBV.

Second, four of the six isolates of ccRNA 1 fast sequenced each consist of two populations of molecules, one of 246 residues and the other of 247 residues, which differ in the presence or absence of a C at residue 198 (Fig. 5, Table 1). Similar sequence heterogeneities have also been reported for CEV[19] and the viroid-like RNA of VTMoV[13]. The relative proportions of the two ccRNA 1 fast subspecies vary between different isolates as listed in Table 1. For the two ccRNA 2 fast isolates sequenced, the relative proportions of the two forms are the same as those of the corresponding ccRNA 1 fast. In contrast to the ccRNA 1 and 2 fast species, no similar sequence heterogeneity has been observed in nine isolates of ccRNA 1 slow and the one sequenced isolate of ccRNA 2 slow (Table 1). Each isolate of the ccRNA slow species thus consists entirely of either one subspecies or the other; in all except one case, the C at the position corresponding to ccRNA 1 fast residue 198 was absent. The various sequence differences between the ccRNA isolates do not seem to correlate with differences in geographical location.

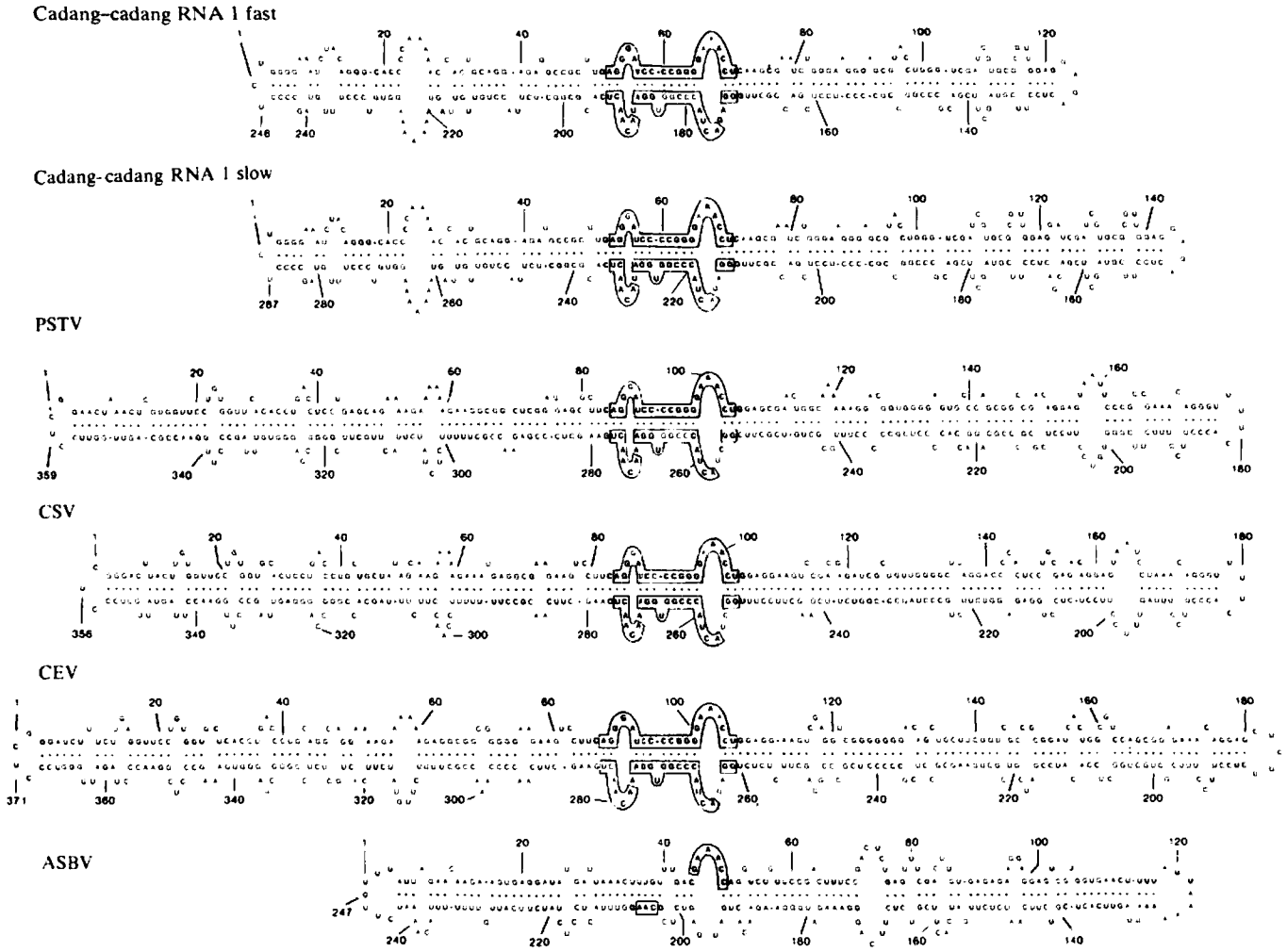

Fig. 2 Sequences and predicted secondary structures of the Baao 54 isolate of ccRNA 1 fast and ccRNA 1 slow are shown with those of PSTV[18], CSV[10], CEV[12] and ASBV[11]. The structures are aligned under the central conserved regions of these viroids (boxed). For sequence determination, nucleic acids were extracted from the fronds of a single infected coconut palm[13], and ccRNAs purified by preparative polyacrylamide gel electrophoresis[4]. Purified ccRNAs were sequenced essentially as described in ref. 10. Thus, 5 μg circular ccRNA in 20 μl 600 mM NaCl, 10 mM $MgCl_2$, at 0 °C was digested with RNase T_1 (Sigma) at 2,000 units ml^{-1} to produce specific full-length linear molecules, or with 3 units ml^{-1} RNase U_2 (Sankyo) or 5 ng ml^{-1} RNase A (Sigma) to produce smaller linear fragments. These fragments were 5'-radiolabelled using [γ-^{32}P]ATP and T4 polynucleotide kinase or, after treatment with calf intestinal phosphatase[34], 3'-radiolabelled using [3'-^{32}P]dpCp and T4 RNA ligase[14], and fractionated on a 80×20×0.05 cm 6% polyacrylamide gel containing 8 M urea and TBE buffer (90 mM Tris-borate *p*H 8.3, 2 mM EDTA). Bands located by autoradiography were excised and eluted[10], and sequenced using partial enzymatic cleavage[10]. The sequences of numerous overlapping fragments were assembled to give the complete primary structure of each circular molecule. For S_1 nuclease mapping of the secondary structures[17], 5'- or 3'-radiolabelled full-length linear fragments, obtained as described above by RNase T_1 digestion, were suspended in 20 μl 200 mM NaCl, 0.5 mM $ZnSO_4$, 50 mM sodium acetate *p*H 4.6, containing 5 μg *Escherichia coli* tRNA carrier, and incubated at 37 °C for 10 min with 0.1, 1 or 10 units of S_1 nuclease (Boehringer). The reaction mixture was extracted with phenol, the RNA precipitated with ethanol and fractionated by polyacrylamide gel electrophoresis in TBE, 8 M urea. Products of partial enzymatic sequencing reactions of the same ccRNA were run as markers, thus allowing sites of S_1 nuclease sensitivity to be located. Data so derived were used to construct the secondary structure models of ccRNA 1 fast and ccRNA 1 slow.

Structural similarities between ccRNAs, viroids and virusoids

The ccRNAs share two regions of sequence homology, each of about 20 nucleotides, with the viroids PSTV, CSV and CEV (Fig. 2). The latter three viroids are closely related, sharing about 50% sequence homology. These two conserved regions are base-paired in the predicted structures of the native molecules to form highly conserved secondary structures.

The conserved regions shared by the ccRNAs correspond to regions of PSTV, CSV and CEV postulated to be involved in base pairing of viroid complementary RNAs with a plant small nuclear RNA (snRNA) in a manner analogous to that proposed for the interaction of mRNA intron–exon splice junctions with mammalian Ula snRNA[19,20], and in the formation of a stabilizing stem-loop structure in the viroid complements[19]. The proposed interaction between viroid complements and snRNA is postulated to reflect the origin of viroids from an intron ancestor[20] or as a basis for pathogenesis[19,20] but not to be directly involved in viroid replication. However, the complementary RNA of at least one viroid, ASBV, is incapable of base pairing with a Ula-like snRNA or of formation of the conserved stem-loop structure, despite up to 18% sequence homology between ASBV and other viroids[11]. It is possible that the central conserved regions of native viroids reflect functional similarities related to viroid replication rather than to the postulated snRNA binding.

Interestingly, the virusoids (encapsidated viroid-like RNAs) of VTMoV[13], SNMV[13], and subterranean clover mottle virus (J.H., C. Davies and R.H.S., unpublished data) also contain highly conserved sequences which are central to their rod-like native structures and which share only the pentanucleotide sequence, GAAAC, with all sequenced viroids, including ASBV.

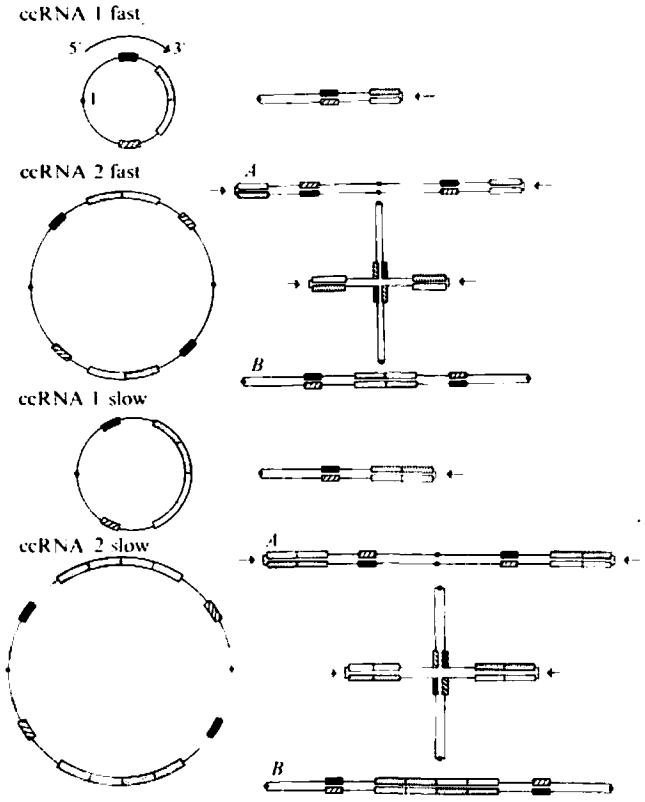

Fig. 3 Schematic representation of the sequences and predicted structure relationships between the ccRNAs. The circular sequences of the four ccRNAs are shown with black boxed and cross-hatched boxed regions representing the sequences highly conserved between the ccRNAs. PSTV, CSV and CEV (Baao 54 ccRNA 1 fast residues 52–71 and 171–194, respectively). The white boxed and stippled boxed regions represented those sequences duplicated within the ccRNA 1 slow species (Baao 54 ccRNA 1 fast residues 103–123 and 124–143, respectively). Positions corresponding to residue 1 of ccRNA 1 fast or ccRNA 1 slow are indicated by black dots. Both ccRNA 2 fast and ccRNA 2 slow molecules are dimers of their respective ccRNA 1 forms. Each ccRNA 2 can potentially form either of two rod-like conformers, *A* or *B*, as well as a large number of cruciform-shaped intermediates of which one is shown for ccRNA 2 fast and ccRNA 2 slow. Each ccRNA 1 species possesses a single, highly accessible site for RNase T_1 cleavage located on a hairpin loop (between Baao 54 ccRNA 1 fast residues 124 and 125 and between ccRNA 1 slow residues 145 and 146); these sites are indicated by arrows. Each ccRNA 2 species possesses two such accessible sites for RNase T_1 cleavage and arrows indicate where these sites also occur on hairpin loops in the different ccRNA 2 conformers.

Replication of ccRNAs

As PSTV, CSV, CEV and the ccRNAs are capable of autonomous replication, the enzymes involved are of considerable interest. However, no viroid-encoded translation products have been found *in vitro*[21,22] or *in vivo*[23]. Although PSTV, CSV and CEV share ~50% sequence homology, none of these viroids nor their putative complementary RNAs can theoretically encode similar translation products[10,12], even assuming the existence of translatable linear viroid RNAs *in vivo*[24,25]. Possible protein-coding regions similar to those of other viroids are not found in the ccRNAs or their complements nor are there any AUG initiation codons present. As a 5'-proximal AUG initiation codon seems essential for initiation of translation of eukaryotic RNA[26,27], it seems highly unlikely that the ccRNAs can code for any functional polypeptide product. All evidence, therefore, indicates that ccRNAs and other viroids must rely entirely on host components for their replication.

Larger than unit-length complementary (−) RNA intermediates have been detected in PSTV- and CEV-infected tissues[28–30], and recently an oligomeric series of RNAs of ASBV (+) have been detected in infected avocado tissue (G. Bruening, A. R. Gould, P. J. Murphy and R.H.S., unpublished results). Rolling circle mechanisms have been postulated for the synthesis of oligomeric complementary RNAs from circular viroid templates[28,30], and oligomeric viroid (+) sequences could be simply generated by transcription of multimeric (−) strand templates. Unit-length linear viroid produced by either specific transcription or cleavage of oligomeric viroid RNAs must be ligated to produce the final circular product. Such a model for viroid replication, involving oligomeric RNA intermediates, could readily account for the formation of the dimeric forms of both ccRNA 1 fast and ccRNA 1 slow; that is, ccRNA 2 fast and ccRNA 2 slow, respectively. Rate-limiting steps during the transcription or the possible processing of viroid transcripts would allow the dimeric ccRNAs 2 to accumulate over the monomeric ccRNA 1 species. The presence or absence of higher multimers of (+) ccRNAs or the complementary (−) ccRNAs in infected tissue has yet to be determined.

ccRNA slow variants and the time course of infection

In the initial stages of cadang-cadang disease, only the fast forms of ccRNA 1 and ccRNA 2 are present in infected palms and it is only after a further 24–30 months that the slow variants

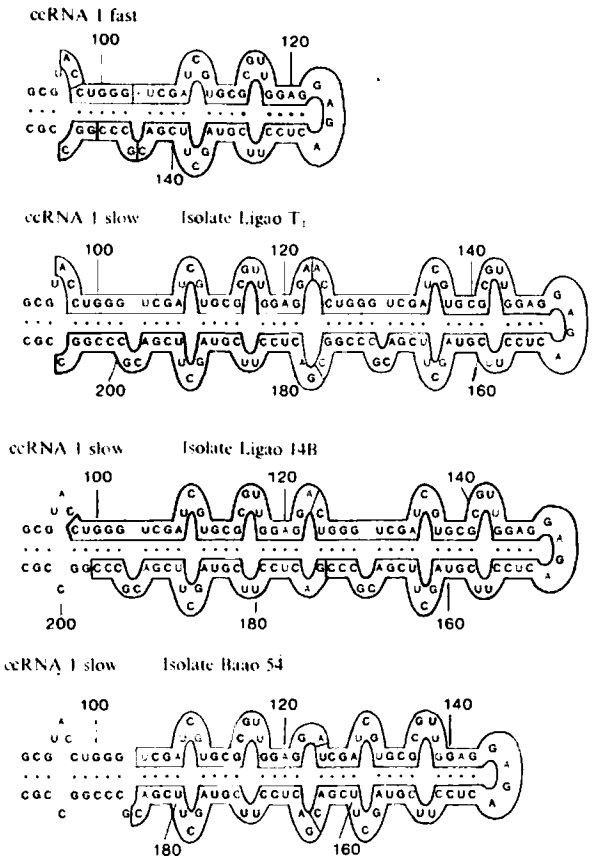

Fig. 4 Sequence variation between ccRNA 1 slow of three ccRNA isolates. The sequences and structures of various ccRNA 1 fast and ccRNA 1 slow isolates were determined as described in Fig. 2 legend. As essentially all sequence variation occurred at the right-hand end of the ccRNA 1 slow molecules, only this region is shown. Boxed regions represent those sequences which are duplicated in the ccRNA 1 slow molecules and which are 41 (isolate Baao 54), 50 (isolate Ligao 14B) or 55 residues (isolate Ligao T_1) long. All sequenced ccRNA 1 slow isolates correspond to one of these forms (Table 1).

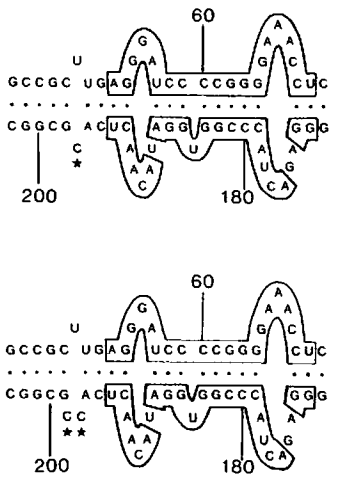

Fig. 5 Portions of the predicted native structures for the two forms of ccRNA 1 fast molecules. Purified ccRNA isolates were found to consist of either one or a mixture of two RNA species which differed in the presence or absence of an extra C residue at the position corresponding to ccRNA 1 fast residue 197 or 198. Stars indicate the sequence differences. Boxed regions indicate those sequences common to PSTV, CSV and CEV.

of ccRNA 1 and ccRNA 2 first appear and in the following years predominate[5]. These data, plus preliminary evidence that the ccRNA fast species are more infectious than the ccRNA slow species[4], are consistent with the *de novo* generation of the ccRNA slow variants during each cadang-cadang disease infection. This proposition is supported by the following sequence data.

(1) The ccRNA 1 slow forms differ from ccRNA 1 fast by the insertion of a single repeated sequence (Fig. 2) and could be simply generated from the ccRNA 1 fast by processing and/or transcription mechanisms.

(2) The ccRNA 1 slow isolates can differ in the size of their inserted sequence repeats (Fig. 4, Table 1) suggesting separate origins for these ccRNA slow variants.

(3) While most ccRNA fast isolates contain a sequence heterogeneity at residue 198, and consist of varying ratios of the 246- and 247-residue species, each of the nine sequenced ccRNA slow isolates consists of a single homogeneous population, either with or without a C residue at the position homologous to ccRNA 1 fast residue 198, and with only one size of repeated sequence (Table 1).

These data are consistent with the generation of ccRNA slow forms from ccRNA fast by single, rare sequence duplication events occurring separately in each cadang-cadang-infected palm. All ccRNA slow molecules would, therefore, originate from single parent molecules and may accumulate in preference to ccRNA fast species due to a competitive advantage in replication.

Origin of cadang-cadang disease

The ccRNAs share biological properties and sequence and structural homology with viroids so that application of the term coconut cadang-cadang viroid is fully justified. However, whereas other viroids consist of a single predominant infectious RNA species, CCCV consists of several variant RNA species. It is feasible that CCCV may have arisen from a pre-existing viroid and that mutation or infection of new hosts, such as the coconut palm and related species[31], resulted in production of the variant ccRNAs by abberant transcription and/or processing mechanisms which normally occur faithfully in the replication of other viroids. The outbreak and subsequent apparent rapid spread of cadang-cadang disease in the Philippines this century[32] is consistent with such an origin of the ccRNAs. Future work with coconut palms inoculated with purified ccRNA species should allow a study of the generation of the ccRNA variants and may help elucidate the nature of viroid replication.

This work was supported by the Australian Research Grants Scheme and the UNDP/FAO Coconut Research Project. We thank Dr B. Zelazny for advice, Dr D. Riesner for unpublished data, Dr A. J. Gibbs for suggesting the term virusoids for encapsidated viroid-like RNAs, and Jenny Rosey, Lisa Waters and Julita Imperial for assistance.

Received 28 June; accepted 16 August 1982.

1. Ocfemia, G. P. *Philipp. Agric.* **26**, 338–340 (1937).
2. Bigornia, A. E. & Phillipp, J. *Coconut Stud.* **2**, 5–33 (1977).
3. Boccardo, G., Beaver, R. G., Randles, J. W. & Imperial, J. S. *Phytopathology* **71**, 1104–1107 (1981).
4. Imperial, J. S., Rodriguez, J. B. & Randles, J. W. *J. gen. Virol.* **56**, 77–85 (1981).
5. Mohamed, N. A., Haseloff, J., Imperial, J. S. & Symons, R. H. *J. gen. Virol.* (in the press).
6. Mohamed, N. A. & Imperial, J. S. (in preparation).
7. Randles, J. W., Rillo, E. P. & Diener, T. O. *Virology* **74**, 128–139 (1976).
8. Randles, J. W. & Hatta, T. *Virology* **96**, 47–53 (1979).
9. Gross, H. J. & Riesner, D. *Angew. Chem. int. Edn.* **19**, 231–243 (1980).
10. Haseloff, J. & Symons, R. H. *Nucleic Acids Res.* **9**, 2741–2752 (1981).
11. Symons, R. H. *Nucleic Acids Res.* **9**, 6527–6537 (1981).
12. Viswader, J. E., Gould, A. R., Bruening, G. E. & Symons, R. H. *FEBS Lett.* **137**, 288–292 (1982).
13. Haseloff, J. & Symons, R. H. *Nucleic Acids Res.* **12**, 3681–3691 (1982).
14. England, T. E. & Uhlenbeck, O. C. *Nature* **275**, 560–561 (1978).
15. Kramer, F. R. & Mills, D. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5334–5338 (1978).
16. Tinoco, I., Uhlenbeck, O. C. & Levine, M. D. *Nature* **230**, 362–367 (1971).
17. Wurst, R. M., Vournakis, J. N. & Maxam, A. M. *Biochemistry* **17**, 4493–4499 (1978).
18. Gross, H. J. *et al. Nature* **273**, 203–208 (1978).
19. Gross, H. J. *et al. Eur. J. Biochem.* **121**, 249–257 (1982).
20. Diener, T. O. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5014–5015 (1981).
21. Davies, J. W., Kaesberg, P. & Diener, T. O. *Virology* **61**, 281–286 (1974).
22. Semancik, J. S., Conjero, V. & Gerhart, J. *Virology* **80**, 218–221 (1977).
23. Conjero, V. & Semancik, J. S. *Virology* **77**, 221–232 (1977).
24. Kozak, M. *Nature* **280**, 82–85 (1979).
25. Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. *Eur. J. Biochem.* **114**, 221–227 (1981).
26. Baralle, F. E. & Brownlee, G. G. *Nature* **274**, 84–87 (1978).
27. Kozak, M. *Cell* **15**, 1109–1123 (1978).
28. Branch, A. D., Robertson, H. D. & Dickson, E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6381–6385 (1981).
29. Rohde, W. & Sänger, H. L. *Biosci. Rep.* **1**, 327–336 (1981).
30. Owens, R. A. & Diener, T. O. *Proc. natn. Acad. Sci. U.S.A.* **79**, 113–117 (1982).
31. Randles, J. W., Boccardo, G. & Imperial, J. S. *Phytopathology* **70**, 185–189 (1980).
32. Zelazny, B. *Acta phytopathol. Acad. Sci. Hung.* **14**, 115–126 (1979).
33. Randles, J. W. *Phytopathology* **65**, 163–167 (1975).
34. Efstratiadis, A. *et al. Nucleic Acids Res.* **4**, 4165–4174 (1977).