Virus Research 262 (2019) 48–53

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Endogenous pararetrovirus sequences are widely present in Citrinae genomes

Huiwen Yu, Xia Wang, Zhihao Lu, Yuantao Xu, Xiuxin Deng, Qiang Xu*

Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

Keywords:
Endogenous pararetrovirus (EPRV)
Citrinae
Citrus
Evolution

ABSTRACT

Endogenous pararetroviruses (EPRVs) are characterized in several plant genomes and their biological effects have been reported. In this study, hundreds of EPRV segments were identified in six Citrinae genomes. A total of 1034 EPRV segments were identified in the genomes of sweet orange, 2036 in pummelo, 598 in clementine mandarin, 752 in Ichang papeda, 2060 in citron and 245 in atalantia. Genomic analysis indicated that EPRV segments tend to cluster as hot spots in the genomes, particularly on chromosome 2 and 5. Large numbers of simple repeats and transposable elements were identified in the 2-kb flanking regions of the EPRV segments. Comparative genomic analysis and PCR experiments showed that there are highly conserved EPRV segments and species-specific EPRV segments between the Citrinae genomes. Phylogenetic analysis suggested that the integration events of EPRVs could initiate in a common progenitor of Citrinae species and repeatedly occur during the Citrinae divergence.

1. Introduction

Pararetroviruses, a class of retroelements, are similar to retroviruses which encapsidate RNA but distinct from the latter in that double-stranded DNA is harbored in pararetrovirus (PRVs) (Staginnus and Richert-Pöggeler, 2006). Cauliflower mosaic virus is a typical pararetrovirus (Marco and Howell, 1984). A number of pararetrovirus-like sequences with nearly perfect identity were detected in plant genomes (Hull et al., 2000), and they were termed as endogenous pararetroviruses (EPRVs), such as *banana steak virus* (eBSV) (Harper et al., 1999; Ndowora et al., 1999), *tobacco vein clearing virus* (TVCV) (Lockhart et al., 2000), *dahlia mosaic caulimovirus* (DMV) (Eid et al., 2011; Pahalawatta et al., 2008), endogenous Dioscorea bacilliform viruses (eDBVs) (Seal et al., 2014; Umber et al., 2014) and rice tungro bacilliform virus-like (eRTBVL) sequences (Chen et al., 2014). The EPRV sequences were possibly integrated into plant genomes in ancient times, and they are fossil records of ancient pararetrovirus sequences (Chen and Kishima, 2016). Therefore, the EPRV sequences can be used as markers to elucidate the phylogenetic relationship between plant species, and the known evolutionary history of the plant can also be used to study the evolution of EPRVs (Gayral et al., 2010). Due to the presence of EPRV sequences in plant genomes, there were two functions related to the host. On the one hand, EPRV sequences were the reservoir of virus in some plants (Chabannes and Iskra-Caruana, 2013), such as *N. tomentosiformis* EPRV (*Nto*EPRV) in tobacco (Gregor et al., 2004) and eBSV (Chabannes et al., 2013; Gayral et al., 2008; Lheureux et al., 2003), which can be activated and cause infectious symptoms. On the other hand, EPRV sequences showed potential contribution to virus resistance in plants, such as tobacco (*Nicotinana tabacum*) endogenous pararetrovirus (TEPRV) (Mette et al., 2002) and *Grapevine leafoll-associated virus-8* (GLRaV-8) in grapevine genome (Bertsch et al., 2009).

Citrus fruit crops are perennial woody plants. There are many well-known citrus cultivars grown worldwide, such as sweet orange (*Citrus sinensis*), mandarin (*Citrus reticulata*), grapefruit (*Citrus paradisi*), pummelo (*Citrus maxima*) and lemon (*Citrus limon*). Citrus crops show wide sexual compatibility, even on the inter-genus level. This characteristic causes difficulty in the classification and phylogenetic analysis of Citrinae species. Previous studies distinguished primitive citrus, near citrus and true citrus based on the botanical characteristics (Swingle, 1967). Many studies reported the relationship among the citrus species based on molecular markers (Barkley et al., 2006; Federici et al., 1998; Garcia-Lor et al., 2013; Nicolosi et al., 2000). The relationship and the evolutionary history of citrus crops have been comprehensively evaluated by using the nuclear and chloroplast genomes. Furthermore, a high-quality pummelo (*C. maxima*) genome via PacBio-based single molecule sequencing, and genomes of atalantia (*Atalantia buxifolia*), Ichang papeda (*Citrus ichangesis*), citron (*Citrus medica*), sweet orange and clementine mandarin (*Citrus clementina*) have been reported (Carbonell-Caballero et al., 2015; Pfeil and Crisp, 2008; Wang et al., 2017; Wu et al., 2014; Xu et al., 2013).

* Corresponding author.
E-mail address: xuqiang@mail.hzau.edu.cn (Q. Xu).

https://doi.org/10.1016/j.virusres.2018.05.018
Received 12 September 2017; Received in revised form 20 May 2018; Accepted 20 May 2018
Available online 21 May 2018
0168-1702/ © 2018 Elsevier B.V. All rights reserved.

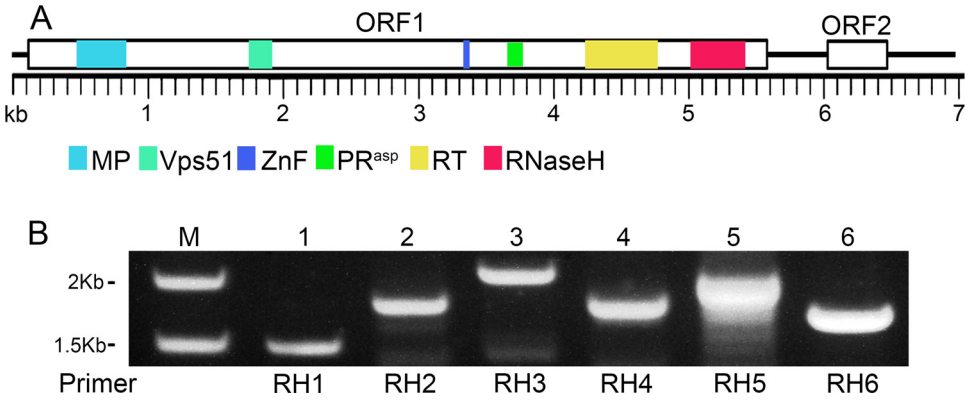

Fig. 1. (A) Simplified structure of citrus EPRV, two ORFs and the related motifs. MP, movement protein; Vps51, Vps51 super family; ZnF, Zinc finger; PR^{asp}, aspartic protease; RT, reverse transcriptase. The motifs were highlighted according to the length in scale. (B) Gel bands of primers RH1, RH2, RH3, RH4, RH5 and RH6 amplified from atalantia (1), Ichang papeda (2), citron (3), pummelo (4), sweet orange (5), and clementine mandarin (6), respectively.

EPRV sequences in the citrus genomes were first identified in the *Citrus tristeza virus* (CTV) resistance locus of *Poncitrus trifoliata*, which is homologous to petunia vein-cleaning virus (PVCV) (Harper et al., 2003; Richert-Pöggeler and Shepherd, 1997; Richert- Pöggeler et al., 2003; Yang et al., 2003). Carrizo citrange (*Citrus sinensis* [L.] Osb. × *Poncirus trifoliata* [L.] Raf.) (Roy et al., 2014) is one of the widely used rootstocks of citrus. Three similar genomes of EPRVs, i.e. CarEPRV1, CarEPRV2 and CarEPRV3 (accession no. KF800043, KF800044 and KF800045), were obtained from Carrizo citrange genome. EPRV (CitPRV) was also detected in the sweet orange cultivar "Valencia" recently (Matsumura et al., 2017). However, comprehensive genomic analyses of EPRVs in citrus are still scarce.

To deepen the knowledge of the EPRVs in the citrus crops, EPRV segments from the genomes of primitive, wild and cultivated citrus were identified in this study. A phylogenetic tree was built from the sequences of RNaseH (RH) domain. The possibility that the EPRV sequences were integrated in the Citrinae genomes during the divergence of the Citrinae species is discussed.

2. Materials and methods

2.1. Materials and EPRV sequences amplification

Leaves of Citrinae species (primitive species: atalantia (*Atalantia buxifolia*), also known as Chinese box orange; wild species: Ichang papeda (*C. ichangesis*) and Mangshan mandarin (*Citrus reticulata*); cultivars: citron (*C. medica*), pummelo (*C. maxima*), sweet orange (*C. sinensis*), clementine mandarin (*C. clementina*) and Ponkan mandarin (*Citrus reticulata*); Australian finger lime (*Microcitrus australasica*), Australian desert lime (*Eremocitrus glauca*), trifoliate (*Poncirus trifoliata*), kumquat (*Fortunella margarita*) and Hongkong kumquat (*Fortunella hindsii*) were sampled from the greenhouse of the National Center of Citrus Breeding, Huazhong Agricultural University (HZAU), Wuhan, China. Total DNA was extracted with CTAB methods (Cheng et al., 2003).

Primers were designed by Primer Premier 6.0 software (PREMIER Biosoft, Canada); the references of conserved sequences which contain the RNaseH (RH) domain were based on the BLAST results. Polymerase chain reaction (PCR) was performed in a volume of 50 µl containing 50 ng of genomic DNA, 1 U Phanta Max Super-Fidelity DNA Polymerase (Vazyme, Nanjing, China), 25 µl 2 × Phanta Max Buffer, 10 mM dNTP Mix, 20 µM of each primer pair and ddH_2O to 50 µl. Amplification was carried out as follows: 5 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 15 s at 56 °C and 2 min at 72 °C, and a final extension at 72 °C for 5 min. Amplification products were cloned in pTOPO-Blunt Simple vector (Aidlab, Beijing, China). Single clone was selected for Sanger sequencing.

2.2. Data analysis

A representative structure of the citrus EPRV was described in this study based on the study of Roy et al. and the EPRV (CarEPRV1, CarEPRV2 and CarEPRV3) sequences obtained from Carrizo citrange (Roy et al., 2014) were used as the query sequences to respectively BLAST against the genomes of six Citrinae species (atalantia, Ichang papeda, citron, pummelo, sweet orange and clementine mandarin). The overlaps of EPRV segments were excluded from the total number. The EPRV segments whose identities were greater than or equal to 80% and whose lengths were more than 300 bp were used for further analysis. The 2 kb sequences flanking the rearranged sites of EPRV segments were analyzed by RMBlast version 2.2.27 (Bigot, 2012). Results of all-by-all BLASTP were used as input to calculate the synteny ratio for pairs of Citrinae genomes using the i-ADHoRe software (version 2.4) (Simillion et al., 2008). The domain of aspartic protease (PR^{asp}) was predicted by PROSITE (http://prosite.expasy.org/). MEGA version 6.06 (Tamura et al., 2013) was used to build the phylogenetic tree by maximum likelihood method. RNaseH (RH) domain sequences of CarEPRV1, CarEPRV2 and CarEPRV3 were highly conserved and the RH domain sequence of CarEPRV1 was selected for phylogenetic analysis.

3. Results

3.1. Identification of EPRV sequences from citrus genomes

Citrus EPRV involve movement protein (MP) motif, Vps51 super family motif, zinc finger motif (ZnF), aspartic protease (PR^{asp}) motif, reverse transcriptase (RT) motif and RNaseH (RH) motif, and a representative structure was described (Fig. 1A). Hundreds of EPRV segments were found in the Citrinae genomes, with e value ranging from 5.0e^{-52} to 0 and identities ranging from 80% to 98.34% (Supplemental Table S1). A total of 245 EPRV segments were found in the genome of atalantia, 752 in the genome of papeda, 2060 in the genome of citron, 2036 in the genome of pummelo, 1034 in the genome of sweet orange and 598 in the genome of clementine mandarin (Table 1).

Table 1
Total number of EPRV segments in Citrinae genomes when the CarEPRV1, CarEPRV2, CarEPRV3 were used as the query sequences, respectively. The overlaps of EPRV segments were excluded from the total number. AtEPRV stands for EPRV from atalantia, IpEPRV for Ichang papeda, CiEPRV for citron, PuEPRV for pummelo, SoEPRV for sweet orange, ClEPRV for Clementine mandarin.

Reference	AtEPRV	IpEPRV	CiEPRV	PuEPRV	SoEPRV	ClEPRV
CarEPRV1	159	603	1619	1525	900	464
CarEPRV2	194	692	1781	1928	919	545
CarEPRV3	235	735	2015	2019	1010	594
Total	245	752	2060	2036	1034	598

Table 2
The primers which were applied to identify the EPRVs present in citrus genomes. RH1 was designed from atalantia, RH2 from Ichang papeda, RH3 from citron, RH4 from pummelo, RH5 from sweet orange and RH6 from clementine mandarin.

Primer name	Corresponding sample	Forward	Reverse
RH1	Atalantia	GCCATCATGCCCTGGTCTACA	GGTGAATGTTGGTGGCAAGAGG
RH2	Ichang papeda	TGACCATGAGAGCCACCAGAAG	TGGAGACACATCACCAGTTGCT
RH3	Citron	TTCCAGCATTCCTAGACCTCCTCT	GTAGCATACAAGGCGGTGAGACA
RH4	Pummelo	CTCCTTCAGACCTCCTGCTTGC	GCCATGTCCTGTGTCCAGTAGT
RH5	Sweet orange	TGACCATGAGAGCCACCAGAAG	AAGCACGCCGACAAGGACTT
RH6	Clementine mandarin	GTCTCCTTCAGATCTCTTGCTTGCTA	GCCTGTGGACGATGAATGGCTATA

Conserved sequences that contain RH domain were selected to design primers for confirming the presence of EPRV sequences in the citrus genomes. Forward primers located in the EPRV-flanking sequences and reverse primers located in the EPRV sequences region were designed, and six primer pairs (RH1, RH2, RH3, RH4, RH5 and RH6) were designed for the six Citrinae species (atalantia, Ichang papeda, citron, pummelo, sweet orange and clementine mandarin), respectively (Table 2, Supplemental Fig. S1). The PCR products confirmed the presence of EPRV sequences in all the six genomes (Fig. 1B).

3.2. The distribution of EPRV sequences in citrus genomes

The pummelo genome is a high-quality genome assembled from PacBio long reads (average of 10 kb), which facilitates the genomic analyses of repetitive sequences and retroelement analyses (Wang et al., 2017). The locations of EPRV segments in pummelo genome were determined as shown in Fig. 2. The distribution pattern indicates that the EPRV sequences were unevenly distributed across nine chromosomes and they were tightly clustered in some regions, particularly on chromosome 2 and 5. The 2-kb sequences flanking the EPRV segments in the six Citrinae genomes were further characterized, respectively. Simple repeats were the most abundant sequence types around EPRV segments in all the Citrinae genomes, with a number of 246 in atalantia, 504 in papeda, 1284 in citron, 1495 in pummelo, 796 in sweet orange and 574 in clementine. Transposable elements (TE) were also found around these regions and there were more types of TE in cultivated citrus (6 types) than in primitive citrus (3 types) (Table 3). EPRV segments were found in the predicted gene *Cg1g024630* of pummelo which harbored sequences that are probably related to *Poncirus trifoliata* CTV resistance gene locus (Supplemental Fig. S2).

3.3. Syntenic and phylogenetic analysis of EPRV sequences in Citrinae

Phylogenetic analysis was based on the RH domain of EPRV. Syntenic analyses of EPRV sequences that contained RNaseH (RH) domain in primitive species (atalantia), wild species (Ichang papeda) and cultivated species (pummelo) were performed. A total of 26 RH domain sequences (length > 300 bp and identity > 75%) were identified from different loci of atalantia. One syntenic block with the RH domain shared between the primitive species of atalantia and the cultivated species of pummelo was identified. The syntenic block was located in Scaffold 5407 (between gene *sb22504* and *sb22594*) of atalantia and on chromosome 2 (between gene *Cg2g012220* and *Cg2g012650*) of pummelo. Three syntenic blocks were identified between Ichang papeda and pummelo. Two species-specific blocks with RH domain were identified in atalantia but absent in both pummelo and Ichang papeda. The six primer sequences (Table 2) directed the amplification of PCR products in the other five species, respectively (Fig. 3). EPRV sequences (amplified by RH4 from pummelo and RH5 from sweet orange) were detected in all the six species. However, EPRV sequence from atalantia (amplified by RH1 from atalantia) only existed in the atalantia and Ichang papeda genomes, and EPRV sequence from clementine mandarin (amplified by RH6 from clementine mandarin) only existed in the citron, sweet orange and clementine mandarin.

In all, 605 RH domain sequences from the six Citrinae (atalantia, Ichang papeda, citron, pummelo, sweet orange, and clementine mandarin) genomes were isolated for phylogenetic analysis (Supplemental Fig. S3). Phylogenetic analysis showed that these RH domain sequences can be divided into two groups, group 1 contained EPRVs from five Citrinae species except atalantia, group 2 contained EPRVs from all the six Citrinae species. RH4 and RH5 were further used to amplify sequences from 7 Citrinae species (Ponkan mandarin, Australian finger lime, Australian desert lime, Mangshan mandarin, trifoliate, kumquat and Hongkong kumquat). The conserved EPRV sequences amplified by RH4 were used to isolate the RH domain sequences. Thirteen RH domain sequences were obtained and displayed e values ranging from $7.0e^{-70}$ to $1.0e^{-166}$, and identities ranging from 82% to 94% between these Citrinae species. A maximum likelihood phylogenetic tree was built based on the 13 nucleotide sequences of amplified RH domain and the RH domain sequence from Carrizo citrange (Roy et al., 2014) (Fig. 4). RH domain sequences of primitive species (atalantia) and wild species (Ichang papeda and Mangshan mandarin) were clustered away from other Citrinae species. RH domain sequence from Carrizo citrange was clustered with RH domain sequences of sweet orange and trifoliate. RH domain sequences from Australian finger lime and Australian desert lime were clustered together.

4. Discussion

This study shows that EPRV sequences are widely present in the Citrinae genomes, not only in primitive species (atalantia) and wild species (Ichang papeda), but also in cultivated species, such as sweet orange and clementine mandarin (Fig. 1). The availability of genomes of different Citrinae species (atalantia, pummelo, papeda, citron, sweet orange and clementine mandarin) make it easy to identify the locations of EPRV segments. The PacBio long-read assembly of the pummelo genome provides a good framework for the identification and characterization of EPRV sequences (Fig. 2). There were thousands of EPRV segments in many seed plant genomes (Diop et al., 2018). The genome-wide analysis indicated that there were hundreds of EPRV segments in the Citrinae genomes (Table 1). The number of EPRV segments in the Citrinae genomes was larger than that in petunia (100 segments) (Richert- Pöggeler et al., 2003). Isolation of RH domain sequences in the study further indicated that the other Citrinae species of *Fortunella*, *Eremocitrus*, *Poncirus*, *Microcitrus* also host the EPRV sequences.

A syntenic block with the RH domain sequence was identified in both the primitive atalantia and the cultivated species of pummelo, which suggested that there were conserved RH domain sequences in the Citrinae species. The EPRV sequences amplified by RH1 from atalantia showed that the amplified segments may be present only in atalantia and Ichang papeda genomes. However, the other EPRV sequences that amplified by RH4 from pummelo indicated that EPRV sequences may integrate into most of the Citrinae species/genomes (Fig. 3).

RH domain sequences were isolated from the Citrinae species and these sequences were highly conserved. Phylogenetic analysis showed that the EPRV from primitive species atalantia and wild species of Ichang papeda and Mangshan mandarin were clustered away from those of the other species (Fig. 4). The relationship of EPRVs from

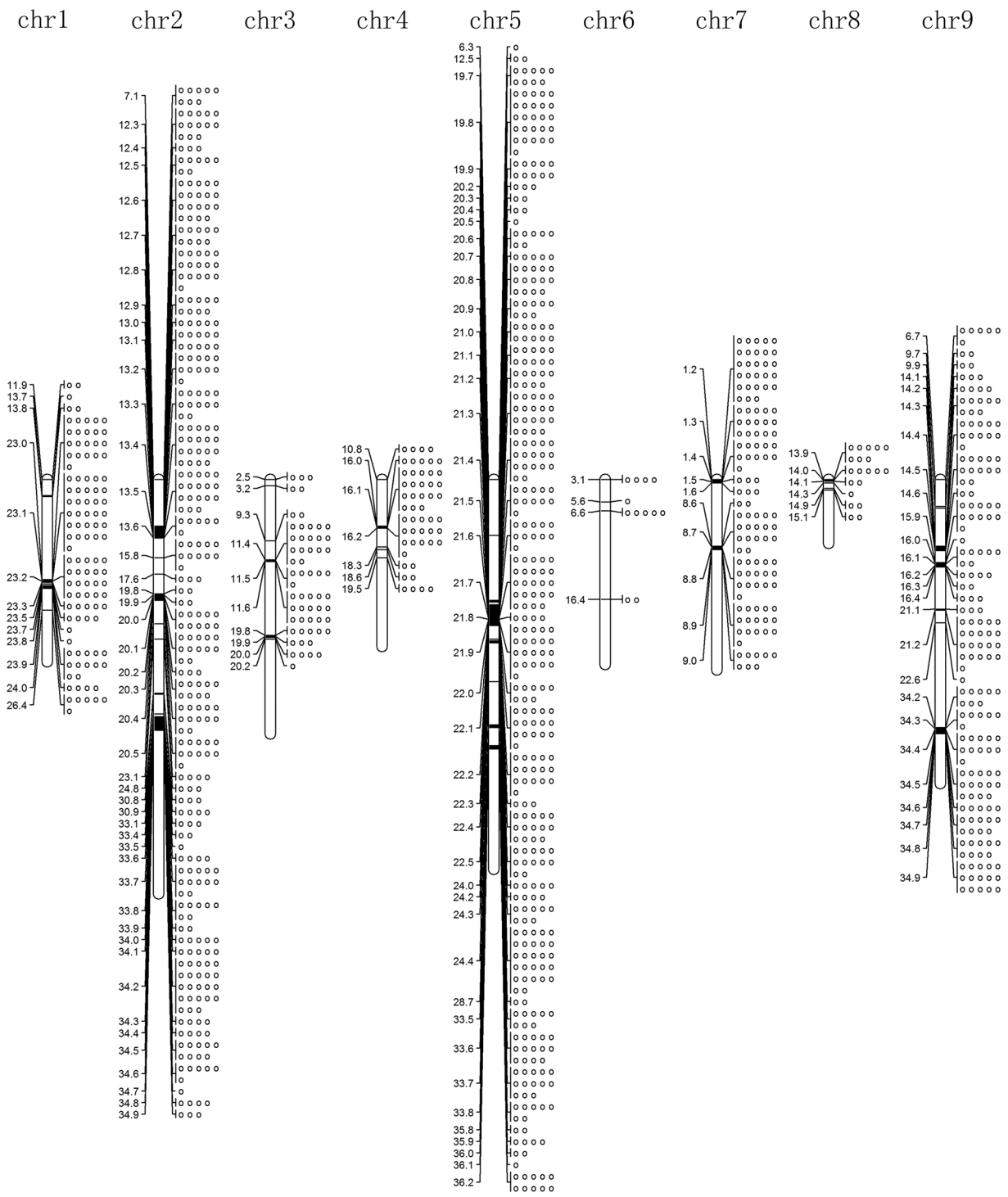

Fig. 2. Distributions of EPRV segments on the 9 chromosomes of pummelo. EPRV segments showed concentrated distribution on the chromosomes, especially on chromosome 2 and 5. Each 'o' represents one EPRV segment present at the chromosome locus.

atalantia and Ichang papeda were similar to Wang et al. (2017) that the atalantia and Ichang papeda were clustered away from other cultivated species. This suggested that there may have been an ancient integration event before the atalantia speciation, which was estimated to be approximately 15 million years ago (Carbonell-Caballero et al., 2015; Pfeil and Crisp, 2008). The particular presence of EPRV sequences in atalantia, Ichang papeda (amplified by RH1 from atalantia), and citron, sweet orange, clementine mandarin (amplified by RH6 from clementine mandarin) suggested that there were probably independent integration events during the evolution of Citrinae species as well. The time EPRVs integrated into the Citrinae genomes was possibly earlier than that of banana (Gayral et al., 2010) and rice (Chen et al., 2014). which were estimated at 0.64 Ma and 0.16 Ma respectively. RH domain sequence from Carrizo citrange was clustered with parental trifoliate and sweet orange, which provided evidence that the EPRV sequences can be transmitted to the offspring via hybridization.

A number of EPRV segments in the Citrinae genome were flanked by repetitive elements, especially simple repeats and LTR_*Gypsy* (Table 3). Simple repeats, such as AT repeats, can trap the episomal DNA in rice genomes (Kunii et al., 2004; Liu et al., 2012) and *Gypsy* elements often

Table 3
Summary of interspersed repeats and low complexity DNA sequences in 2-kb sequences which flank EPRV segments.

TE classification	Atalantia		Papeda		Citron		Pummelo		Sweet orange		Clementine	
	number of elements	length (bp)	number of elements	length (bp)	number of elements	length (bp)	number of elements	length (bp)	number of elements	length (bp)	number of elements	length (bp)
LINE	–	–	2	869	5	1672	4	696	4	1178	2	532
LTR_Copia	9	3879	47	17783	45	18038	59	27939	60	26089	26	11290
LTR_Gypsy	85	49746	119	66950	205	118251	229	148096	182	82420	97	64675
DNA_hAT	2	120	18	1493	14	2767	10	2374	12	1598	3	212
Tc1	–	–	2	591	1	239	9	1863	3	775	4	474
En-Spm	–	–	–	–	1	631	–	–	1	546	2	550
Simple repeats	246	11660	504	25215	1284	62532	1495	64758	796	41544	574	26814
Low complexity	66	3504	145	8395	396	22813	320	17479	159	8722	171	9051

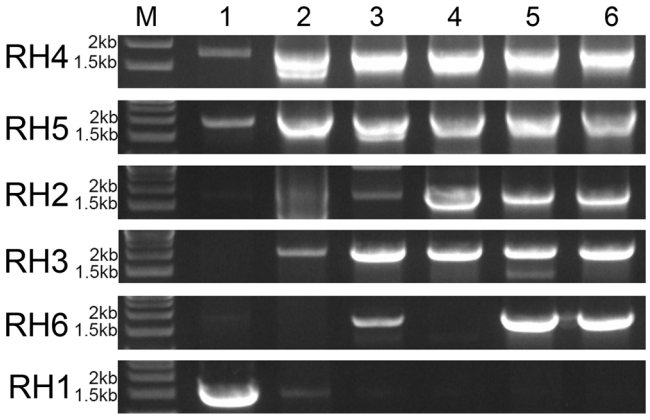

Fig. 3. Electrophoretic analysis of PCR products from primers RH4, RH5, RH2, RH3, RH6, and RH1, which were designed from EPRV sequence of pummelo, sweet orange, Ichang papeda, citron and atalantia, respectively. The DNA samples of atalantia (1), Ichang papeda (2), citron (3), pummelo (4), sweet orange (5) and clementine mandarin (6) were indicated on the top of the figure.

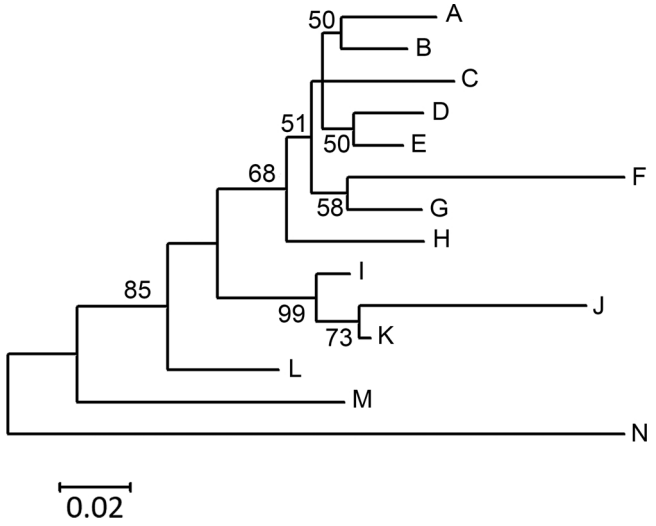

Fig. 4. Phylogenetic analysis of RNaseH (RH) domain sequences from Citrinae species by maximum likelihood methods. A to kumquat, B to Hongkong kumquat, C to clementine mandarin, D to pummelo, E to Ponkan mandarin, F to Australian desert lime, G to Australian finger lime, H to citron, I to sweet orange, J to Carrizo citrange (Roy et al., 2014), K to trifoliate, L to Mangshan mandarin, M to Ichang papeda and N to atalantia. Bootstrap values (1000 replicates) with only values > 50% are shown on the branches.

flanked the *Nto*EPRV of tobacco (Gregor et al., 2004). Therefore, EPRV sequences probably tended to be trapped by flanking repetitive elements when rearranged into the Citrinae genomes. The EPRV segments in the gene structure suggested that the EPRV sequences may play potential functions in the Citrinae species. CitPRV was detected in the sweet orange symptomatic of citrus sudden death from Brazil, which suggests that the EPRV is likely to be associated with the symptoms of citrus sudden death (Matsumura et al., 2017). Therefore, EPRV sequences may play two functions in the Citrinae life cycle. One is that the EPRV sequences as the normal part of Citrinae genomes, serve as a genetic pool for the generation of new genes or become tamed as the elements like TE; the other is that if the entire EPRV sequences have integrated into the Citrinae genomes, it could be reactivated under certain conditions to infect the host. Therefore, the identification of the EPRV sequences from Citrinae lays a foundation for understanding the function and evolution of EPRVs.

5. Conclusions

Hundreds of segments were integrated into Citrinae genomes. They clustered as hot regions in the rearrangement events on the Citrinae chromosomes. TE elements, especially simple repeats, flanked the EPRV segments. The ancient integration events happened probably before the speciation of Citrinae species, and there were other integration events during the evolution of Citrinae species as well.

Authors' contributions

Huiwen Yu designed the primers, made the phylogenetic analysis and drafted this manuscript. Xia Wang did the bioinformatic analysis. Zhihao Lu and Yuantao Xu isolated the RNaseH domain sequences from the Citrinae species. Qiang Xu and Xiuxin Deng designed and coordinated the project. All authors proofread the final manuscript.

Conflicts of interest

No conflicts of interest exist for any of the authors.

Acknowledgements

This project was supported by the National Natural Science Foundation of China (31572105 and 31330066), the key project of Hubei provincial Natural Science Foundation (2015CFA031), and the Ministry of Education of China (IRT13065).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: https://doi.org/10.1016/j.virusres.2018.05.018.

References

Barkley, N.A., Roose, M.L., Krueger, R.R., Federici, C.T., 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theor. Appl. Genet. 112, 1519–1531.

Bertsch, C., Beuve, M., Dolja, V.V., Wirth, M., Pelsy, F., Herrbach, E., Lemaire, O., 2009. Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. Biol. Direct 4, 21.

Bigot, Y., 2012. Mobile Genetic Elements: Protocols and Genomic Applications, vol. 859. Humana Press, pp. 1–308.

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., Dopazo, J., 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. Mol. Biol. Evol. 32, 2015–2035.

Chabannes, M., Baurens, F.C., Duroy, P.O., Bocs, S., Vernerey, M.S., Rodier-Goud, M., Barbe, V., Gayral, P., Iskra-Caruana, M.L., 2013. Three infectious viral species lying in wait in the banana genome. J. Virol. 87, 8624–8637.

Chabannes, M., Iskra-Caruana, M.-L., 2013. Endogenous pararetroviruses—a reservoir of virus infection in plants. Curr. Opin. Virol. 3, 615–620.

Chen, S., Kishima, Y., 2016. Endogenous pararetroviruses in rice genomes as a fossil record useful for the emerging field of palaeovirology. Mol. Plant Pathol. 17, 1317–1320.

Chen, S., Liu, R., Koyanagi, K.O., Kishima, Y., 2014. Rice genomes recorded ancient pararetrovirus activities: virus genealogy and multiple origins of endogenization during rice speciation. Virology 471, 141–152.

Cheng, Y., Guo, W., Yi, H., Pang, X., Deng, X., 2003. An efficient protocol for genomic DNA extraction from Citrus species. Plant Mol. Biol. Rep. 21, 177–178.

Diop, S.I., Geering, A., Alfama-Depauw, F., Loaec, M., Teycheney, P.Y., Maumus, F., 2018. Tracheophyte genomes keep track of the deep evolution of the Caulimoviridae. Sci. Rep. 8, 572.

Eid, S., Saar, D.E., Druffel, K.L., Pappu, H.R., 2011. Plant pararetroviral sequences in wild Dahlia species in their natural habitats in Mexican mountain ranges. Plant Pathol. 60, 378–383.

Federici, C., Fang, D., Scora, R., Roose, M., 1998. Phylogenetic relationships within the genus Citrus (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. Theor. Appl. Genet. 96, 812–822.

Garcia-Lor, A., Curk, F., Snoussi-Trifa, H., Morillon, R., Ancillo, G., Luro, F., Navarro, L., Ollitrault, P., 2013. A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. Ann. Bot. 111, 1–19.

Gayral, P., Blondin, L., Guidolin, O., Carreel, F., Hippolyte, I., Perrier, X., Iskra-Caruana, M.-L., 2010. Evolution of endogenous sequences of *banana streak virus*: what can we learn from banana (*Musa* sp.) evolution? J. Virol. 84, 7346–7359.

Gayral, P., Noa-Carrazana, J.-C., Lescot, M., Lheureux, F., Lockhart, B.E., Matsumoto, T., Piffanelli, P., Iskra-Caruana, M.-L., 2008. A single *Banana streak virus* integration event in the banana genome as the origin of infectious endogenous pararetrovirus. J. Virol. 82, 6697–6710.

Gregor, W., Mette, M.F., Staginnus, C., Matzke, M.A., Matzke, A.J., 2004. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. Plant Physiol. 134, 1191–1199.

Harper, G., Osuji, J.O., Heslop-Harrison, J.S., Hull, R., 1999. Integration of Banana Streak Badnavirus into the *Musa* genome: molecular and cytogenetic evidence. Virology 255, 207–213.

Harper, G., Richert-Pöggeler, K.R., Hohn, T., Hull, R., 2003. Detection of petunia vein-clearing virus: model for the detection of DNA viruses in plants with homologous endogenous pararetrovirus sequences. J. Virol. Methods 107, 177–184.

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

Kunii, M., Kanda, M., Nagano, H., Uyeda, I., Kishima, Y., Sano, Y., 2004. Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: implications for integration and evolution. BMC Genom. 5, 80.

Lheureux, F., Carreel, F., Jenny, C., Lockhart, B., Iskra-Caruana, M., 2003. Identification of genetic markers linked to banana streak disease expression in inter-specific *Musa* hybrids. Theor. Appl. Genet. 106, 594–598.

Liu, R., Koyanagi, K.O., Chen, S., Kishima, Y., 2012. Evolutionary force of AT-rich repeats to trap genomic and episomal DNAs into the rice genome: lessons from endogenous pararetrovirus. Plant J. 72, 817–828.

Lockhart, B.E., Menke, J., Dahal, G., Olszewski, N.E., 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. J. Gen. Virol. 81, 1579–1585.

Marco, Y., Howell, S.H., 1984. Intracellular forms of viral DNA consistent with a model of reverse transcriptional replication of the cauliflower mosaic virus genome. Nucleic Acids Res. 12, 1517–1528.

Matsumura, E., Coletta-Filho, H., Nouri, S., Falk, B., Nerva, L., Oliveira, T., Dorta, S., Machado, M., 2017. Deep sequencing analysis of RNAs from Citrus plants grown in a Citrus sudden death-affected area reveals diverse known and putative novel viruses. Viruses 9, 92.

Mette, M.F., Kanno, T., Aufsatz, W., Jakowitsch, J., van der Winden, J., Matzke, M.A., Matzke, A.J.M., 2002. Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. EMBO J. 21, 461.

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

Nicolosi, E., Deng, Z., Gentile, A., La Malfa, S., Continella, G., Tribulato, E., 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor. Appl. Genet. 100, 1155–1166.

Pahalawatta, V., Druffel, K., Pappu, H., 2008. A new and distinct species in the genus *Caulimovirus* exists as an endogenous plant pararetroviral sequence in its host, Dahlia variabilis. Virology 376, 253–257.

Pfeil, B.E., Crisp, M.D., 2008. The age and biogeography of Citrus and the orange subfamily (Rutaceae: Aurantioideae) in Australasia and New Caledonia. Am. J. Bot. 95, 1621–1631.

Richert-Pöggeler, K.R., Shepherd, R.J., 1997. Petunia vein-clearing virus: a plant pararetrovirus with the core sequences for an integrase function. Virology 236, 137–146.

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

Roy, A., Shao, J., Schneider, W.L., Hartung, J.S., Brlansky, R.H., 2014. Population of endogenous pararetrovirus genomes in Carrizo citrange. Genome Announc. 2 e01063-01013.

Seal, S., Turaki, A., Muller, E., Kumar, P.L., Kenyon, L., Filloux, D., Galzi, S., Lopez-Montes, A., Iskra-Caruana, M.-L., 2014. The prevalence of badnaviruses in West African yams (*Dioscorea cayenensis-rotundata*) and evidence of endogenous pararetrovirus sequences in their genomes. Virus Res. 186, 144–154.

Simillion, C., Janssens, K., Sterck, L., Van de Peer, Y., 2008. i-ADHoRe 2.0: an improved tool to detect degenerated genomic homology using genomic profiles. Bioinformatics 24, 127–128.

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

Swingle, W.T., 1967. The botany of Citrus and its wild relatives. Citrus Ind. 1, 190–430.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Umber, M., Filloux, D., Muller, E., Laboureau, N., Galzi, S., Roumagnac, P., Iskra-Caruana, M.L., Pavis, C., Teycheney, P.Y., Seal, S.E., 2014. The genome of African yam (*Dioscorea cayenensis-rotundata* complex) hosts endogenous sequences from four distinct badnavirus species. Mol. Plant Pathol. 15, 790–801.

Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., Yu, H., Yang, X., Lan, H., Wang, N., Wang, L., Xu, J., Jiang, X., Xie, Z., Tan, M., Larkin, R.M., Chen, L.L., Ma, B.G., Ruan, Y., Deng, X., Xu, Q., 2017. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat. Genet. 49, 765–772.

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., Takita, M.A., Labadie, K., Poulain, J., Couloux, A., Jabbari, K., Cattonaro, F., Fabbro, C.D., Pinosio, S., Zuccolo, A., Chapman, J., Grimwood, J., Tadeo, F.R., Estornell, L.H., Muñoz-Sanz, J.V., Ibanez, V., Herrero-Ortega, A., Aleza, P., Pérez-Pérez, J., Ramón, D., Brunel, D., Luro, F., Chen, C., Farmerie, W.G., Desany, B., Kodira, C., Mohiuddin, M., Harkins, T., Fredrikson, K., Burns, P., Lomsadze, A., Borodovsky, M., Reforgiato, G., Freitas-Astúa, J., Quetier, F., Navarro, L., Roose, M., Wincker, P., Schmutz, J., Morgante, M., Machado, M.A., Talon, M., Jaillon, O., Ollitrault, P., Gmitter, F., Rokhsar, D., 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. 32, 656–662.

Xu, Q., Chen, L.L., Ruan, X., Chen, D., Zhu, A., Chen, C., Bertrand, D., Jiao, W.B., Hao, B.H., Lyon, M.P., Chen, J., Gao, S., Xing, F., Lan, H., Chang, J.W., Ge, X., Lei, Y., Hu, Q., Miao, Y., Wang, L., Xiao, S., Biswas, M.K., Zeng, W., Guo, F., Cao, H., Yang, X., Xu, X.W., Cheng, Y.J., Xu, J., Liu, J.H., Luo, O.J., Tang, Z., Guo, W.W., Kuang, H., Zhang, H.Y., Roose, M.L., Nagarajan, N., Deng, X.X., Ruan, Y., 2013. The draft genome of sweet orange (*Citrus sinensis*). Nat. Genet. 45, 59–66.

Yang, Z.N., Ye, X.R., Molina, J., Roose, M.L., Mirkov, T.E., 2003. Sequence analysis of a 282-kilobase region surrounding the Citrus Tristeza virus resistance Gene (*Ctv*) Locus in *Poncirus trifoliata* L. Raf. Plant Physiol. 131, 482–492.