Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

ENDOGENOUS PARARETROVIRUSES SEQUENCES CAN BE USED AS MARKERS TO DIFFERENTIATE FOUR AUBRIETA SPECIES

OSAMAH NADHIM ALISAWI*
Department of Plant protection, College of Agriculture, University of Kufa, Najaf, Iraq.

JOTYAR J. MUHAMMED
Department of Forestry, College of Agricultural Engineering sciences, University of Duhok, Duhok, Kurdistan region-Iraq.

PAT HESLOP-HARRISON
Department of Genetics and genome biology, University of Leicester, Leicester, United Kingdom.

Abstract

The genus Aubrieta Adan. (Brassicaceae) is widely distributed and diverges across different elevations. We aimed to study aspects of the genome organization and components to understand evolution and differentiation. Endogenous pararetroviruses (EPRVs) were examined in whole-genome of four Aubrieta species using high-throughput DNA sequencing and bioinformatics. Two genera of caulimovirid sequences have been found in the four examined genomes, caulimoviruses, and florendoviruses with four members each named as Caulimovirus-AAn, Caulimovirus-AEu, Caulimovirus-AGr, Caulimovirus-ASc, AanaV, AeurV, AgraV, and AscaV. The full length of Caulimovirus-AAn, Caulimovirus-AEu, Caulimovirus-AGr, and Caulimovirus-ASc were 7579, 6726, 7223, and 6609 bp, while the florendoviruses, AanaV, AeurV, AgraV, and AscaV were 6675, 6888, 6702, and 6638 bp respectively. The integrants encode four coding domains; movement protein (MP), two domains of reverse transcriptase (RT and RVT), and RNaseH (RH), and except Caulimovirus- AEu, all caulimoviruses are inverted from 3' to 5', while the florendoviruses are arranged from 5' to 3'. Variable numbers of genome proportions and copies have been recorded for these integrants reporting A. eurobsens as more accessible genome in the case of florendoviruses, while caulimovirus-like sequence was most abundant in the genome of the A. anamasica. The genome of A. scardica was limited in EPRVs existence comparing to A. anamasica which having a good marker to separate these species. The phylogenetic tree confirms the close relationships of each group of the EPRVs as they are arranged next to their genus members.

Keyword: Aubrieta Genomes, Bioinformatics, Endogenous Pararetroviruses, Next Generation Sequencing (NGS).

1. INTRODUCTION

The Brassicaceae (Mustard family) includes many commercially important vegetables, condiments, oilseeds, and fodder crops, along with *Arabidopsis thaliana*, used as a model plant from molecular biology (Warwick 2011). This family including wild plants has been infected by a wide range of plant viruses such as *Turnip mosaic virus* (TuMV), *Cucumber mosaic virus*, *Brassica yellows virus*, *Pelargonium zonate spot virus* and *Arabidopsis halleri partitivirus* 1 (Krajacic *et al*., 2001; Holub 2008; Kamitani *et al*., 2019). Furthermore, it is recognized as the main host of *Cauliflower mosaic virus*, the typical member of the family Caulimovirideae (Shafiq *et al*., 2020). *Aubrieta spp*., a

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

perennial herbaceous plant of the Brassicaceae family, has been reported to be a host to multiple plant viruses like what Shukla and Schmelzer (1970) mentioned about *Aubrieta deltoidea* that has symptomless infection resulting from *Cucumber mosaic virus*. Fletcher (1987) recorded that *Aubrieta* is one of *Cucumber mosaic virus* hosts, in a study aimed to register new hosts for plant viruses in New Zealand. Furthermore, *Aubrieta* cvs Blue Emperor and Gloriosa showed flower color-breaking symptoms caused by *Cucumber mosaic virus*, lacking satellite RNA-5 (Krajacic *et al*., 2001). Weeds of this family are known to host *Turnip mosaic virus* (TuMV). However, *Turnip mosaic virus* was recently reported in Iran, it was detected in Brassicaceae reservoir weeds (532 samples) of eight provinces of Iran displaying virus-like symptoms in 64% of the tested plants such as *Rapistrum rugosum*, *Sisymberium loeselii*, *S. irio*, and *Hirschfeldia incana* (Farzadfar *et al*., 2005; Farzadfar *et al*., 2009). Plant virus studies tend to focus on crop diseases, and natural vegetation viruses are little studied (Jones 2016). Nevertheless, infections with *Cucumber mosaic virus* (CMV), *Turnip mosaic virus* (TuMV), *Pelargonium zonate spot virus*, *Brassica yellows virus*, and *Arabidopsis halleri partitivirus 1* were isolated from the two perennial species, *Arabidopsis halleri* subsp. *gemmifera* and *Rorippa indica* (Kamitani *et al*., 2019). A partial sequence of a putative novel virus was constructed from *Arabis fragellosa* by *de novo* assembly, and viral species composition and infection rates varied according to plant species and site composition, as well as virus transmission across multiple years has been confirmed for *A. halleri*. The phylogenetic analysis of TuMV and CMV found that viruses from wild Brassicaceae formed a clade with strains from crop plants. Thus, wild and crop plants shared the same viruses. In natural plant populations, virus distribution is determined by the combination of virus and host life histories. The study of viral distribution in natural plant communities enhances our understanding of the ecology of plant viruses (Kamitani *et al*., 2019). Endogenous pararetroviruses (EPRVs), taxon *Caulimoviridae*, are retroelements that contain reverse transcriptase, and likely arose around 1.6 billion years ago from RNA to DNA. Their origins are probably traced to retroviruses and LTR retrotransposons rather than relying on nuclear DNA replication after transcription in the host nucleus, viral DNA is synthesized in the cytoplasm on terminally redundant RNA, including inter-and intra-molecular recombination (Richert-Pöggeler *et al*., 2021). The elements have colonized plant genomes extensively as a result of endogenization events (Geering *et al*., 2014). Most EPRVs can no longer replicate through reverse transcription or infect cells due to fragmentation and rearrangement. Extant EPRVs are associated with closely related retrotransposons and often persist for millions of years in plant genomes (Schmidt *et al*., 2021). Plant genomes increasingly show EPRV presence as a result of advances in genome sequencing techniques (Geering *et al*. 2014); however, EPRVs have not been discovered in *Aubrieta* genomes yet. This study conducted to reveal integrated elements of endogenous pararetroviruses in four genomes belong to four species of *Aubrieta* using next generation sequencing and bioinformatics techniques.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

2. PROCEDURE

2.1. Plant material

Fresh leaves of four *Aubrieta* species; *A. gracilis*, *A. scardica*, *A. erubescens,* and *A. anamasica* were collected from the cultivated plants in Leicester University Botanic Garden. One gram of young leaves was put in small plastic bags with silica gel (1:20) for 24 hours at 60°C for drying.

2.2. DNA isolation from *Aubrieta* species

To extract DNA, the young leaves of the four *Aubrieta* species were used using the cetyl-trimethylammonium bromide (CTAB) method (Doyle and Doyle 1990), with minor modifications (The CTAB buffer was incubated for 45 min and the DNA was spun down at 735 g for 5 min). The extracted samples were sent for sequencing at the Interdisciplinary Centre for Biotechnology Research, University of Florida, USA. The species were sequenced by using the Hiseq500 2x150bp reads technique based on the manufacturer's procedure.

2.3. Graph-based clustering of raw reads

The graph-based read clustering program RepeatExplorer was used to explore and characterize EPRV clusters and repetitive DNA sequences in the data of last-generation sequencing (Novák *et al*., 2013). The RepeatExplorer2 clustering (Galaxy Version 2.3.8.1) was applied, and the selected taxon and protein domain database version (REXdb) was viridplantae version 3.0. The generated clusters from the Repeat Explorer were extracted, and then the whole contigs of each cluster were submitted to the Repbase dataset (Jurka *et al*., 2005), and the Basic Local Alignment Search Tool (Altschul *et al*., 1990). Further, these sequences were aligned to published plant virus sequences (DPVweb) (Adams and Antoniw 2005). Finally, the suggested viruses were identified at the levels of genus and species by sequence alignments.

2.4. Map to reference

The Illumina raw reads were aligned against the identified viruses and showed the number of assembled reads, the total used reads, and also the frequently overlapped reads in one report. This data was applied to calculate copy numbers and genome proportions as follows: 1- Copy number: number of assembled reads x read length/reference sequence length.2- Genome proportion: number of aligned reads / numbers of total NGS reads x 100 (Mustafa *et al*., 2018).

2.5. Phylogenetic analysis

MEGA 11 (Tamura *et al*., 2013) was used with the maximum likelihood (ML) method to find a robust phylogeny model. A Clustal W alignment was applied to extract sequences (about 7000bp for each). Then, the tree was reconstructed with General Time

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

Reversible (GTR). Fifteen endogenous viruses were applied to reconstruct the phylogeny tree, and *Petunia vein clearing virus* (PVCV) was the outgroup member.

3. RESULTS

The numbers of Illumina clean 150bp sequence reads were 16121154, 15454208, 13441852, and 8436374 for *Aubrieta anamasica*, *Aubrieta eurobsens*, *Aubrieta gracillis* and *Aubrieta scardica* respectively. The graph-based clustering with RepeatExplorer revealed two clusters of EPRVs in the four genomes of *Aubrieta.* The clusters were extracted and mapped against the genome clean reads using the Burrows-Wheeler algorithm to obtain the complete sequence of each EPRV through *Aubrieta* genome clean reads. The analysis showed that two genera of caulimovirid sequences were found in the four examined genomes, *Caulimovirus* and *Florendovirus* with four members each. The *Caulimovirus* elements were named based on Repbase regulation as *Caulimovirus-AAn, Caulimovirus-AEu, Caulimovirus-AGr,* and *Caulimovirus-ASc,* while florendoviruses were named according to Geering *et al*. (2014) as *AanaV, AeurV, AgraV, AscaV.* The assembled reads of the caulimovirus-like sequences were 28647, 13005, 5086, 5905 for *Caulimovirus-AAn*, *Caulimovirus-AEu*, *Caulimovirus-AGr*, and *Caulimovirus-ASc*, while the florendovirus-like sequences were 18694, 25465, 12442, 4716 for *AanaV*, *AeurV*, *AgraV*, and *AscaV* respectively (Table 1). The full length of *Caulimovirus-AAn*, *Caulimovirus-AEu*, *Caulimovirus-AGr*, and *Caulimovirus-ASc* was 7579, 6726, 7223, and 6609 bp, while it was 6675, 6888, 6702, and 6638 bp for *AanaV*, *AeurV*, *AgraV*, and *AscaV* respectively. All sequences have been deposited in the Repbase dataset (https://www.girinst.org/repbase/). These components encode four coding domains; movement protein (MP), two domains of reverse transcriptase (RT and RVT), and RNaseH (RH). Except for *Caulimovirus-AEu,* all caulimoviruses were inverted from 3’ to 5’ while the florendoviruses were more consistent and arranged from 5’ to 3’ (Fig. 1). The genome proportion of *Caulimovirus-AAn*, *Caulimovirus-AEu*, *Caulimovirus-AGr*, and *Caulimovirus-ASc* were 0.177, 0.08, 0.03, and 0.06, while the copy numbers were 566, 290, 106, and 122 respectively. The genome proportion of florendoviruses, *AanaV*, *AeurV*, *AgraV*, and *AscaV* were 0.11, 0.16, 0.09, and 0.05 with 463, 674, 308, and 106 copy numbers respectively (Table 1). The phylogeny confirms the close relationship of each group of EPRVs as they are arranged next to their genus members (Fig. 1).

4. DISCUSSION

Endogenous pararetrovirus EPRV sequences incorporated in the host nuclear genome have been considered as elements with high impacts for the plant. Using bioinformatic tools, we investigated in their nature, diversity, and organization with high throughput sequencing techniques within *Aubrieta* genomes. In general, the integrants are less understood in plants than they are in animals (eg Mustafa *et al*., 2022), despite the fact that they have been present in plant genomes since 20-34 MYA (Geering *et al.*, 2014). Here, we highlighted more facts in a new host plant, *Aubrieta,* to understand the

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

widespread nature of this genome component and assist with explaining the genetic and geographical diversity and phylogeny. Interestingly, EPRVs were clearly distinguished in their genome proportions and copy numbers cross different species taking into consideration complete genome sizes. (Table 1). *A. eurobsens* had relatively more copies of the florendovirus-like sequence (*AeurV*) integration while caulimovirus-like sequences (*Caulimovirus-AAn*) were most abundant in the genome of *A. anamasica* (Table 1). *A*. *scardica* was limited in EPRVs existence, unlike *A. anamasica* and acting as a good marker sequence to separate these species. According to Alisawi (2019), petunia genomes were highly variable in proportions of EPRV sequences, and these sequences were widely distinguished over their loci on chromosomes, suggesting these elements could be a new marker to recognize plant species. Further, genome diversification could happen with time through fixation, accumulation, and incomplete excision of inserted retrotransposons (Carvalho *et al.,* 2010; Saeidi *et al*., 2008; Salina *et al*., 2011; Wessler, 2006; Heitkam and Schmidt 2009). Additionally, the widespread distribution within chromosomes, high copy numbers, and the ubiquitous nature of retrotransposons has made these components quite typical for the development of DNA-based markers. (Flavell *et al*., 1992; Teo *et al*., 2005). As they are very closely related to retrotransposons, endogenous pararetroviruses integrate into host genomes by hitchhiking on the retrotransposons and the hybrid between the two is potentially capable of undergoing real transposition through this route (Hohnn 1994; Froissart *et al*., 2005; Richert-Pöggeler *et al*., 2003; Gregor *et al*., 2004, Staginnus *et al*., 2007). Although EPRVs have integrated from a viral sequence within genomes, their existence has a huge impact on host genome such as methylation status and chromosomal rearrangements (Hohn *et al*., 2008). The florendoviruses have been more consistent in genome organization than caulimovirus-like sequences, although phylogenetic relationship was confirmed their similar sequences and coding regions (Fig. 1). In recent years, the molecular phylogenetic of the genus *Aubrieta* has become well known. Phylogenetically, the *Aubrieta* are positioned in the tribe Arabideae (Warwick *et al.*, 2010) and clustered with the subalpine genus *Arabis* (Karl *et al*., 2012). Muhammed (2017) analysed that *Aubrieta* species were divided into five gene pools, from Neat Eat to Sicily crossing Anatolia, Bulgaria and Greece. In this study, the *Caulimovirus* and *Florendovirus* elements within *Aubrieta* genomes were analysed phylogenetically. The results showed that the tree is polytomic with two significant clades (100%) bootstrap. The first clade comprises of the *A. eurobescens* subclade, and *A. gracilis*, *A. anamasica* and *A. scardica* subclade (Fig 1). The Greece costal endemic *A. eurobescens* is clearly separated as a clade due to their geographical position in the North East of Greece. This result agrees with Yüzbaşıoğlu *et al*. (2015) when combined nuclear (ITS) and chloroplast (trnL-F) genes. On the second clade, *A. scardica* was cluster with *A. eurobescens* (92%) bootstrap on florendovirus, whereas cluster with *A. anamasica* (99%) bootstrap on *Caulimovirus*. This variation probably happened as a result of the overlapping distribution features of the crypic taxa *A. scardica*, and the overall results partially agreed with Koch *et al.* (2017) and Muhammed (2017). The present research will subsequently use the chromosomal and genetic

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

maps that are planned to achieve by the future genome works through molecular mapping and *in situ* hybridization to explain how are these genomes differentiated and geographically distributed.

Table 1: List of EPRVs in *Aubrieta* genomes shows genome proportions and copy numbers extracted by map to reference tool

EPRV	Host	Genus	Genome proportion	Copy number	Assembled reads	clean reads
Caulimovirus-AAn	Aubrieta anamasica	Caulimovirus	0.177	566	28647	16121154
Caulimovirus-AEu	Aubrieta eurobsens	Caulimovirus	0.08	290	13005	15454208
Caulimovirus-AGr	Aubrieta gracillis	Caulimovirus	0.03	106	5086	13441852
Caulimovirus-ASc	Aubrieta scardica	Caulimovirus	0.06	122	5905	8436374
AanaV	Aubrieta anamasica	Florendovirus	0.11	463	18694	16121154
AeurV	Aubrieta eurobsens	Florendovirus	0.16	674	25465	15454208
AgraV	Aubrieta gracillis	Florendovirus	0.09	308	12442	13441852
AscaV	Aubrieta scardica	Florendovirus	0.05	106	4716	8436374

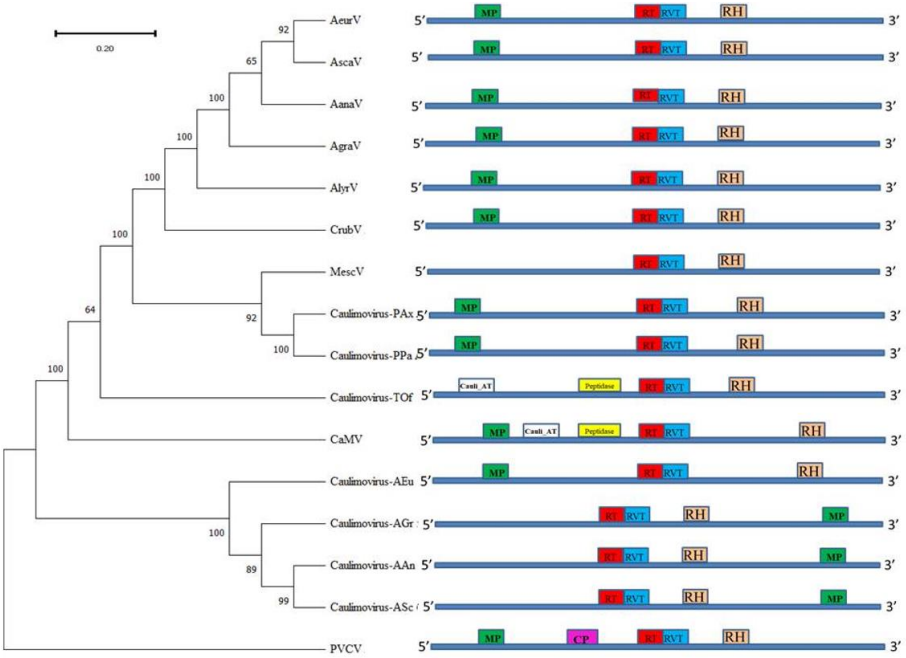

Fig.1 The Maximum Likelihood (ML) *Caulimovirus* and *Florendovirus* trees of four *Aubrieta* taxa. The number above each branch is the bootstrap if it has > 50% support. The genome arrangement shows four coding regions (MP, RT, RVT, and RH) in each integrant, and two more domains (Cauli-AT and Peptidase) are found in CaMV and *Caulimovirus-TOf*. The outgroup member is *Petunia vein clearing virus* (PVCV).

References

1. Alisawi, O. N. (2019): *Virus integration and tandem repeats in the genomes of Petunia* (Doctoral dissertation, University of Leicester).

ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

2. Al-Shehbaz, IA. (2012): A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon,* 61: 931–954.
3. Ančev, M. and Goranova, V. (2009): *Aubrieta* (Brassicaceae) in the Bulgarian flora. *Phytologia Balcanica*,15(1), 43-50.
4. Bendich, A. J. (2004): Circular chloroplast chromosomes: the grand illusion. *The Plant Cell*,16(7), 1661-1666.
5. Carvalho, M., Ribeiro, T., Viegas, W., Morais-Cecílio, L., & Rocheta, M. (2010): Presence ofenv-like sequences inQuercus suber retrotransposons. *Journal of applied genetics*, *51*(4), 461-467.
6. Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L. and Jansen, R. K. (2006): The complete chloroplast genome sequence of Pelargonium× hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution*,23(11), 2175-2190.
7. Demesure, B., Sodzi, N. and Petit, R. (1995): A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular ecology*,4(1), 129-134.
8. Duminil, J. and Di Michele, M. (2009): Plant species delimitation: a comparison of morphological and molecular markers. *Plant Biosystems*,143(3), 528-542.
9. Doyle, J. and Doyle, J. (1990): Isolation of DNA from small amounts of plant tissues. *BRL Focus* 12, V15.
10. Flavell, A.J., Dunbar, E., Anderson, R., Pearce, S.R., Hartley, R., Kumar, A. (1992): Ty1– copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Research*, 20, 3639-44.
11. Fletcher, J. D. (1987): New plant disease records in New Zealand: additional hosts of alfalfa mosaic virus and cucumber mosaic virus. *New Zealand journal of agricultural research*, *30*(4), 505-506.
12. Franzke, A., German, D., Al-Shehbaz, IA., Mummenhoff, K. (2009): *Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon*, 58: 425–437.
13. Franzke, A., Lysak, MA., Al-Shehbaz, IA., Koch, MA., Mummenhoff, K. (2011): Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science,* 16: 108–116.
14. 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 Physiology,* 134, 1191-9.
15. Haudry, A., Platts, A. E., Vello, E., Hoen, D. R., Leclercq, M., Williamson, R. J., Forczek, E., Joly-Lopez, Z., Steffen, J. G. and Hazzouri, K. M. (2013): An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics*, 45(8), 891-898.
16. Heitkam, T., & Schmidt, T. (2009): BNR–a LINE family from Beta vulgaris–contains a RRM domain in open reading frame 1 and defines a L1 sub-clade present in diverse plant genomes. *The Plant Journal*, *59*(6), 872-882.
17. Heslop-Harrison, J.S. (2000): Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *The Plant Cell*, 12(5), 617-635.
18. Hollingsworth, P. M., Graham, S. W., Little, D. P. (2011): Choosing and using a plant DNA barcode. *PloS one*,6(5), e19254.
19. Holub, E. B. (2008): Natural history of Arabidopsis thaliana and oomycete symbioses. *The Downy Mildews-Genetics, Molecular Biology and Control*, 91-109.

ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

20. Hohn, T. (1994): Recombination of a plant pararetrovirus: Cauliflower mosaic virus. In: *Homologous Recombination and Gene Silencing in Plants* (pp. 25-38). Springer, Dordrecht.
21. Hohn, T., Richert-Pöggeler, K.R., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C.H., Teycheney, P.-Y., Iskra-Caruana, M.-L., Hull, R. (2008): Evolution of integrated plant viruses. In: *Plant Virus Evolution* (pp. 53-81). Springer, Berlin, Heidelberg.
22. Jones, R. A. C. (2016): Future scenarios for plant virus pathogens as climate change progresses. *Advances in virus research*, *95*, 87-147.
23. Kamitani, M., Nagano, A. J., Honjo, M. N., Kudoh, H. (2019): A survey on plant viruses in natural Brassicaceae communities using RNA-seq. *Microbial ecology*, 78(1), 113-121.
24. Karl, R., Kiefer, C., Ansell, S., Koch, MA. (2012): Systematics and evolution of arctic-alpine *Arabis alpina L*. (Brassicaceae) and its closest relatives in the eastern Mediterranean. *American Journal of Botany*, 99: 778–794.
25. Karl, R., Koch, MA. (2013): A world-wide perspective on crucifer speciation and evolution: phylogeny, biogeography and trait evolution in tribe Arabideae. *Annals of Botany*,112: 983–1001.
26. Koch, M.A., Karl, R. and German, D.A., (2017): Underexplored biodiversity of Eastern Mediterranean biota: systematics and evolutionary history of the genus *Aubrieta* (Brassicaceae). *Annals of botany*, *119*(1), pp.39-57.
27. Krajacic, M., Mamula, D., Horvath, J. (2001): Flower colour breaking in two *Aubrieta* cultivars caused by cucumber mosaic Cucumovirus (CMV) lacking satellite RNA-5. *Acta Phytopathologicaet Entomologica Hungarica Hungary*.
28. Koch, M., Bishop, J. and Mitchell-Olds, T. (1999): Molecular systematics and evolution of *Arabidopsis* and *Arabis*. *Plant Biology*,1(5), 529-537.
29. Lysak, M. A. and Koch, M. A. (2011): Phylogeny, genome, and karyotype evolution of crucifers (Brassicaceae), in Schmidt, R. and Bancroft, I., eds. *Genetics and Genomics of the Brassicaceae*, pp. 1-31. Berlin: Springer.
30. 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. *The EMBO Journal*, 22, 4836-45.
31. Richert-Pöggeler, K. R., Vijverberg, K., Alisawi, O., Chofong, G. N., Schwarzacher, T., Heslop-Harrison, J. S. (2021): Participation of multifunctional RNA in replication, recombination and regulation of endogenous plant pararetroviruses (EPRVs). *Frontiers in Plant Science*, *12*, 1148.
32. Parkin, I. A., Gulden, S. M., Sharpe, A. G., Lukens, L., Trick, M., Osborn, T. C. and Lydiate, D. J. (2005): Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics*,171(2), 765-781.
33. Muhammed, J.J. (2017): *Systematic and Genomic Studies in the Genus Aubrieta (Brassicaceae)* (Doctoral dissertation, University of Leicester).
34. Mustafa, S. I., Schwarzacher, T., & Heslop-Harrison, J. S. (2022): The nature and chromosomal landscape of endogenous retroviruses (ERVs) integrated in the sheep nuclear genome. *DNA*, *2*(1), 86-103.
35. Taberlet, P., Gielly, L., Pautou, G. and Bouvet, J. (1991): Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant molecular biology*,17(5), 1105- 1109.
36. Teo, C.H., Tan, S.H., Ho, C.L., Faridah, Q.Z., Othman, Y.R., Heslop-Harrison, J.S., Kalendar, R. Schulman, A.H. (2005): Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. *Journal of Plant Biology*, 48, 96-105.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 12-2022
DOI 10.17605/OSF.IO/9U6AN

37. Shukla, D. D., and Schmelzer, K. (1970): Studies on viruses and virus diseases of cruciferous plants. I. Viruses in some ornamentals. *Acta Phytopathologica Academiae Scientiarum Hungaricae*, 5(2-4), 137-144.

38. Warwick, SI., Mummenhoff, K., Sauder, C., Koch, MA., Al-Shehbaz, IA. (2010): Closing the gaps: Phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Systematics and Evolution*, 285: 209–232.

39. Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Zhang, Z. (2011): The genome of the mesopolyploid crop species *Brassica rapa*. *Nature genetics*, *43*(10), 1035-1039.

40. Yüzbaşıoğlu, S., Koch, M.A. and Al-Shehbaz, I.I., (2015): Proof of a knowledge database concept. *Aubieta ekimii* (Brassicaceae), a new species from NW Anatolia (Turkey): morphological and molecular support. *Plant Systematics and Evolution*, 301(8), 2043-2055

41. Farzadfar, S., Tomitaka, Y., Ikematsu, M., Golnaraghi, A. R., Pourrahim, R., Ohshima, K. (2009): Molecular characterisation of Turnip mosaic virus isolates from Brassicaceae weeds. *European journal of plant pathology*, *124*(1), 45.

42. Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S. Michalakis, Y. (2005): Recombination every day: abundant recombination in a virus during a single multicellular host infection. *PLoS Biology*, 3, 389-95.

43. Warwick, S. I. (2011): Brassicaceae in agriculture. In *Genetics and Genomics of the Brassicaceae* (pp. 33-65). Springer, New York, NY.

44. Saeidi, H., Rahiminejad, M. R., & Heslop-Harrison, J. S. (2008): Retroelement insertional polymorphisms, diversity and phylogeography within diploid, D-genome Aegilops tauschii (Triticeae, Poaceae) sub-taxa in Iran. *Annals of Botany*, *101*(6), 855-861.

45. Salina, E. A., Sergeeva, E. M., Adonina, I. G., Shcherban, A. B., Belcram, H., Huneau, C., Chalhoub, B. (2011): The impact of Ty3-gypsy group LTR retrotransposons Fatima on B-genome specificity of polyploid wheats. *BMC plant biology*, *11*(1), 1-14.

46. Schmidt, N., Seibt, K. M., Weber, B., Schwarzacher, T., Schmidt, T., Heitkam, T. (2021): Broken, silent, and in hiding: Tamed endogenous pararetroviruses escape elimination from the genome of sugar beet (*Beta vulgaris*). *Annals of botany*, *128*(3), 281-299.

47. Shafiq, M., Qurashi, F., Mushtaq, S., Hussain, M., Hameed, A., Haider, M. S. (2020): DNA plant viruses: biochemistry, replication, and molecular genetics. In *Applied Plant Virology* (pp. 169-182). Academic Press.

48. Staginnus, C., Gregor, W., Mette, M.F., Teo, C.H., Borroto-Fernandez, E.G., Machado, M.L.d.C., Matzke, M. Schwarzacher, T. (2007):Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. *BMC Plant Biology*, 7, 24-40.

49. Wessler, S. R. (2006): Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences*, *103*(47), 17600-17601.