AoB PLANTS http://aobplants.oxfordjournals.org/

Open access – Research article

Detection and analysis of endogenous badnaviruses in the New Zealand flora

David J. Lyttle, David A. Orlovich and Paul L. Guy*

Department of Botany, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand

Received: 30 November 2010; **Returned for revision:** 31 January 2011; **Accepted:** 8 March 2011; **Published:** 15 March 2011

Citation details: Lyttle DJ, Orlovich DA, Guy PL. 2011. Detection and analysis of endogenous badnaviruses in the New Zealand flora. *AoB PLANTS* 2011 plr008 doi:10.1093/aobpla/plr008.

Abstract

Background and aims Badnaviruses and their host-integrated DNA occur in tropical crops and a few northern temperate species. Following the discovery of a badnavirus on a subantarctic island with floristic links to New Zealand, we postulated that badnaviruses exist in the New Zealand flora. Badnavirus reverse transcriptase (RT) sequences consist of variable regions flanked by highly conserved regions. This study used RT sequences to detect and characterize badnavirus sequences in the New Zealand flora and to investigate their utility for the study of broader aspects of plant biology.

Methodology Molecular diversity of RT sequences was analysed using polymerase chain reaction and denaturing gradient gel electrophoresis (DGGE). In a study of the genus *Melicytus*, internal transcribed spacer (ITS) sequences were compared with the RT data.

Principal results No freely replicating badnaviruses were detected but more than half of the species (37/60) contained RT sequences. Phylogenetic analysis of 21 RT sequences formed monophyletic groups distinct from other species and from badnaviruses. No frameshift mutations occurred in any of the sequences translated *in silico*. More detailed study of the genus *Melicytus* indicated broader applications for our approach. Analysis of RT sequences revealed the presence of a previously unrecognized species (confirmed using ITS). Inheritance of DGGE profiles by *Melicytus ramiflorus* seedlings suggested that this species may undergo apomixis.

Conclusions The presence of integrated badnavirus sequences in a wide range of taxa from this Southern Hemisphere flora indicates that these sequences may be common in many temperate regions. Potential to activate viruses from these sequences should be considered when placing these species in tissue culture or under other forms of abiotic or genomic stress. Analysis of endogenous RT sequences shows potential for the study of systematics, phylogenetics and plant reproductive biology.

Introduction

Members of the plant virus family *Caulimoviridae* possess a circular double-stranded DNA genome of ~7–8 kb. *Caulimoviridae* are pararetroviruses as they replicate through an obligatory RNA intermediate but, unlike animal retroviruses, do not need to integrate into the host genome to complete their replication cycle. The *Caulimoviridae* is divided into six genera: *Caulimovirus*, *Soymovirus*, *Cavemovirus*, *Petuvirus*, *Badnavirus* and *Tungrovirus* (Fauquet *et al.* 2005). Three of these genera

* Corresponding author's e-mail address: paul.guy@botany.otago.ac.nz

have been shown to exist as integrated forms that can, under conditions of genomic or environmental stress, excise from the host genome and generate episomal virus infection (Harper *et al.* 2002; Hansen and Heslop-Harrison 2004). Such endogenous pararetroviruses (EPRVs) form a distinct class of retroelements and have been found in a variety of crop and ornamental plants. For example, banana contains, and taro and yucca most likely contain, integrated badnavirus sequences (Clover *et al.* 2003; Yang *et al.* 2003; Geering *et al.* 2005*a*, *b*; Gambley *et al.* 2008) as well as freely replicating, transmissible, badnaviruses. Eudicots such as *Citrus*, *Theobroma*, *Ribes* and *Rubus* contain badnaviruses, but it is not known whether they contain integrated badnavirus sequences. A number of eudicots do, however, contain integrated forms of other genera of *Caulimoviridae* (Hohn *et al.* 2008).

Although a few badnaviruses have been isolated from cultivated plants from northern temperate regions, badnaviruses are generally considered to be of tropical or subtropical origin. Recently a new species of badnavirus, *Stilbocarpa mosaic bacilliform virus* (SMBV), was reported infecting *Stilbocarpa polaris* on subantarctic Macquarie Island (Skotnicki *et al.* 2003). Although the subantarctic islands are among the most remote places on Earth, of the 21 species of dicotyledons that occur on Macquarie Island, 19 are also part of the New Zealand flora (Allan 1982; Hnatiuk 1993). The unexpected discovery of SMBV indicated that the New Zealand flora may harbour unique viruses and pararetroviral sequences that have been evolving in isolation for a long period of their evolutionary history. To date no badnaviruses have been recorded in the New Zealand flora (Pearson *et al.* 2006).

New Zealand's flora has strong affinities with the floras of other Southern Hemisphere landmasses and there has been much debate over whether these affinities reflect ancient Gondwanan connections or more recent dispersal events. Molecular phylogenies show that many plant lineages are recent arrivals in New Zealand, which have diversified and then travelled to other Southern Hemisphere landmasses (Winkworth *et al.* 2002).

We chose to study *Melicytus* (Violaceae) in detail as the genus appears to have evolved in New Zealand and dispersed north into the south-western Pacific and west into Australia and Tasmania (Mitchell *et al.* 2009). *Melicytus ramiflorus* is a small tree found in lowland forest throughout New Zealand. It extends offshore to the Three Kings Islands, Kermadec Islands and Norfolk Island. Closely allied species are found in Tonga, Samoa, Solomon Islands, Vanuatu and Fiji (Kellogg and Weitzman 1985). *Melicytus alpinus* is a widely distributed shrub found mainly on the dryer eastern side of the South Island where it is found predominantly in grassland and shrubland (Mitchell *et al.* 2009). Species related to *M. alpinus* are found offshore in eastern and South Australia and Tasmania. Other species of *Melicytus* have more restricted distributions, with *Melicytus chathamicus* being confined to the Chatham Island group which lies 700 km east of the South Island (Mitchell *et al.* 2009).

In this paper, we used specific polymerase chain reaction (PCR) primers to amplify a 530-nucleotide sequence of badnavirus reverse transcriptase (RT) to investigate the diversity and abundance of badnavirus sequences in the New Zealand flora. We found that often individual plants contained mixed populations of the target sequence. These mixed sequences were closely related and were resolved by denaturing gradient gel electrophoresis (DGGE) giving characteristic band patterns for each species examined. In one species (*M. ramiflorus*), DGGE band patterns were shown to be inherited.

Materials and methods

Study species

Unless otherwise stated, plants showing no visible signs of disease were collected from a range of habitats in the vicinity of Dunedin, East Otago, Lower South Island, New Zealand. Approximately 10 % of samples were mechanically inoculated to subsets of the species listed in Guy *et al.* (1984) in an attempt to detect latent virus infection. Extracts of the species listed with a superscript in Table 1 (200 mM phosphate buffer pH 6.1, 0.5 % sodium sulphite, 0.5 % 2-mercaptoethanol, 2 % polyethylene glycol 6000, 2 % polyvinyl pyrrolidone) were clarified by centrifugation (10 000 × *g*/20 min) and then adjusted to 6 % PEG-6000 final concentration. The clarified extracts were incubated at 4 °C for 2 h and centrifuged (10 000 × *g*/20 min). The precipitate was resuspended (10 mM phosphate buffer pH 7.4, 1.5 % sodium chloride, 2 % Triton X-100) overnight at 4 °C. Following centrifugation (5000 × *g*/20 min) the supernatant was centrifuged at 100 000 × *g*/2 h and the resulting pellets were resuspended in distilled water and then examined for virus particles in a transmission electron microscope.

Extraction of DNA

Total genomic DNA was extracted from fresh or frozen leaf tissue. A 50 mg sample of leaf tissue was frozen in liquid nitrogen and placed in a lysing matrix tube containing a garnet matrix and 1/4″ ceramic sphere (Lysing Matrix A, Q-BIO gene, MP Biomedicals, LLC,

Table 1 **Badnavirus RT sequences in New Zealand plant species.**

Family	Species	PCR	Number of badnavirus-positive samples/collections examined	Sequence (GenBank Accession)
Apocynaceae	*Parsonsia heterophylla*[a]	+	2/2	FJ900050, FJ900051
Apiaceae	*Stilbocarpa polaris*[a]	−	0/5	
Araliaceae	*Raukaua anomalus*	−	0/1	
	Schefflera digitata	−	0/1	
Asteraceae	*Brachyglottis sciadophila*	−	0/1	
	Celmisia haastii	+	1/1	
	Celmisia prorepens	−	0/1	
	Celmisia viscosa	+	1/1	
Cornaceae	*Griselinia littoralis*	+	1/1	
Escalloniaceae	*Carpodetus serratus*	−	0/1	
Poaceae	*Chionochloa rubra* subsp. *cuprea*	+	1/1	
	Chionochloa rubra subsp. *rubra* var. *inermis*[a]	+	1/1	FJ900044
Icacinaceae	*Pennantia corymbosa*	−	0/2	
Malvaceae	*Hibiscus trionum*	+	1/1	
	Hoheria angustifolia	−	0/1	
Moraceae	*Streblus microphylla*	−	0/1	
Myrsinaceae	*Myrsine australis*	+	2/2	
	Myrsine chathamica	+	2/2	
	Myrsine coxii	+	1/1	
	Myrsine divaricata	+	4/4	
	Myrsine nummularifolia	+	1/1	
	Myrsine salicinia	+	2/2	FJ900049
Myrtaceae	*Kunzea ericioides*	−	0/1	
	Meterosideros diffusa	−	0/1	
	Neomyrtus obcordata	−	0/1	
Pittosporaceae	*Pittosporum tenuifolium*	−	0/1	
Polygonaceae	*Muehlenbeckia australis*	+	1/1	
Ripogonaceae	*Ripogonum scandens*	−	0/4	
Rosaceae	*Acaena juvenca*	−	0/1	
	Rubus cissioides	−	0/1	
Rubiaceae	*Coprosma* 'Beatson's Gold'	+	1/1	
	Coprosma aerolata	+	1/1	
	Coprosma cheesemanii	+	1/1	FJ900045
	Coprosma ciliata	+	1/1	
	Coprosma crassifolia	+	5/5	FJ900046
	Coprosma decurva	−	0/1	
	Coprosma elatirioides	+	1/1	FJ900047
	Coprosma foetidissima	+	1/1	
	Coprosma lucida[a]	+	3/3	FJ900048

Continued

Table 1 ***Continued***

Family	Species	PCR	Number of badnavirus-positive samples/collections examined	Sequence (GenBank Accession)
	Coprosma parviflora	−	0/1	
	Coprosma propinqua	+	2/2	
	Coprosma rhamnoides	+	1/1	
	Coprosma robusta	+	1/1	
	Coprosma rotundifolia	+	2/2	
	Coprosma rugosa	−	0/1	
	Coprosma virescens	+	1/1	
Plantaginaceae	*Veronica sect. Hebe dieffenbachii*	+	1/1	
	V. sect. Hebe odora	+	1/1	
	V. sect. Hebe salicifolia	+	1/1	
Solanaceae	*Solanum laciniatum*	−	0/1	
Violaceae	*Melicytus alpinus*[a]	+	13/18	FJ900052–FJ900061
	Melicytus chathamicus	+	2/3	FJ900062, FJ900063
	Melicytus flexuosus[a]	+	2/3	FJ900064
	Melicytus lanceolatus	−	0/3	
	Melicytus macrophyllus	+	1/1	
	Melicytus micranthus	+	1/1	FJ900065
	Melicytus obovatus	−	0/1	
	Melicytus ramiflorus[a]	+	19/19	FJ900066–FJ900072
Urticaceae	*Urtica australis*	−	0/1	
Winteraceae	*Pseudowintera colorata*	+	1/1	

[a]Species examined in the electron microscope for virus particles.

Solon, OH, USA). The frozen sample was disrupted in a Retsch mixer mill Type 301 (Retsch GmbH & Co. KG, Haan, Germany) using three cycles of 30 s at 30 cycles per second. Following this step, the DNA was purified using a DNAeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions.

Polymerase chain reaction amplification

All PCR amplifications were performed in a 20 μL volume containing 1–50 ng of template, 10 pmol each primer and 17 μL of Reddy Mix PCR master mix containing 1.5 mM $MgCl_2$ (Abgene Limited, Epsom, UK). Polymerase chain reaction products were purified using a QIAquick PCR purification kit (Qiagen) prior to sequencing.

The badna RT fragment was amplified using the degenerate primers HafF (5′ ATG CCI TTY GGI ITI AAR AAY GCI CC 3′) and HafR (5′ CCA YTT RCA IAC ISC ICC CCA ICC 3′) (Yang *et al.* 2003). For amplification for DGGE, the HafR primer was modified by including a GC clamp (5′ AGC CGC GCG GCG GGC GGG GCG GGG GCA CGG CCA YTT RCA IAC ISC ICC CCA ICC 3′). Polymerase chain reaction parameters were as follows: one cycle at 95 °C for 5 min, then 35 cycles of 94 °C for 30 s, 52 °C for 1 min, 72 °C for 1 min followed by one cycle of extension at 72 °C for 30 min. For PCRs using the GC clamped primer, the cycle was modified by increasing the denaturing time to 1 min and the extension time to 2 min.

The internal transcribed spacer (ITS) of nuclear ribosomal DNA was amplified using the primers ITS-1 and ITS-4 (White *et al.* 1990). Polymerase chain reaction parameters were one cycle at 95 °C for 5 min, then 35 cycles of 94 °C for 1 min, 48 °C for 45 s, 72 °C for 2 min followed by one cycle of extension at 72 °C for 10 min.

Denaturing gradient gel electrophoresis

Denaturing gradient gel electrophoresis was performed using the DCode Universal Mutation Detection System (Bio-Rad Laboratories, Hercules, CA, USA). Gradient gels containing 8 % polyacrylamide (37.5:1 ratio of acrylamide:bis-acrylamide) were formed using a peristaltic pump (Gilson Minipuls) and a gradient mixing device (Hoefer SG30) with a denaturant gradient of 25–40 %. Gels were run in 1× Tris-acetate-EDTA buffer at 75 V, 400 mA for 16 h at 60 °C. Gels were stained for 60 min in MilliQ water containing 0.5 mg of ethidium bromide per litre and destained for 30 min in MilliQ water. Gels were placed on a UV transilluminator and the digital images were recorded with a Kodak Gel Logic Imaging System (Eastman Kodak Co., Rochester, NY, USA). Bands of interest were excised from DGGE gels and DNA eluted by soaking in 100 μL of water (MilliQ) overnight. Aliquots of the eluate were re-amplified using the HafF and HafR primers and purified prior to sequencing.

DNA sequencing

Samples were sequenced by the Allan Wilson Centre Genome Service, Massey University, using the Applied Biosystems BigDye™ Terminator Version 3.1 Ready Reaction Cycle Sequencing Kit (Life Technologies Corporation, Carlsbad, CA, USA).

Analysis of badnavirus RT sequences

All available badnavirus RT fragment sequences from GenBank (http://www.ncbi.nlm.nih.gov/) were aligned with the sequences obtained in this study using ClustalW. Mismatched sequences and additional sequence representatives with >95 % identity to existing sequences were removed from the alignment. The final nucleotide alignment was then derived from the corresponding translated amino acid alignment that had been adjusted manually. The aligned dataset contained 534 nucleotide characters, of which 448 were parsimony informative. A maximum parsimony phylogenetic analysis was done using PAUP* version 4.0b10 (Swofford 2003), with *Rice tungro bacilliform virus* (RTBV GenBank: AF113831) as the outgroup. One thousand heuristic search replicates were done, using the tree bisection-reconnection branch-swapping algorithm and a random addition of taxa for each replicate, saving 100 trees per replicate. Uninformative characters were excluded and gaps were treated as missing data. Branch support was assessed by bootstrap analysis, with 1000 bootstrap replicates. Each bootstrap replicate involved two replicate heuristic searches with random addition of taxa, saving a maximum of 50 trees per random addition replicate.

Analysis of *Melicytus* ITS region

Internal transcribed spacer sequences were aligned using ClustalW. The dataset contained 704 nucleotide characters and eight binary characters, representing indels coded as present (1) or absent (0). A Bayesian inference phylogenetic analysis was done using MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003), using a general time reversible model with gamma-distributed rate variation across sites and a proportion of invariable sites. Four hundred and fifty thousand generations were run, sampling every hundredth generation. The standard deviation of split frequencies was 0.008268 after 450 000 generations. Trees and branch length samples were summarized after discarding the first 25 % (1125) of the samples.

Results

New Zealand plant species tested for badnavirus sequences

A total of 60 species belonging to 28 genera and 22 families were tested (Table 1). A total of 37 species gave a positive result for the 530 bp fragment of RT with the remaining 23 being negative. No positives were found from species belonging to the families Apiaceae, Araliaceae, Escalloniaceae, Icacinaceae, Moraceae, Myrtaceae, Ripogonaceae, Rosaceae, Solanaceae and Urticaceae that were examined. Reverse transcriptase DNA sequences were detected in both endemic and other native species with wider distributions. Reverse transcriptase sequences were found in plant species from vegetation types ranging from tussock grassland communities to shrublands and forests.

The RT sequences reported here were detected in a monocot (*Chionochloa rubra* subsp. *rubra* var. *inermis*), a liane (*Parsonsia heterophylla*) and small trees and shrubs. *Chionochloa rubra* subsp. *rubra* var. *inermis* is a member of the widespread and ecologically important genus of snowgrasses (Pirie *et al.* 2008). Reverse transcriptase sequences were present in four widespread species of *Myrsine* and two Chatham Islands endemics *Myrsine chathamica* and *M. coxii*. Thirteen out of the 16 species of *Coprosma*, a large and taxonomically diverse genus of shrubs, surveyed contained RT sequences. The three species of *V.* sect. *Hebe* all contained RT sequences. *Veronica salicifolia* is a lowland forest species native to New Zealand and coastal Chile, *V. dieffenbachii* is endemic to the Chatham Islands and *V. odora* is a subalpine species.

Stilbocarpa mosaic bacilliform virus was not detected in *S. polaris* collected from Auckland and Campbell Island (no collections from Macquarie Island). None of

the *Stilbocarpa* plants showed symptoms of virus infection and no RT sequences were amplified using the Haf primers or the primers of Skotnicki *et al.* (2003).

Melicytus ramiflorus was the only species sampled that displayed virus symptoms. Extensive electron microscope examination of *M. ramiflorus* sap, pellets prepared from extracts by differential centrifugation and ultrathin sections failed to detect any badnavirus particles. Particles of an unidentified icosohedral virus were observed in healthy and symptomatic trees (D. J. Lyttle and P. L. Guy, unpubl. res.). However, there was no correlation between the presence of RT sequence and virus symptoms. The failure to find particles with characteristic badnavirus morphology suggests that virus DNA is present only as integrated sequence.

Virus particles were not detected in any of the preparations from the other species examined in the electron microscope. No mechanically transmissible viruses were detected during this study.

Phylogenetic diversity of endogenous badnavirus sequences

BLAST searching (Altschul *et al.* 1997) showed that the sequences reported here did not match any known badnavirus sequence. When the 21 RT DNA sequences from 12 species belonging to four genera (*Melicytus* 17, *Coprosma* 7, *Parsonsia* 2, *Chionochloa* 1, *Myrsine* 1) were aligned with known badnavirus sequences from GenBank, each species was shown to contain distinct sequences (Fig. 1).

With the exception of the *Coprosma propinqua* sequence, *Coprosma* sequences were monophyletic (clade A, Fig. 1) with good bootstrap support. Another monophyletic group combined the *M. alpinus* and *M. chathamicus* sequences. All the *M. ramiflorus* sequences, with the exception of *M. ramiflorus* Wa6, grouped together with strong bootstrap support. *Melicytus ramiflorus* and *Melicytus flexuosus* sequences were less closely related. Two *P. heterophylla* sequences were closely related.

Conservation of amino acid sequences in endogenous badnavirus RT

All the endogenous badnavirus sequences that were determined were translated into the corresponding amino acid sequence and aligned. After duplicate sequences were removed, this gave a total of 21 distinct amino acid sequences. No frameshift mutations occurred in any of the sequences and a conservative pattern of amino acid changes was observed.

Analysis of badnavirus RT sequence diversity from *Melicytus*

Badnavirus RT sequences from the *Melicytus* spp. were investigated to determine the sequence diversity present within individual plants, among individuals of the same species and between different species in the genus. It was not possible to establish an unambiguous sequence directly from the PCR product for every collection investigated as a number of the PCR products contained a heterogeneous mixture of sequences. These mixtures were subjected to DGGE to resolve the oligonucleotide fragments.

The 530 bp RT fragment from *M. ramiflorus* was resolved into >25 distinct bands by DGGE (Fig. 2A). Overall, band patterns from different individuals were very similar in appearance and showed many bands in common. However, differences were apparent as some individuals contained additional bands or conversely lacked other bands. For example, the DGGE band pattern from a plant collected from Karamea (lane 4) was similar though distinct from eastern coastal Otago plants (lanes 2, 3) although fewer bands were observed overall. A plant from Rangitoto Island, Auckland (lane 6) showed three distinctive strong bands in the lower part of the gel but otherwise was similar to the remaining collections. Four plants collected from the same locality on the Otago Peninsula (lanes 7, 8, 9, 10) showed individual differences but contained many common bands. There were no consistent differences in the band patterns from plants collected from the southern part of the South Island (Dunedin), the northern part of the South Island (Karamea) and the northern part of the North Island (Auckland) that could be attributed to geographic location.

Inheritance of badnavirus fragment patterns

Six plants grown from seed collected from a single parent (Mel3) were compared in this analysis. As *M. ramiflorus* is a dioecious species only the female parent is known for this particular cross. Badnavirus fragment patterns for the female parent and the six progeny plants were generated using DGGE (Fig. 2B). The progeny plants contained all the badnavirus fragments present in the maternal plant or a subset of them. Three seedlings, P1, P3 and P5, showed DGGE patterns identical to that of the parent. Each of the remaining three seedlings (P2, P4 and P6) all showed a subset of the parental bands.

Analysis of variation in the *M. alpinus* species complex using badnavirus fragment patterns

Analysis of badnavirus fragment patterns for the *M. alpinus* collections showed considerable variation within the species (Fig. 3). Plants identified in the field

Fig. 1. **Phylogenetic tree showing the relationships of New Zealand endogenous badnavirus sequences to known badnaviruses.**

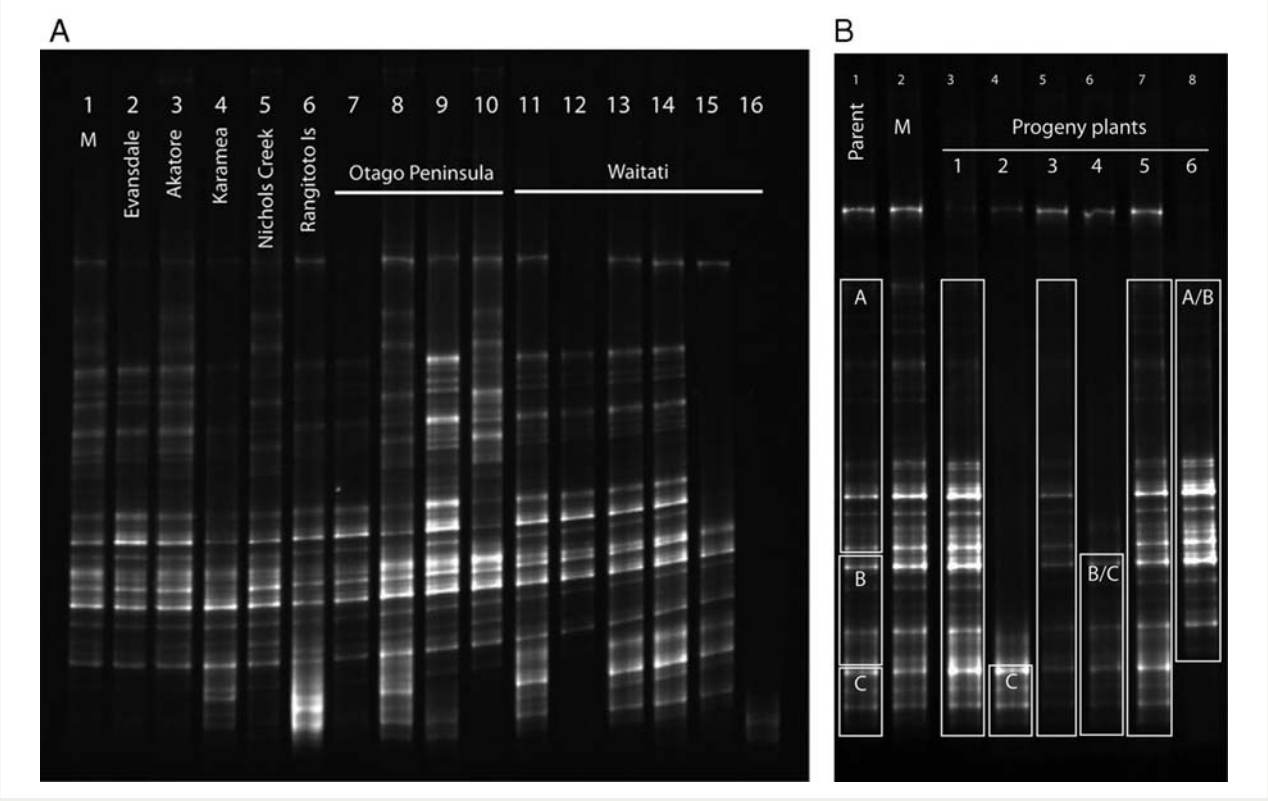

Fig. 2. (A) Denaturing gradient gel electrophoresis analysis of RT fragments from different collections of *M. ramiflorus* from different geographic localities. 1: Marker: Otago Peninsula (Mel1); 2: Evansdale Glen; 3: Akatore; 4: Karamea (West Coast); 5: Nichols Creek (Dunedin); 6: Rangitoto Island (Auckland); 7: Otago Peninsula (OP-1); 8: Otago Peninsula (OP-2); 9: Otago Peninsula (OP-3); 10: Otago Peninsula (OP-4); 11: Waitati (Wa1); 12: Waitati (Wa2); 13: Waitati (Wa3); 14: Waitati (Wa4); 15: Waitati (Wa5); 16: Waitati (Wa6). (B) Denaturing gradient gel electrophoresis of RT fragments from different collections of *M. ramiflorus* from one maternal and six progeny plants. The parent plant contained a complete set of fragments (labelled groups A, B and C). Progeny P1, P3 and P5 also contained the complete set, whereas progeny P2 contained only group C, progeny P4 contained groups B and C, and progeny P6 contained groups A and B. 1: Waitati (Mel-3) female parent; 2: Marker: Waitati (Mel-4); 3–8: Progeny plants: 3: Waitati (P1); 4: Waitati (P2); 5: Waitati (P3); 6: Waitati (P4); 7: Waitati (P5); 8: Waitati (P6).

as *M. alpinus* collected from six coastal sites in eastern Otago (Akatore Creek, Hooper's Inlet, Nugget Point, Pyramids, Sandymount and Taiaroa Head) and from four inland sites (Swampy Summit, Middlemarch, Ida Range and Pisa Range) were subjected to PCR and DGGE analysis. DNA samples from the Swampy Summit plants (4/4) and Sandymount plants (2/2) did not amplify with the HafF and HafR primers. The remaining DNA samples all gave PCR products of the expected size. When the corresponding GC clamped primers were used for the amplification of the RT fragment, the plants collected from the coastal sites (Akatore Creek, Hooper's Inlet, Nugget Point, Pyramids, Sandymount and Taiaroa Head) and one inland site (Middlemarch) showed identical DGGE band profiles (Fig. 3, type 1). Two of the inland plants, Pisa Range 1 and Ida Range, shared a second unique DGGE band profile (Fig. 3, type 2). The two remaining plants collected from the Pisa Range showed band profiles that were different from each other and from the other two groups. The two specimens of *M. chathamicus* examined gave identical DGGE patterns (results not shown).

Congruence between host phylogeny and endogenous badnavirus profiles

To clarify the phylogenetic relationships among the different species of *Melicytus* examined for badnavirus sequences in this study, the sequence of the ITS region was determined for 34 collections of various *Melicytus* spp. (Table 2) using the primers ITS-1 and ITS-4.

Two major clades with weak support (posterior probability value 0.92) were resolved. One clade contained

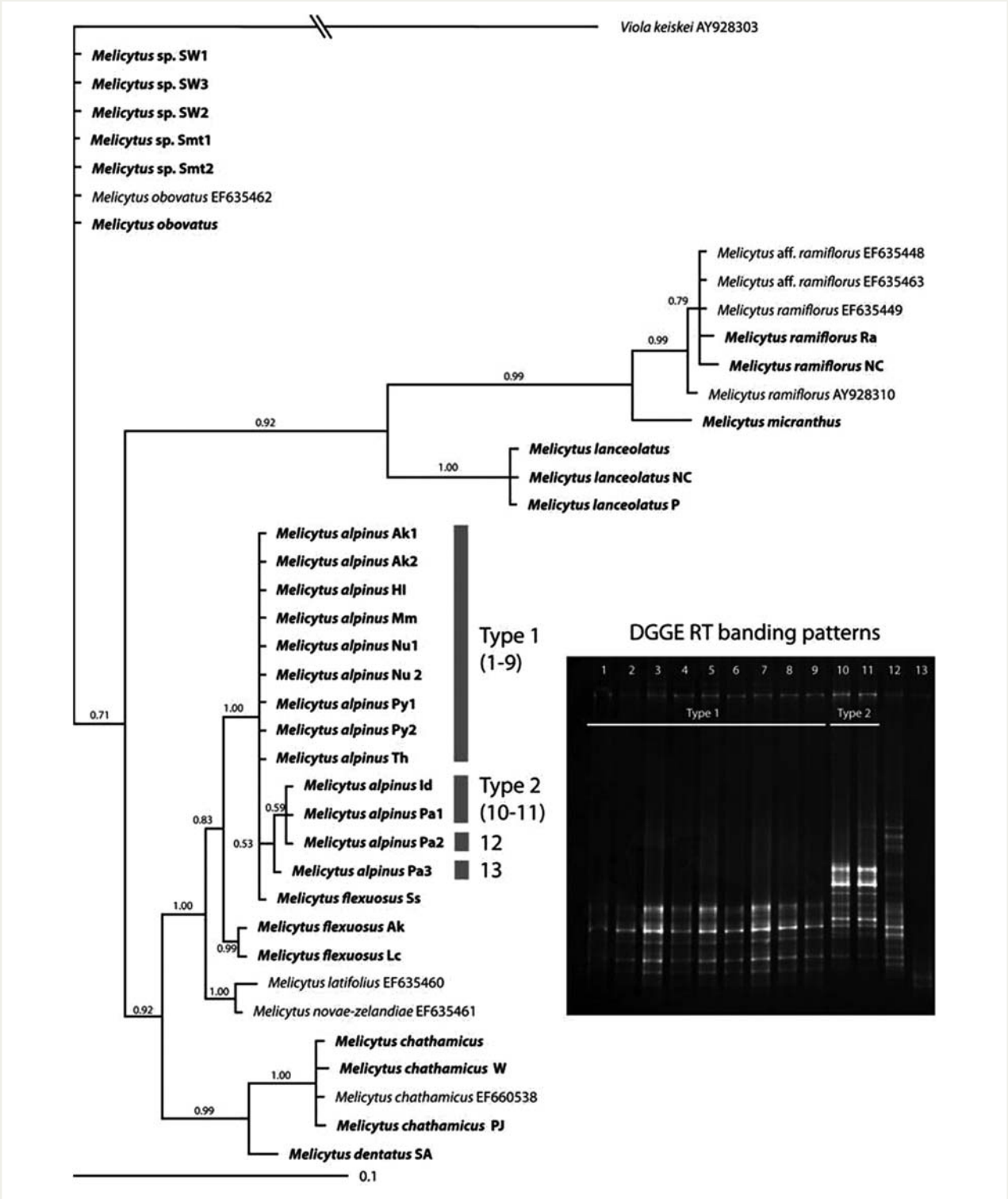

Fig. 3. Phylogenetic tree showing the relationships of various *Melicytus* species based on their ITS sequences and the DGGE band patterns of the RT fragments from *M. alpinus*. For *M. alpinus*, two predominant fragment patterns are distinguishable in the gel (designated type 1 (lanes 1–9) and type 2 (lanes 10–11)). 1: *M. alpinus* Akatore Creek 1; 2: *M. alpinus* Akatore Creek 2; 3: *M. alpinus* Hooper's Inlet; 4: *M. alpinus* Middlemarch; 5: *M. alpinus* Nugget Point 1; 6: *M. alpinus* Nugget Point 2; *M. alpinus* Pyramids 1; 8: *M. alpinus* Pyramids 2; 9: *M. alpinus* Taiaroa Head; 10: *M. alpinus* Ida Range; 11: *M. alpinus* Pisa Range 1; 12: *M. alpinus* Pisa Range 2; 13: *M. alpinus* Pisa Range 3.

Table 2 ***Melicytus*** **specimens used for ITS sequence and DGGE comparisons.**

Identifier	Species	Locality	Otago Herbarium sheet no.	GenBank Accession	RT[a]	DGGE	ITS indel
alpAk1	*Melicytus alpinus*	Akatore Creek	OTA 060541	FJ900084	Present	alp Ty1	indel3T
alpAk2	*Melicytus alpinus*	Akatore Creek	OTA 060542	FJ900083	Present	alp Ty1	indel3T
alpHI	*Melicytus alpinus*	Hooper's Inlet, Otago Peninsula	OTA 061459	FJ900085	Present	alp Ty1	indel3T
alpId1	*Melicytus* aff *alpinus*	Ida Range	OTA 060539	FJ900088	Present	alp Ty2	?
alpMm	*Melicytus alpinus*	Crater, Middlemarch	No voucher	FJ900082	Present	alp Ty1	indel3T
alpNu1	*Melicytus alpinus*	Nugget Point, South Otago	OTA 060848	FJ900081	Present	alp Ty1	indel3T
alpNu2	*Melicytus alpinus*	Nugget Point, South Otago	OTA 060849	FJ900079	Present	alp Ty1	indel3T
alpPa1	*Melicytus alpinus*	Locharburn, Pisa Ra	OTA 061461	FJ900089	Present	alp Ty2	indel 1a ~200
alpPa2	*Melicytus alpinus*	Locharburn, Pisa Ra	OTA 061462	FJ900090	Present	alp Ty1 variant	No indel
alpPa3	*Melicytus alpinus*	Locharburn, Pisa Ra	OTA 061463	FJ900094	Present	variant	indel/hybrid
alpPy1	*Melicytus alpinus*	Pyramids, Otago Peninsula	No voucher	FJ900086	Present	alp Ty1	indel3T
alpPy2	*Melicytus alpinus*	Pyramids, Otago Peninsula	OTA 060540	FJ900080	Present	alp Ty1	indel3T
alpSmt1	*Melicytus* aff *obovatus*	Sandymount, Otago Peninsula	OTA 061455	FJ900102	Absent	Null	No indel
alpSmt2	*Melicytus* aff *obovatus*	Sandymount, Otago Peninsula	No voucher	FJ900103	Absent	Null	No indel
alpSw1	*Melicytus* aff *obovatus*	Swampy Summit	No voucher	FJ900091	Absent	Null	No indel
alpSw2	*Melicytus* aff *obovatus*	Swampy Summit	OTA 061453	FJ900093	Absent	Null	No indel
alpSw3	*Melicytus* aff *obovatus*	Swampy Summit	OTA 061454	FJ900092	Absent	Null	No indel
alpTh	*Melicytus alpinus*	Taiaroa Head, Otago Peninsula	OTA 061460	FJ900087	Present	alp Ty1	indel3T
chath	*Melicytus chathamicus*	Landcare garden	No voucher	FJ900098	Present	Chath	
chathPJ	*Melicytus chathamicus*	Landcare garden	No voucher	FJ900100	Absent	Null	
chathW	*Melicytus chathamicus*	Waitangi, Chatham Islands	OTA 061464	FJ900099	Present	Chath	
denSA	*Melicytus dentatus*	South Australia	No voucher	FJ900101	Present	Den	
flexAk	*Melicytus flexuosus*	Akatore Creek	OTA 060545	FJ900096	Present	Flex	No indel
flexLc	*Melicytus flexuosus*	Landcare garden	No voucher	FJ900097	Present	ND	
flexSs	*Melicytus* aff *flexuosus*	Swampy Spur, Dunedin	OTA 061469	FJ900095	Absent	Null	indel3T
lan	*Melicytus lanceolatus*	Landcare garden	No voucher	FJ900076	Absent	Null	
lanNC	*Melicytus lanceolatus*	Nichols Creek, Dunedin	No voucher	FJ900077	Absent	Null	
lanP	*Melicytus lanceolatus*	Papatowai, South Otago	OTA 060858	FJ900078	Absent	Null	
mic	*Melicytus micranthus*	Motokarara Nursery	No voucher	FJ900075	Present	ND	
obo	*Melicytus obovatus*	Matai Nursery, Waimate	No voucher	FJ900104	Absent	Null	
ram NC	*Melicytus ramiflorus*	Nichols Creek, Dunedin	No voucher	FJ900074	Present	ram	
ramRa	*Melicytus ramiflorus*	Rangitoto Island, Auckland	No voucher	FJ900073	Present	ram	

[a]RT sequence.

M. lanceolatus, *M. ramiflorus* and *M. micranthus*. The second comprised *M. alpinus*, *M. flexuosus* and *M. novae zelandiae* from New Zealand and *M. chathamicus* (Chatham Islands), *M. latifolius* (Norfolk Island) and *M. dentatus* (South Australia). The two most geographically separated species, *M. dentatus* and *M. chathamicus*, had the most closely related sequences.

In this analysis *Melicytus obovatus* (Cook's Strait), together with five specimens of *M. alpinus* collected from Sandymount and Swampy Summit near Dunedin, formed an unresolved basal polytomy. Plants from all three locations had identical ITS sequences. Although the small-leaved specimens were morphologically similar to *M. alpinus*, their ITS sequence and lack of RT sequence separates them from *M. alpinus sensu stricto*.

Analysis of individual ITS sequences of *Melicytus* DNAs showed that two types of variation were present. Many DNA templates showed single-base polymorphisms (Table 2) where the sequencing chromatograms showed alternate bases at a number of positions. This type of polymorphism was consistent in terms of nucleotide ratios at each position and was scored if it was observed in both the forward and reverse sequences for a given sample. A second polymorphism identified was the presence of an indel that caused nucleotide shifts giving overlapping peaks in the sequencing chromatogram. Two different indels were initially identified as a character present in some species from the alignment of all the *Melicytus* ITS sequences. The 3T indel occurred in sequences from *M. flexuosus*, and the TGA indel occurred in sequences from the *M. obovatus* clade.

The 3T indel was found as a hybrid polymorphism in a group of specimens collected from mainly coastal sites. The simplest hypothesis to account for the 3T indel polymorphism is that the plants in which it occurs were originally derived from hybrids between two pre-existing species. From the results of this study it is likely that the coastal clade of *M. alpinus* arose as a hybrid between *M. alpinus sensu stricto* (alpId, alpPa3 and alpPa2) and *M. flexuosus*. In the same manner we consider that the specimens Sw1, Sw2 and Sw3 are likely to be of hybrid origin between an *M. obovatus*-like progenitor and *M. flexuosus* which is still growing nearby. The predominant sequence from these plants shows a TGA indel characteristic of *M. obovatus*, but a lesser amount of the 3T indel of *M. flexuosus* is also observed. In positions where the sequences of *M. obovatus* and *M. flexuosus* diverge, the predominant base is that of *M. obovatus* and the minor base is from *M. flexuosus*. The closely related plants Smt1 and Smt2 are isolated from any currently existing *M. flexuosus* population and do not show the introgression of *M. flexuosus* sequences. Each group shows a badna RT DGGE pattern that is entirely consistent with the ITS sequence data.

Discussion

In spite of its Gondwanan origins, many of New Zealand's plant lineages are more recent arrivals (Winkworth *et al.* 2002) with many species but few genera being endemic. The New Zealand genera which contained RT sequences also occur in other regions: S.E. Asia, Papua New Guinea, Australia, New Zealand, Pacific Islands, Sub Antarctic Islands and South America. Based on our results, we suggest that endogenous badnaviruses may be common in these regions. *Melicytus dentatus* collected in South Australia contained RT sequences.

Denaturing gradient gel electrophoresis proved a useful technique to analyse the sequence diversity present in many of the samples: cutting out individual bands and re-amplifying them led to additional sequence information. The most complex band pattern was found in *M. ramiflorus*. When PCR products from this species were analysed by DGGE, over 25 bands were observed in the majority of the samples. Some bands showed minor nucleotide changes but the amino acid sequence was conserved, while others showed additional nucleotide changes that resulted in five amino acid changes (*in silico*). Denaturing gradient gel electrophoresis has been used widely in environmental sampling to uncover the diversity of species within various ecosystems (Singh *et al.* 2004; Bull and Stach 2007; Tannock 2008), but this study shows that it is a useful technique to take a snapshot of the diversity within and between genomes.

The presence of endogenous badnaviruses, often as multiple copies, in so many of the species tested and at high incidences, indicates that they may confer a selective advantage. One such advantage could be protection from infection by related viruses via RNA interference (Covey and Al-Kaff 2000; Hull *et al.* 2000; Noreen *et al.* 2007). Multiple copies could have arisen by multiple integrations of badnaviruses into the host genome or by the single integration of concatemers whose component sequences then diverged. Typical of island floras, there is a high incidence of polyploidy in New Zealand species (Richards 1997), which could further increase copy number; however, recent studies with *Nicotiana* spp. (Gregor *et al.* 2004; Skalická *et al.* 2005) have shown selective deletion of repetitive sequences in *some* polyploids.

The inheritance of endogenous badnavirus fragment patterns was investigated by examining plants grown from the seed of a single female plant of *M. ramiflorus*. Three DGGE band patterns were identical to that of the

maternal plant but 3/6 progeny contained only a subset of the maternal bands. There is strong presumptive evidence that many genera of New Zealand native woody plants including *Melicytus* reproduce by facultative apomixis (B. P. J. Molloy, pers. comm., 2008) and this has been shown definitively for the genus *Coprosma* (Heenan *et al.* 2003). The conservation of endogenous badnavirus DGGE band patterns in *M. ramiflorus* is consistent with the existence of apomictic races that would give rise to additional variants whenever out-crossing occurs.

The *M. alpinus* complex includes a diverse assemblage of forms that encompasses a considerable amount of environmental and genetic variation, ranging from a low-growing spiny cushion, typical of the species, to shrubs and dwarf alpine forms growing in high-altitude screes. There are several unnamed shrubby taxa in the *M. alpinus* complex present in New Zealand (Molloy and Clarkson 1996; Mitchell *et al.* 2009). It is likely that apomixis also occurs in these small-leaved species and may be in part responsible for the existence of the bewildering array of forms found in the *M. alpinus* complex. The presence or absence of RT sequences and consistent differences in the RT fragment patterns were observed, and in conjunction with ITS sequences revealed that *M. alpinus* is composed of several distinct taxa.

Integrated badnavirus sequences are diverse (Geering *et al.* 2005*a*, *b*) and have undergone reassortment and given rise to infection during tissue culture (Ndowora *et al.* 1999). Recent work showed that the proliferation stage of micropropagation triggers the episomal replication (Dallot *et al.* 2001). Wounding and various abiotic changes (Lockhart *et al.* 2000; Richert-Pöggeler *et al.* 2003) are known to activate EPRVs and indeed other genomic sequences previously regarded as pseudogenes (Summerfield *et al.* 2008). Considering the growing use of tissue culture and other biotechnologies for plant conservation (Engelmann 1991; Harding *et al.* 1997; Harding 2004; Bapat *et al.* 2008) and the increasing prevalence of these 'two-faced travellers in the plant genome' (Staginnus and Richert-Pöggeler 2006), it now seems prudent to screen wild germplasm for endogenous badnaviruses and other endogenous *Caulimoviridae* (Geering *et al.* 2010) before tissue culture.

Conclusions and forward look

Although our preliminary focus was detecting badnaviruses and endogenous badnaviruses at large in the New Zealand flora, the detection and analysis of badnavirus RT sequences in combination with DGGE were useful for much wider applications in plant systematics, phylogenetics and reproduction biology.

Accession numbers

Reverse transcriptase (see Table 1 Accession Nos FJ90044–72) and ITS (see Table 2 Accession Nos FJ90073–101) sequences are deposited with GenBank. *Melicytus* spp. specimens have been lodged with the University of Otago Herbarium (see Table 2).

Sources of funding

This work was supported by the University of Otago and the Brenda Shore Trust (New Zealand).

Contributions by the authors

D.L., D.O. and P.G. conceived and designed the experiments. D.L. and D.O. did the experiments. D.L., D.O. and P.G. analysed the data and prepared the manuscript.

Acknowledgements

Thanks to Drs Peter Johnson, Landcare, and Kelvin Lloyd, Wildlands, for collecting *Melicytus* and *Chionochloa*, respectively, to Richard Easingwood at the Otago Centre for Microscopy for electron microscopy and to anonymous reviewers for constructive criticism.

Conflicts of interest statement

None declared.

References

Allan HH. 1982. *Flora of New Zealand, Volume 1*. Wellington: Government Printer, 1085 pp.

Altschul SF, Madden TL, Schäffer AA, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Bapat VA, Yadav SR, Dixit GB. 2008. Rescue of endangered plants through biotechnological applications. *National Academy Science Letters-India* **31**: 201–210.

Bull AT, Stach JEM. 2007. Marine actinobacteria: new opportunities for natural product search and discovery. *Trends in Microbiology* **15**: 491–499.

Clover GRG, Pearson MN, Elliott DR, Tang Z, Smales TE, Alexander BJR. 2003. Amplification of badnavirus-like sequences from *Yucca elephantipes*. *Australasian Plant Pathology* **32**: 563–564.

Covey SN, Al-Kaff NS. 2000. Plant DNA viruses and gene silencing. *Plant Molecular Biology* **43**: 307–322.

Dallot S, Acuna P, Rivera C, Ramirez P, Cote F, Lockhart BEL, Caruana ML. 2001. Evidence that the proliferation stage of micropropagation procedure is determinant in the expression of *Banana streak virus* integrated into the genome of the FHIA 21 hybrid (*Musa* AAAB). *Archives of Virology* **146**: 2179–2190.

Engelmann F. 1991. In vitro conservation of tropical plant germplasm–a review. *Euphytica* **57**: 227–243.

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA. 2005. *Virus taxonomy, classification and nomenclature of viruses. 8th Report of the International Committee on the Taxonomy of Viruses*. San Diego: Elsevier Academic Press, 1259pp.

Gambley CF, Geering ADW, Steele V, Thomas JE. 2008. Identification of viral and non-viral reverse transcribing elements in pineapple (*Ananas comosus*), including members of two new badnavirus species. *Archives of Virology* **153**: 1599–1604.

Geering ADW, Olszewski NE, Harper G, Lockhart BEL, Hull R, Thomas JE. 2005*a*. Banana contains a diverse array of endogenous badnaviruses. *Journal of General Virology* **86**: 511–520.

Geering ADW, Pooggin MM, Olszewski NE, Lockhart BEL, Thomas JE. 2005*b*. Characterisation of banana streak Mysore virus and evidence that its DNA is integrated in the b genome of cultivated *Musa*. *Archives of Virology* **150**: 787–796.

Geering ADW, Scharaschkin T, Teycheney P-Y. 2010. The classification and nomenclature of endogenous viruses of the family *Caulimoviridae*. *Archives of Virology* **155**: 123–131.

Gregor W, Mette MF, Staginnus C, Matzke ME, Matzke AJM. 2004. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. *Plant Physiology* **134**: 1191–1199.

Guy PL, Dale JL, Adena MA, Gibbs AJ. 1984. A taxonomic study of the host ranges of tymoviruses. *Plant Pathology* **33**: 337–346.

Hansen C, Heslop-Harrison JS. 2004. Sequences and phylogenies of plant pararetroviruses, viruses, and transposable elements. *Advances in Botanical Research Incorporating Advances in Plant Pathology* **41**: 165–193.

Harding K. 2004. Genetic integrity of cryopreserved plant cells: a review. *Cryoletters* **25**: 3–22.

Harding K, Benson EE, Clacher K. 1997. Plant conservation biotechnology: an overview. *Agro Food Industry hi-tech* **8**: 24–29.

Harper G, Hull R, Lockhart B, Olszewski N. 2002. Viral sequences integrated into plant genomes. *Annual Review of Phytopathology* **40**: 119–136.

Heenan PB, Molloy BPJ, Bicknell RA, Luo C. 2003. Levels of apomictic and amphimictic seed formation in a natural population of *Coprosma robusta* (Rubiaceae) in Riccarton bush, Christchurch, New Zealand. *New Zealand Journal of Botany* **41**: 287–291.

Hnatiuk RJ. 1993. Macquarie Island. *Flora of Australia* **50**: 53–62.

Hohn T, Richert-Pöggeler KR, Staginnus C, *et al.* 2008. Evolution of integrated plant viruses. In: Roosinck MJ, ed. *Plant Virus Evolution*. Berlin: Springer, 53–81.

Hull R, Harper G, Lockhart B. 2000. Viral sequences integrated into plant genomes. *Trends in Plant Science* **5**: 362–365.

Kellogg EA, Weitzman AL. 1985. A note on the Oceanic species of *Melicytus* (Violoaceae). *Journal of the Arnold Arboretum* **66**: 491–502.

Lockhart BE, Menke J, Dahal G, Olszewski NE. 2000. Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *Journal of General Virology* **81**: 1579–1585.

Mitchell AD, Heeenan PB, Murray BG, Molloy BPJ, de Lange PJ. 2009. Evolution of the south-western Pacific genus *Melicytus* (Violaceae): evidence from DNA sequence data, cytology and sex expression. *Australian Systematic Botany* **22**: 143–157.

Molloy BPJ, Clarkson BD. 1996. A new, rare species of *Melicytus* (Violaceae) from New Zealand. *New Zealand Journal of Botany* **34**: 431–440.

Ndowora T, Dahal G, LaFleur D, Harper G, Hull R, Olszewski NE, Lockhart B. 1999. Evidence that badnavirus infection in *Musa* can originate from integrated pararetroviral sequences. *Virology* **255**: 214–220.

Noreen F, Akbergenov R, Hohn T, Richert-Poeggeler KR. 2007. Distinct expression of *Petunia vein clearing virus* and the DNA transposon *dTph1* in two *Petunia hybrida* lines is correlated with differences in histone modification and siRNA production. *Plant Journal* **50**: 219–229.

Pearson MN, Clover GRG, Guy PL, Fletcher JD, Beever RE. 2006. A review of the plant viruses and mollicute records for New Zealand. *Australasian Plant Pathology* **35**: 217–252.

Pirie MD, Humphreys AM, Galley C, Barker NP, Verboom GA, Orlovich D, Draffin SJ, Lloyd K, Baeza CM, Negritto M, Ruiz E, Sanchez JHC, Reimer E, Linder HP. 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution* **48**: 1106–1119.

Richards AJ. 1997. *Plant breeding systems*, 2nd edn. London: Chapman & Hall.

Richert-Pöggeler KR, Noreen F, Schwarzacher T, Harper G, Hohn T. 2003. Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. *EMBO Journal* **22**: 4836–4845.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Singh BK, Millard P, Whiteley AS, Murrell JC. 2004. Unravelling rhizosphere–microbial interactions: opportunities and limitations. *Trends in Microbiology* **12**: 386–393.

Skalická K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A. 2005. Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. *New Phytologist* **166**: 291–303.

Skotnicki ML, Selkirk PM, Kitajima E, McBride TP, Shaw J, Mackenzie A. 2003. The first subantarctic plant virus report: Stilbocarpa mosaic bacilliform badnavirus (SMBV) from Macquarie island. *Polar Biology* **26**: 1–7.

Staginnus C, Richert-Pöggeler KR. 2006. Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends in Plant Science* **11**: 485–491.

Summerfield TC, Toepel J, Sherman LA. 2008. Low oxygen induction of normally cryptic *psbA* genes in cyanobacteria. *Biochemistry* **47**: 12939–12941.

Swofford DL. 2003. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*. Version 4. Sunderland, MA: Sinauer Associates.

Tannock GW. 2008. The search for disease-associated compositional shifts in bowel bacterial communities of humans. *Trends in Microbiology* **16**: 488–495.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols, a Guide to Methods and Applications*. San Diego: Academic Press, 315–324.

Winkworth RC, Wagstaff SJ, Glenny D, Lockhart PJ. 2002. Plant dispersal news from New Zealand. *Trends in Ecology & Evolution* **17**: 514–520.

Yang IC, Hafner GJ, Dale JL, Harding RM. 2003. Genomic characterisation of taro bacilliform virus. *Archives of Virology* **148**: 937–949.