Journal of General Virology (2003), **84**, 751–756 DOI 10.1099/vir.0.18755-0

Short Communication

Identification of neutral mutants surrounding two naturally occurring variants of *Potato spindle tuber viroid*

R. A. Owens,[1] S. M. Thompson[1] and M. Kramer[2]

Molecular Plant Pathology Laboratory[1] and Biometrical Consulting Service[2], USDA/ARS, Beltsville, MD 20705, USA

Correspondence
Robert Owens
owensr@ba.ars.usda.gov

Received 8 August 2002
Accepted 25 October 2002

Single point mutations in the pathogenicity domain of *Potato spindle tuber viroid* (PSTVd) can have a dramatic effect on disease expression, and only three substitutions are required for the spontaneous conversion of the type strain PSTVd-Intermediate to the rapidly replicating, highly pathogenic variant RG1 (Gruner *et al*., *Virology* **209**, 60–69, 1995). To identify available evolutionary pathways linking these two variants, we mutagenized five positions in an infectious cDNA copy of PSTVd-Intermediate and screened the resulting mixture of 768 sequences for neutral or near-neutral mutants. Numerical simulations based on the bioassay data indicate that the 23 variants recovered represent >80 % of all such sequences. RG1 was the only naturally occurring variant recovered, and the overall pattern of sequence changes observed indicates that PSTVd-Int occupies a comparatively steep peak within the fitness landscape.

Viroids are the smallest autonomously replicating pathogens yet described – small (246–399 nt), unencapsidated, circular RNAs that lack mRNA activity yet are able to induce a full range of disease symptoms in susceptible host plants (Flores *et al*., 1997; Owens, 1999). Like other RNA genomes, viroids propagate *in vivo* as quasispecies, populations of similar but non-identical sequences resulting from the inherently high error rate of RNA replication (Eigen, 1993). Evolution of viroid and viral satellite RNA sequences appears to be constrained by the need to preserve certain secondary structural features (Fraile & Garcia-Arenal, 1991; Fontana & Schuster, 1998; Ambrós *et al*., 1998); nevertheless, cloned *Potato spindle tuber viroid* (PSTVd) variants often give rise to complex populations of variants within a single passage (Góra-Sochacka *et al*., 1997). Such populations contain all possible single and double mutants as well as many triple mutants.

Previous studies of viroid sequence variation have addressed several aspects of host–pathogen interaction. Differences in overall sequence homology have been used to divide isolates of *Hop stunt viroid* (HSVd) into host-specific groups (Shikata, 1990; Kofalvi *et al*., 1997; Sano *et al*., 2001; Amari *et al*., 2001). Co-variation in nucleotide sequence provides evidence for the role of alternative structures such as secondary hairpins I (Polivka *et al*., 1996) and II (Loss *et al*., 1991; Qu *et al*., 1993) in viroid replication. A recent study by Matousek *et al*. (2001) provides insight into the role of environmental stresses in generating viroid sequence diversity. These authors showed that heat treatment of hop plants infected with *Hop latent viroid* (HLVd) results in a rapid decrease in viroid titre and a dramatic increase in sequence variability. The prevalence of HLVd variants containing three or four changes suggests that mutations accumulate during successive replication cycles. For genomes like that of influenza virus A where certain codons are subject to positive selection, phylogenetic methods can be used to predict the future course of evolution (Bush *et al*., 1999; Farci *et al*., 2000).

To better understand the selective pressures guiding viroid evolution *in vivo*, we have examined the fitness landscape surrounding two well-characterized, naturally occurring strains of PSTVd. PSTVd is the type member of the genus *Pospiviroid* whose members possess a highly base-paired, rod-like secondary structure. RG1 is a highly pathogenic variant that appeared spontaneously during repeated passage of PSTVd-Int(ermediate) in tomato (Gruner *et al*., 1995). PSTVd-Int and RG1 differ at only three positions within a portion of the molecule known as the pathogenicity domain. Our results indicate that these variants are surrounded by a network of neutral or near-neutral mutants that allows the population to move from peak to peak in the fitness landscape.

Fig. 1 compares the pattern of sequence variation among more than 40 natural isolates of PSTVd (panel A) with that introduced by oligonucleotide-directed randomization of the pathogenicity domain of PSTVd-Int (panel B). Natural variability is concentrated in the lower portion of the pathogenicity domain as well as positions 120 and 123 of the variable domain. To investigate the feasibility of screening large libraries of viroid sequences for variants with altered biological properties, we randomized five

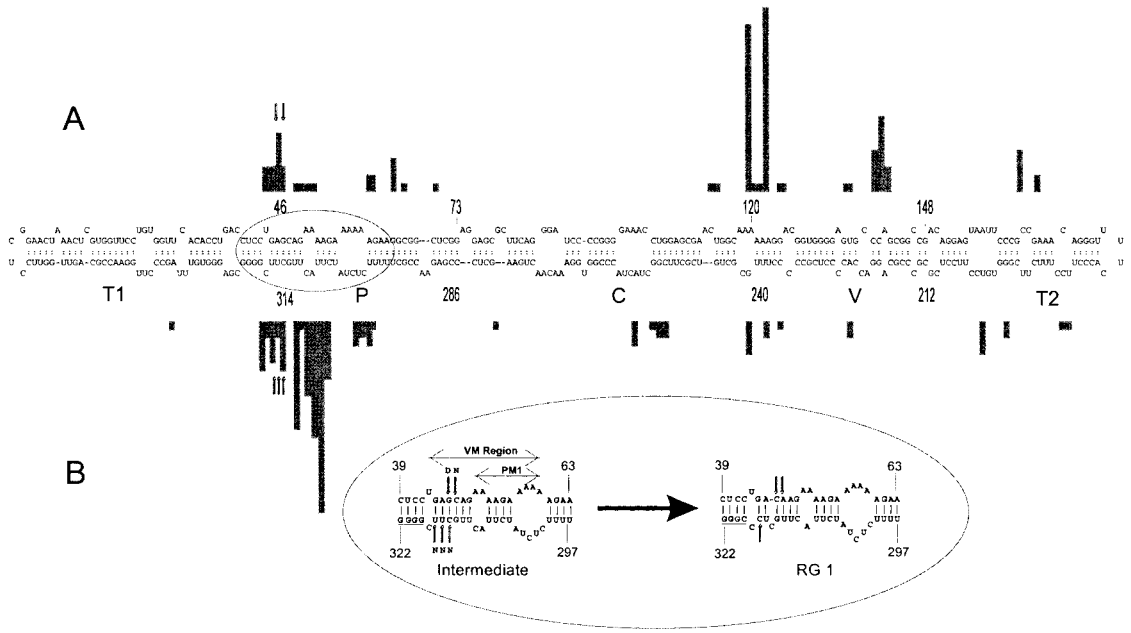

Fig. 1. PSTVd sequence variability. (A) Variation among 48 naturally occurring isolates. Bar heights are proportional to the number of isolates differing from PSTVd-Int (Gross *et al.*, 1978) at individual positions. Numbers indicate boundaries of the five structural/functional domains proposed by Keese & Symons (1985): terminal left (T1), pathogenicity (P), central conserved (C), variable (V) and terminal right (T2) domains. Vertical arrows indicate nucleotide positions mutagenized by oligonucleotide-directed randomization (Reidhaar-Olson & Sauer, 1988) in the present study. (B) Pathogenicity domains of PSTVd-Int and RG1. Horizontal arrows indicate locations of pre-melting loop 1 (PM1) and the virulence-modulating (VM) region as described by Schnölzer *et al.* (1985).

positions in PSTVd-Int. PSTVd-Int and RG1 differ at positions 46, 47 and 317; positions 315 and 316 were also mutagenized to produce a more complex mixture of variants. Wild-type viroid was eliminated from the potential progeny by replacing the G residue at position 46 with an equimolar ratio of A, C and T, resulting in an equimolar mixture of $3 \times 4 \times 4 \times 4 \times 4 = 768$ sequences. Three bioassays were carried out using RNA transcripts synthesized from independent cDNA template preparations to inoculate either Rutgers or Microtom (Meissner *et al.*, 1997) tomato seedlings. Infected plants were identified by dot blot hybridization (Podleckis *et al.*, 1993) and the progeny from 50 individual plants was characterized by RT-PCR (Hu *et al.*, 1996). By examining uncloned PCR products rather than individual cloned cDNAs, we would expect to detect only the predominant PSTVd variant(s) in each infected plant.

Table 1 summarizes the results from automated sequence analysis of uncloned PCR products. In experiment 1, eight variants were recovered from eleven infected plants; i.e. three variants were present in more than one plant. There appeared to be a general tendency for CA to replace GC at positions 46–47, and spontaneous mutations were detected at several nearby positions. These may represent compensatory mutations that offset the deleterious effects of mutations at the targeted positions (Matousek *et al.*, 2001). The ID_{50} for unmutagenized PSTVd-Int was approximately 1 pg μl^{-1}, while that of the mutagenized transcript mixture appeared to be >10-fold higher (results not shown). Variant M3/M1 was previously described by Hammond (1992).

Progeny from the second bioassay included four other variants in addition to M3/M1. One plant contained a mixture of RG1 and M1, another variant first described by Hammond (1992). These two variants differ by only a single U→C change at position 317, suggesting that M1 may represent an evolutionary intermediate between PSTVd-Int and RG1. The most striking result from this experiment, however, was the presence of M1 in ten separate plants. This was a most improbable outcome, and sequence analysis of this particular cDNA template preparation subsequently revealed a marked overrepresentation of C at position 46 as well as very low levels of substitution at the other mutagenized positions (results not shown). No such bias was detected in the template preparations used in the two other bioassays.

The 24 infected plants analysed in the third bioassay contained 16 different PSTVd variants. Variant M24/M7 was recovered from three different plants and seven other variants were recovered twice. Taken together, our three bioassays yielded 23 viable variants of which 20 were previously unknown. Because relatively few of the 768 possible variants were recovered (and several more than once), the total number of neutral mutants seemed likely to be small. Because all but two of the 50 plants analysed

Table 1. PSTVd variants recovered from randomized inocula

Newly described variants deposited in GenBank (accession numbers AF458986–459004).

		Nucleotide positions†									
Variant(s)	**Plants***	**46**	**47**	**60**	**304**	**305**	**314**	**315**	**316**	**317**	**359**
PSTVd-Int		G	C	A	U	A	G	C	U	U	U
Experiment 1											
M33/M1	2	C	A					U	A	C	
M12/M1	2	C	A					A		C	
M3/M1	2	C	A					U		C	
M34/M1	1	C	A					U	G	C	
M23/M7	1	C						G	A	C	
M3/M6a	1	A	A				U	U		C	
M31/M5a	1	A		−A‡				U			
M31/M5b	1	A			C	U		U			
Experiment 2											
M1	10	C	A								
M32/M1a	1	C	A					U	A		A
M31/M5	2	A						U			
M3/M1	1	C	A					U		C	
M1 + RG1	1	C	A							[U/C]§	
Experiment 3											
M22/M7 + M7a	1	C		[A/−A]§				G	A	A	
M24/M7	3	C						G		A	
M25/M7	2	C						G	G	A	
M2/M5	2	A						U			
M11/M8	1	U						A			
M31/M5a	1	A		−A‡				U			
M12/M8	2	U						A		C	
M33/M1	2	C	A					U	A	C	
M12/M1	2	C	A					A		C	
M13/M1	2	C	A					A	A	C	
M3/M1	2	C	A					U		C	
M21/M1	1	C	A					G	A		
M23/M1	1	C	A					G	A	C	
M14/M1	1	C	A					A	G	C	
RG1	1	C	A							C	

*Number of plants containing individual sequence variants.

†Location of mutations – either targeted (positions 46, 47, 315–317) or spontaneous.

‡Nucleotide deleted.

§Heterogeneous nucleotide.

contained only a single variant, their competitive abilities appeared to be ordered. Viroid symptom expression is almost completely suppressed in dwarf tomato varieties like Microtom; thus, the relative severity of each of these variants could not be directly assessed.

To estimate the true number of neutral mutants, we simulated the inoculation and subsequent infection processes based on the following assumptions: (i) variants are initially present in equal numbers in the inoculum; (ii) the number of variants present in an inoculum is a sample from a Poisson distribution (i.e. variants are distributed randomly in the solution and the Poisson parameter depends on solution concentration); and (iii) the competitive ability of variants is ordered (i.e. if two variants are present in the same inoculum, only the most competitive will later be detected). *Sim2*, written in Perl, allows one to specify the number of simulations, number of plants used, true number of infectious variants and proportion of uninfected plants [by setting the Poisson parameter to −ln (proportion of uninfected plants)]. For each simulation, the program (i) creates an inoculum, (ii) draws from the inoculum and applies it to the plant and (iii), if the plant becomes infected, identifies the most competitive variant. For each true

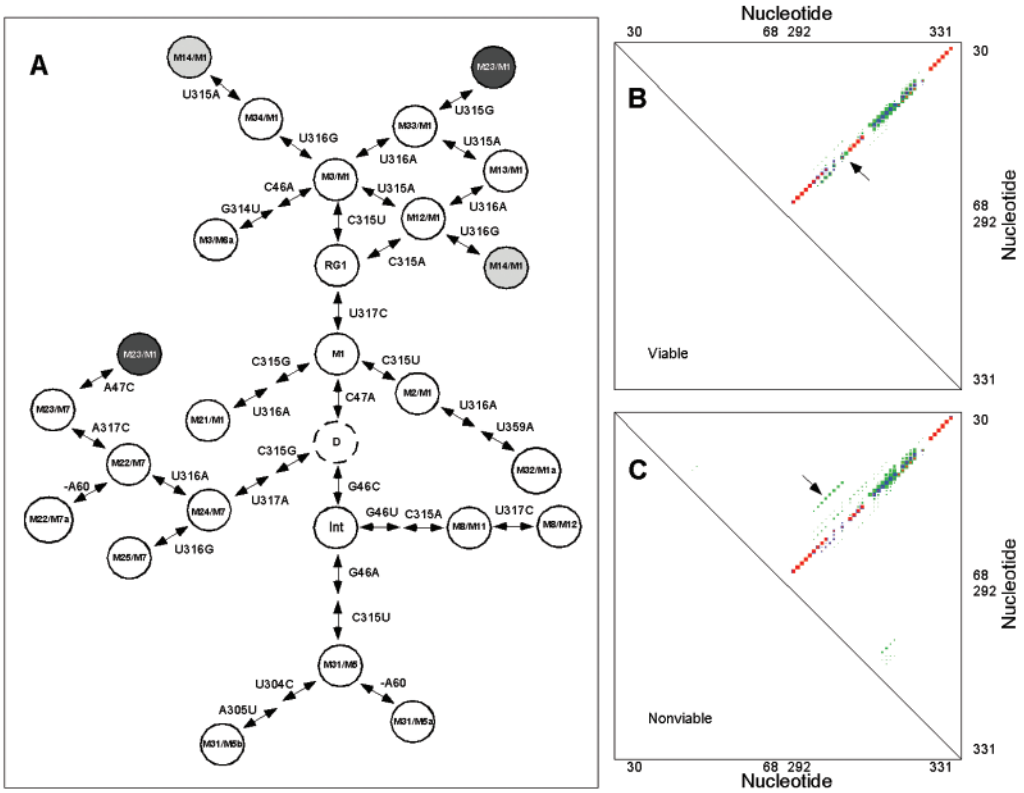

Fig. 2. Sequence and structural relationships among PSTVd sequence variants. (A) Two-dimensional network connecting infectious variants generated by randomization/*in vivo* selection. Note that M14/M1 and M23/M1 (shaded) are each connected to two other variants by only single nucleotide changes. PSTVd-D is a nonviable variant described by Hammond (1992). (B and C) Computer-predicted secondary structures of viable variants including PSTVd-Int and RG1 (B) and 25 randomly selected noninfectious variants (C). Red squares represent base pairs present in all sequences in the data set, and blue squares denote base pairs present in PSTVd-Int. Green squares denote base pairs present in sequences other than PSTVd-Int. Smaller squares indicate less probable base pairings; i.e. suboptimal structures. The nine red squares at the lower left corner of the diagonal denote a 9 bp helix involving nucleotides 60–68 and 292–300; the six red squares in the upper right corner denote a helix at the left side of the pathogenicity domain containing nucleotides 30–35 and 326–331 (see Fig. 1).

number of infectious variants in the range producing results similar to our data, we ran 1000 simulations. When we set the true number of infectious variants at 28, the results matched those of our largest data set (experiment 3) exactly. In this data set, inoculation of 24 tomato seedlings with 1000 pg mutagenized RNA transcripts μl^{-1} resulted in 14 infected plants containing 11 different variants. For these parameters and with the true number of viable sequences set to 28, *Sim2* calculated the mean number of different variants to be $11{\cdot}0 \pm 1{\cdot}92$. *Sim2* also allowed us to estimate the power associated with various experimental designs. We found that high infection rates and a sample size of $\geqslant 100$ plants is necessary for a precise estimate of the true number of neutral mutants.

Possible relationships among mutagenesis-derived and naturally occurring PSTVd sequence variants were analysed using a combination of neighbour-joining and maximum-parsimony (Swofford, 1998) and split decomposition (Huson, 1998) methods. The results indicated the presence of networks rather than bifurcating trees. As shown in Fig. 2(A), noteworthy features of these networks included (i) the complex relationships between certain variants and (ii) the relatively isolated position occupied by PSTVd-Int. In many cases, conversion of one variant to another requires only a single substitution. For example, only a single substitution at position 315 separates RG1 from its two nearest neighbours, and M14/M1 and M23/M1 can each be derived from at least two other variants by single nucleotide substitutions. All three pathways leading away from PSTVd-Int, in contrast, begin with a double change. A two-dimensional representation can depict only the lowest transition orders of such a complex network.

Because so few mutagenesis-derived PSTVd variants were detected, it seemed possible that the competitive sequences might share some common structural feature(s). To address this possibility, we compared the secondary structures of

their pathogenicity domains using ConStruct, a software tool developed by Luck *et al.* (1999). Fig. 2(B) shows the optimal and suboptimal structures for PSTVd-Int, RG1 and the 23 nearby neutral mutants; Fig. 2(C) shows the corresponding structures for 25 randomly chosen noncompetitive sequences. Overall, the two structure distributions were very similar. The two additional base pairs visible on the diagonal in Fig. 2(B) were present in only a single neutral mutant, M31/M5b. Likewise, the eight base pairs above the main diagonal in Fig. 2(C) were present in only one noncompetitive sequence. In PSTVd-Int, nucleotides 46 and 315 are predicted to form a G : C pair in the 6 bp helix visible in the upper right hand corner of Fig. 2(B). Compensatory mutations in variants M11/M8 and M31/M5 result in the conversion of this G : C pair into either a U : A or an A : U pair, but neither change is predicted to have significant structural effects. No structural features unique to competitive variants were obvious.

To describe the informational complexity of RNA and DNA genomes, Eigen introduced the concept of a multidimensional 'sequence space' (Eigen & Biebricher, 1988; Eigen, 1993). Sequences differing at one to *n* positions (where *n* is the total number of nucleotides in the molecule) occupy positions that are located, respectively, one to *n* units from each other. Collapsing this multidimensional space to two dimensions and assigning a fitness value to each sequence results in a fitness landscape consisting of peaks connected by ridges and separated by valleys or planes. Implicit in the *n*-dimensional nature of sequence space is the possibility that sequence changes in distant portions of an RNA molecule may be functionally linked. PSTVd-Int and RG1 differ at only three positions, and randomization of two additional positions could open evolutionary pathways to other previously described variants. If PSTVd-Int and RG1 were located on a flat, comparatively featureless portion of the fitness landscape, one might expect the number of neutral mutants to approach the theoretical population size of 768. The much lower estimate derived from our data (i.e. 28) together with the scarcity of variants containing only a single nucleotide change provides further evidence for a rugged topology surrounding PSTVd-Int. In contrast to the comparatively isolated position of PSTVd-Int, RG1 is separated from its three closest neighbours by only a single nucleotide change.

Although most studies have focused on the diseases they cause, viroid infection can also have beneficial effects. For example, viroid infection can dwarf citrus growing on certain rootstocks, thereby increasing productivity and lowering production costs (reviewed by Hutton *et al.*, 2000). PCR-mediated 'DNA shuffling' has been used to balance the complex viral functions necessary for gene therapy and vaccine applications (Soong *et al.*, 2000), and we are investigating the use of a similar mutagenesis-selection strategy to produce improved *Citrus viroid III* (CVd-III) dwarfing agents (Owens *et al.*, 2003). In preliminary studies, limited sequence randomization resulted in dramatic changes in biological properties. Provided that the appropriate genome regions for mutagenesis can be identified, our results suggest that randomization/selection will yield improved citrus dwarfing agents and other viroid variants with specific biological properties.

REFERENCES

Amari, K., Gomez, G., Myrta, A., Di Terlizzi, B. & Pallás, V. (2001). The molecular characterization of 16 new sequence variants of *Hop stunt viroid* reveals the existence of invariable regions and a conserved hammerhead-like structure on the viroid molecule. *J Gen Virol* **82**, 953–962.

Ambrós, S., Hernández, C., Desvignes, J. C. & Flores, R. (1998). Genomic structure of three phenotypically different isolates of peach latent mosaic viroid: implication of the existence of constraints limiting the heterogeneity of viroid quasispecies. *J Virol* **72**, 7397–7406.

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. (1999). Predicting the evolution of human influenza A. *Science* **286**, 1921–1925.

Eigen, M. (1993). The origin of genetic information: viruses as models. *Gene* **135**, 37–47.

Eigen, M. & Biebricher, C. K. (1988). Sequence space and quasispecies distribution. In *RNA Genetics*, vol. III, *Variability of RNA Genomes*, pp. 211–245. Edited by E. Domingo, J. J. Holland & P. Ahlquist. Boca Raton, FL: CRC Press.

Farci, P., Shimoda, A., Coiana, A. & 9 other authors (2000). The outcome of acute hepatitis C predicted by evolution of the viral quasispecies. *Science* **288**, 339–344.

Flores, R., Di Serio, F. & Hernández, C. (1997). Viroids: the noncoding genomes. *Semin Virol* **8**, 65–73.

Fontana, W. & Schuster, P. (1998). Continuity in evolution: on the nature of transitions. *Science* **280**, 1451–1455.

Fraile, A. & Garcia-Arenal, F. (1991). Secondary structure as a constraint on the evolution of a plant viral satellite RNA. *J Mol Biol* **221**, 1065–1069.

Góra-Sochacka, A., Kierzek, A., Candresse, T. & Zagórski, W. (1997). The genetic stability of potato spindle tuber viroid (PSTVd) molecular variants. *RNA* **3**, 68–74.

Gross, H. J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H. & Sänger, H. L. (1978). Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature* **273**, 203–208.

Gruner, R., Fels, A., Qu, F., Zimmat, R., Steger, G. & Riesner, D. (1995). Interdependence of pathogenicity and replicability with potato spindle tuber viroid. *Virology* **209**, 60–69.

Hammond, R. W. (1992). Analysis of the virulence modulating region of potato spindle tuber viroid (PSTVd) by site-directed mutagenesis. *Virology* **187**, 654–662.

Hu, Y., Feldstein, P. A., Bottino, P. J. & Owens, R. A. (1996). Role of the variable domain in modulating potato spindle tuber viroid replication. *Virology* **219**, 45–56.

Huson, D. H. (1998). SplitsTree: analysing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.

Hutton, R. J., Broadbent, P. & Bevington, K. B. (2000). Viroid dwarfing for high density citrus plantings. *Hortic Rev* **24**, 277–317.

Keese, P. & Symons, R. H. (1985). Domains in viroids: evidence of intermolecular RNA rearrangements and their contribution to viroid evolution. *Proc Natl Acad Sci U S A* **82**, 4582–4586.

Kofalvi, S. A., Marcos, J. F., Cañizares, M. C., Pallás, V. & Candresse, T. (1997). Hop stunt viroid (HSVd) sequence variants

from *Prunus* species: evidence for recombination between HSVd isolates. *J Gen Virol* **78**, 3177–3186.

Loss, P., Schmitz, M., Steger, G. & Riesner, D. (1991). Formation of a thermodynamically metastable structure containing hairpin II is critical for infectivity of potato spindle tuber viroid RNA. *EMBO J* **10**, 719–727.

Luck, R., Graf, S. & Steger, G. (1999). ConStruct: a tool for thermodynamic controlled prediction of conserved secondary structure. *Nucleic Acids Res* **27**, 4208–4217.

Matousek, J. A., Patzak, J., Orctová, L., Schubert, J., Vrba, L., Steger, G. & Riesner, D. (2001). The variability of hop latent viroid as induced upon heat treatment. *Virology* **287**, 349–358.

Meissner, R., Jacobson, Y., Melamed, S., Levyatuv, S., Shalev, G., Ashri, A., Elkind, Y. & Levy, A. (1997). A new model system for tomato genetics. *Plant J* **12**, 1465–1472.

Owens, R. A. (1999). Viroids. In *Encyclopedia of Virology*, 2nd edn, pp. 1928–1937. Edited by A. Granoff & R. G. Webster. New York: Academic Press.

Owens, R. A., Thompson, S. M., Sieburth, P. J. & & Garnsey, S. M. (2003). Limited sequence randomization: testing a strategy to produce improved citrus viroid dwarfing agents. In *Proceedings 14th Conference International Organization of Citrus Virologists*. Riverside,CA: IOCV (in press).

Podleckis, E. V., Hammond, R. W., Hurtt, S. W. & Hadidi, A. (1993). Chemiluminescent detection of potato and pome fruit viroids by digoxygenin-labelled dot blot and tissue blot hybridization. *J Virol Methods* **43**, 147–158.

Polivka, H., Staub, U. & Gross, H. J. (1996). Variation of viroid profiles in individual grapevine plants: novel grapevine yellow speckle viroid 1 mutants show alterations of hairpin I. *J Gen Virol* **77**, 155–161.

Qu, F., Heinrich, C. Loss P., Steger, G., Tien, P. & Riesner, D. (1993). Multiple pathways of reversion in viroids for conservation of structural elements. *EMBO J* **12**, 2129–2139.

Reidhaar-Olson, J. F. & Sauer, R. T. (1988). Combinatorial cassette mutagenesis as a probe for the informational content of protein sequences. *Science* **241**, 53–57.

Sano, T., Mimura, R. & Ohshima, K. (2001). Phylogenetic analysis of hop and grapevine isolates of hop stunt viroid supports a grapevine origin for hop stunt disease. *Virus Genes* **22**, 53–59.

Schnölzer, M., Haas, B., Ramm, K., Hofmann, H. & Sänger, H. L. (1985). Correlation between structure and pathogenicity of potato spindle tuber viroid. *EMBO J* **4**, 2181–2190.

Shikata, E. (1990). New viroids from Japan. *Semin Virol* **1**, 107–115.

Soong, N. W., Nomura, L., Pekrun, K., Reed, M., Sheppard, L., Dawes, G. & Stemmer, W. P. (2000). Molecular breeding of viruses. *Nat Genet* **25**, 436–439.

Swofford, D. L. (1998). PAUP*. Phylogenetic analysis using parsimony (*and other methods) – version 4. Sunderland, MA: Sinauer Associates.