Genetic diversity among endogenous plant pararetroviral sequences from geographically diverse sources of dahlia (*Dahlia* spp).

C.V. Almeyda, K. L. Druffel, S. G. Eid* and H. R. Pappu

Department of Plant Pathology. Washington State University, Pullman, WA-99164.

*Present address: Department of Plant, Soil and Entomological Sciences. University of Idaho, Moscow, ID-83844.

Abstract

Two distinct caulimoviruses, *Dahlia mosaic virus* (DMV) and Dahlia common mosaic virus (DCMV), and an endogenous plant pararetroviral sequence (DvEPRS) were reported in *Dahlia* spp. DvEPRS, previously referred to as DMV-D10, was originally isolated in the US from the cultivated *Dahlia variabilis*, and has been also found in New Zealand, Lithuania and Egypt as well as in wild dahlia species growing in their natural habitats in Mexico. Here we report the complete genome sequences of three new DvEPRS isolates from a Lithuanian cultivar (7159 nt), a New Zealander cultivar (7156 nt) and from the wild dahlia species, *D. rupicola* (7133 nt). The three have the structure and organization typical of a caulimovirus species and showed identities between 71 and 97% at the nucleotide level (nt) among various open reading frames (ORFs) when compared to those of DvEPRS. To better understand the genetic diversity, DvEPRS from cultivated and wild dahlia species were compared. A total of 7 full-length sequences were used for phylogenetic analyses, mutation frequencies, potential recombination events, selection and fitness as evolutionary evidences for genetic diversity. Phylogentic analyses showed one clade of all DvEPRS indicating a lack of clustering by geographical origin. When the various DvEPRS were grouped into two taxa, no difference was observed between those from cultivated and wild dahlia species. Assessment of population genetic parameters found strong negative selection for all ORFs, with the replicase region more variable than other ORFs. Identification of potential recombination events involving parents from different lineages provided strong evolutionary evidence for genetic diversity among various DvEPRS. This study contributes to an increased understanding of molecular population genetics and evolutionary pathways of these reverse transcribing viral elements.

Introduction

DMV and DCMV belong to the plant pararetrovirus group, family *Caulimoviridae,* genus *Caulimovirus.* Both have a circular dsDNA genome with six essential open reading frames (ORFs), ORF I-VI, coding for movement protein (MP), aphid transmission factor (ATF), DNA binding protein (DNAb), coat protein (CP), reverse transcriptase (RT) and inclusion body protein (IB). DvEPRS differs from the previously described dahlia mosaic caulimoviruses and other known caulimoviruses in the lack of the ATF and the fusion of the truncated CP with the RT. DvEPRS was originally isolated in the US as the first EPRS isolated from cultivated dahlias. Recently, EPRSs have been isolated from wild dahlia species (*D. coccinea*, *D. sherffii* and *D. tenuicaulis*) and have been designated as D10-DC, D10-DS and D10-DT.

Plant viruses develop genetic variation by errors occurring during the replication of their genomes. Plant pararetroviruses package their genomes as DNA but replicate via an RNA intermediate undergoing reverse transcription. The error rate of replication is though to be the reason for increased genetic diversity and rapid evolution in viral genomes, providing variants upon which natural selection can act . The two main types of errors are mutation and recombination. Selection is estimated using the ratio of non-synonymous to synonymous substitutions (dN/dS); where positive selection is characterized by dN/dS>1, negative selection is characterized by dN/dS<1, and neutral evolution has a dN/dS=1.

After sequencing three new DvEPRS isolates, we aimed to investigate the genetic diversity of all the endogenous pararetroviral sequences from geographically diverse sources of dahlia (US, New Zealand, Lithuania, Egypt and Mexico) obtained up to now. A total of 7-full length sequences were used for phylogenetic and population genetics analyses in order to determine the level and distribution of genetic polymorphism in cultivated and wild populations and to detect the evolutionary forces (mutation, recombination, selection and fitness) that determine the pattern of genetic variation observed in these populations.

Materials and Methods

Plant Material

- Cultivated dahlia samples (*Dahlia variabilis*) from New Zealand and Lithuania were used to test the presence of EPRS.
- Wild dahlia sample (*Dahlia rupicola*) from Mexico was collected to evaluate the presence of EPRS.

DNA extraction, PCR, cloning and sequencing

- DNA extractions were done using the DNeasy Plant Mini Kit (QIAGEN).
- PCR amplicons were cloned using pGEM-T easy cloning kit (Promega)
- Sequence assembling (Contigs Express,Vector NTI Advance 11 program, BLAST).

Genetic diversity studies

- Sequence dataset: 7 full-length EPRSs from cultivated and wild dahlia spp.
- Phylogenetic analyses were done using MEGA5.
- DNA polymorphism and evolution analyses were done using DNAsp and SNAP programs.
- Recombination analyses were done using RDP3 software.

Results

Molecular characterization of three new EPRSs from Dahlia spp.

DvEPRS	ORF I aa/nt	ORF III aa/nt	ORF IV/ V aa/nt	ORF VI aa/nt	IGR nt
D10-LT (%)	94/95	96/97	95/95	98/97	95
D10-NZ (%)	97/97	95/96	90/90	96/96	86
D10-DR (%)	79/79	64/71	77/85	69/72	56

Table 1. Amino acid and nucleotide sequence identities of each ORFs of EPRS isolated in this study with the corresponding ORFs of DvEPRS isolated in the US.

D10-DR	ORF I aa/nt	ORF III aa/nt	ORF IV/ V aa/nt	ORF VI aa/nt	IGR nt
D10-DC (%)	79/80	65/77	75/83	69/73	59
D10-DS (%)	80/79	67/77	74/82	71/74	55
D10-DT (%)	80/80	67/78	80/86	70/74	57

Table 2. Amino acid and nucleotide sequence identities of each ORF of D10-DR isolated in this study with the corresponding ORFs of previously EPRSs reported in wild dahlia species (D10-DC, D10-DS and D10-DT)

Phylogenetic analyses

Clustering into one lineage of all EPRSs isolated from Dahlia spp. was observed, indicating a lack of clustering by geographical origin (Figure 1). When EPRSs isolated from Dahlia spp. were grouped into two taxa, no difference was observed between those from cultivated and wild dahlia species (Figure 2). Each ORF of EPRS from Dahlia spp. formed one cluster in comparison to selected members of the Caulimoviridae (Figure 3).

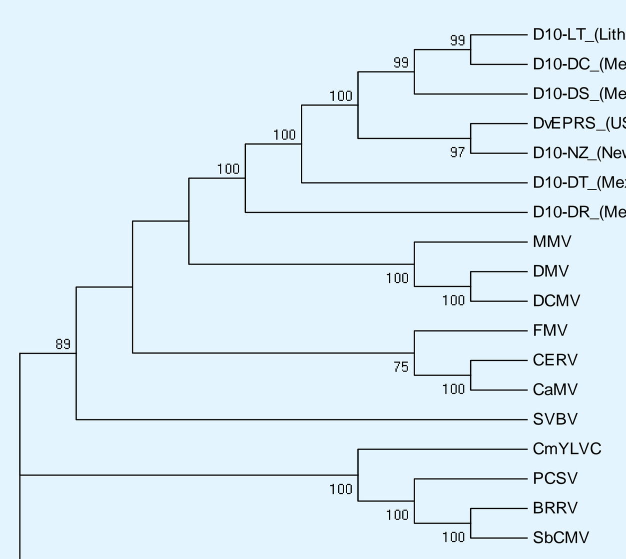

Figure1. Phylogram drawn from Clustal W alingments of the complete genome sequence of selected members of the family Caulimoviridae compared to those of EPRSs isolated from Dahlia spp.

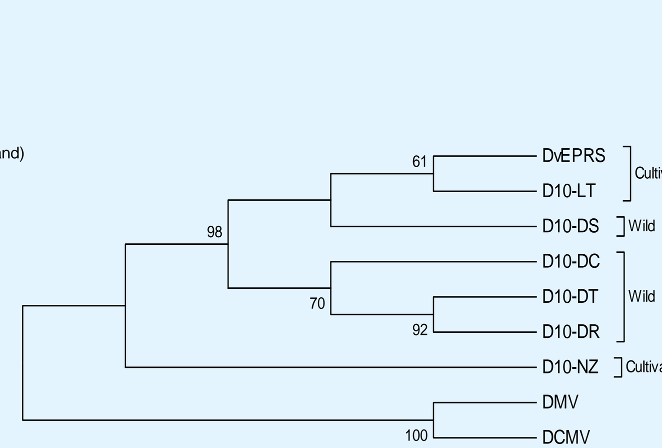

Figure 2. Phylograms drawn from Clustal W alignments of the replicase of EPRSs at the amino acid level when they were grouped into two taxa (cultivated vs. wild).

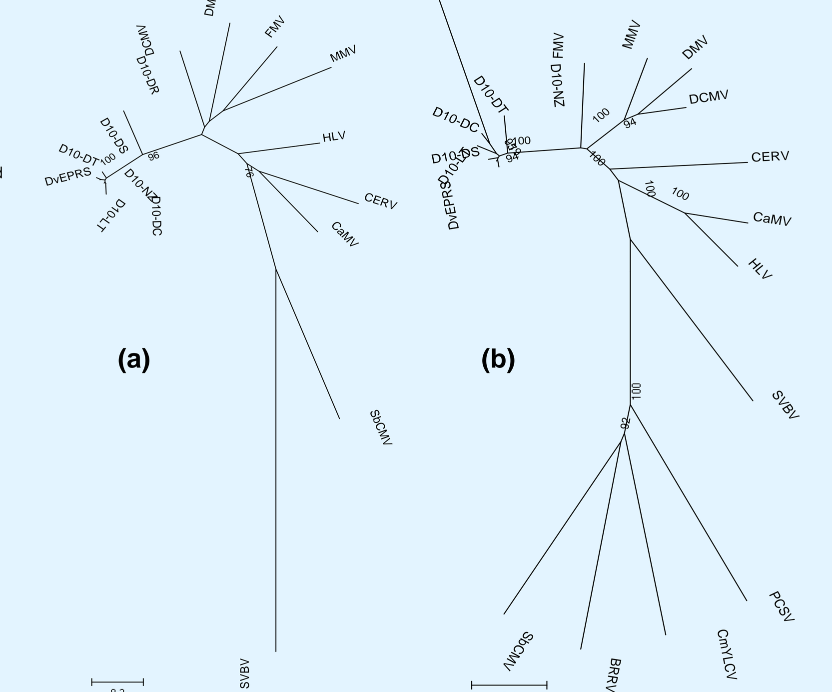

Figure 3. Phylograms drawn from Clustal W alignments of the DNAb (a) and replicase (b) of selected members of the family Caulimoviridae compared to those of EPRSs at amino acid level.

Genetic diversity studies

Highest genetic variation was observed in the replicase (π=0.11577) indicating that replicase was the more variable among all ORFs evaluated. All ORFs were found under negative selection (dN/dS<1).

ORF	Region	*n*	S	Eta	Hd	*π*	Θ	dN	dS	dN/dS	Tajima's D	Fu and Li's D*	Fu and Li's F*
I	MP	7	267	292	1.000	0.09997	0.12075	0.0601	0.4188	0.1435	-1.01226	-0.93187	-1.05396
III	DNAb	7	83	87	1.000	0.09262	0.12203	0.0711	0.3214	0.2212	-1.40536	-1.37028	-1.52722
V	RT	7	695	780	1.000	0.11577	0.13276	0.0729	0.4144	0.1760	-0.75430	-0.61956	-0.72126
VI	IB	7	414	457	1.000	0.10321	0.13008	0.0669	0.4201	0.1592	-1.21621	-1.18050	-1.31892

Table 3. Population genetic parameters and neutrality test calculated for each ORF of EPRSs isolated from dahlia spp.

Where:
n: number of isolates, S: number of segregating sites, Eta: total number of mutations, Hd: haplotype diversity, π: nucleotide diversity, Θ: statistic diversity, d_N : average number of non-synonymous substitutions per site, d_S: average number of synonymous substitutions per site.

Figure 4. Amount and distribution of synonymous (d_S) and non-synonymous (d_N) substitutions for replicase and DNAb sequences of EPRSs isolated from Dahlia spp.

Recombination analyses

Potential recombination events involving parents from different lineages provided strong evolutionary evidence for genetic diversity among various EPRSs isolated from Dahlia spp.

Isolate	Country	Host species	Recombination Site	Major parent	Minor parent	Region
D10-LT	Lithuania	*D.variabilis*	199-265	DvEPRS (USA)	D10-DS (Mexico)	I
			1957-2142	D10-DR (Mexico)	D10-NZ (New Zealand)	IV/V
			2143-2357	D10-DS (Mexico)	D10-NZ (NZ Zealand)	IV/V
			6501-7131	D10-DT (Mexico)	D10-NZ (New Zealand)	IGR
			6690-6864	D10-DC (Mexico)	DvEPRS (USA)	IGR
			7148-905	D10-DS (Mexico)	D10-DT (Mexico)	IGR, I
D10-NZ	New Zealand	*D. variabilis*	697-1794	D10-DC (Mexico)	D10-DT (Mexico)	I, II, III
			2375-2537	D10-DC (Mexico)	D10-DT(Mexico)	IV/V
			2658-3018	D10-DT (Mexico)	D10-DR (Mexico)	IV/V
			3796-4062	D10-DR (Mexico)	DvEPRS (USA)	IV/V
			5335-5558	D10-DC (Mexico)	D10-DS (Mexico)	VI
			6646-6860	DvEPRS (USA)	D10-DC (Mexico)	IGR
D10-DR	Mexico	*D. rupicola*	3294-3370	D10-LT (Lithuania)	DCMV (USA)	IV/V
			4129-4332	D10-DT (Mexico)	DvEPRS (USA)	IV/V
			5417-5445	D10-NZ (New Zealand)	D10-DC (Mexico)	VI
			5780-6177	D10-DT (Mexico)	DMV (USA)	VI
			6243-6311	DvEPRS (USA)	DCMV (USA)	IGR
			6408-26	D10-NZ (New Zealand)	DCMV(USA)	IGR

Table 4. Recombination events occurring in EPRSs isolated from Dahlia spp.

Acknowledgments:
Research was supported in part by the Samuel and Patricia Smith Endowment for Dahlia Virus Research, created by the American Dahlia Society. Funding from the USDA Northwest Nursery Crop Research Center, and the Washington State Commission for Pesticide Registration is gratefully acknowledged.

Washington State University
World Class. Face to Face.

Washington State's Agricultural Experiment Station
Agricultural Research Center