IRANIAN JOURNAL OF BOTANY 30(1), 2024
DOI: 10.22092/ijb.2024.364281

COMPARATIVE ANALYSIS OF TANDEM REPEATS IN FOUR AUBRIETA (BRASSICACEAE) GENOMES

O. Alisawi, J. Muhammed, & P. Heslop-Harrison

Received 2023.12.02; accepted for publication 2024.04.08

Alisawi, O., Muhammed, J. & Heslop-Harrison, P. 2024.06.30: Comparative analysis of tandem repeats in four *Aubrieta* genomes. -*Iran. J. Bot. 30(1): 54-61*. Tehran.

The genus *Aubrieta* Adan. (Brassicaceae) is widely distributed and diverged across different elevations. The genome components and organization of this plant are still less understood. Tandemly repeated sequences were examined in whole genome raw reads of four *Aubrieta* species using Next Generation Sequencing (NGS) and bioinformatics techniques. Six clusters of tandem repeats were found based on RepeatExplorer and TAREAN pipelines; one cluster in *A. pinardii* (ApinSAT1L) and A. *scardica* (AscaSAT95 H), two clusters in *A. erubescens* (AeurSAT132L and AeurSAT230L), and A. *gracilis* (AgraSAT2L and AgraSAT15H) with the GenBank accession numbers (PP391544, PP391547, PP391548, PP391549, PP391545, PP391546) respectively, have been found within all examined genomes. The tandem repeated features were confirmed using de novo assembly contigs. Variable numbers of genome proportions and copies have been recorded for these elements. *Aubrieta erubescens* has a high copy number compared to *A. scardica* which has lower copies. The arrangement of tandem repeat clusters within the genome was tandemly organized except for *A. scardica* (AscaSAT95 H) which was dispersed. Therefore, these genomes can be explained in terms of composition, structure, and evolutionary relationships.

Osamah Nadhim Alisawi, Department of Plant Protection, Faculty of Agriculture, University of Kufa Najaf-Iraq- Jotyar Muhammed (correspondence<mizory@uod.ac>), University of Duhok College of Agricultural Engineering Sciences Kurdistan Region of Iraq Duhok.- Patrick Heslop-Harrison, Department of Genetics and Genome Biology, Institute for Environmental Futures, University of Leicester, Leicester, United Kingdom & Key Laboratory of Plant Resources Conservation and Sustainable Utilization/Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China.

Keywords: Brassicaceae; evolution; repeat sequences; next-generation sequencing (NGS); bioinformatics

تجزیه و تحلیل مقایسه‌ای توالی‌های تکرار شده در ژنوم چهار گونه از جنس Aubrieta (Brassicaceae)

اسامة ناظم العیساوی: گروه آموزشی حفاظت گیاهی، دانشکده کشاورزی، دانشگاه کوفه، نجف، عراق

جوتیار محمد: دانشکده علوم مهندسی کشاورزی، دانشگاه دهوک، منطقه کردستان عراق

پاتریک هسلوپ هاریسون: گروه آموزشی ژنتیک و بیولوژی ژنوم، موسسه آینده محیط زیست، دانشگاه لیستر، انگلستان و آزمایشگاه های حفاظت منابع گیاهی و مصرف پایدار، استان گوانگدونگ چین و باغ گیاه‌شناسی جنوب چین، آکادمی علوم گوانگژو، چین

گونه‌های جنس .Aubrieta Adan (Brassicaceae) به‌طور گسترده در ارتفاعات مختلف پراکنده و از یکدیگر جدا شده‌اند. اجزا و سازمان ژنوم این گیاهان هنوز کمتر شناخته شده است. توالی‌های تکرار شده پشت سر هم در قرائت‌های خام کل ژنوم چهار گونه Aubrieta با استفاده از روش‌های توالی‌یابی نسل بعدی (NGS) و بیوانفورماتیک مورد بررسی قرار گرفتند. با استفاده از نرم افزارهای RepeatExplorer و TAREAN شش دسته توالی‌های تکراری بدست آمدند. یک دسته در A. pinardii (ApinSAT1L) ویک دسته در A. scardica (AscaSAT95 H) و دو دسته در A. erubescens (AeurSAT132L and AeurSAT230L) و دو دسته در A. gracilis (AgraSAT2L and AgraSAT15H)، به ترتیب با

شماره ردیف‌های بانک ژن PP391548, PP391547, PP391544, PP391546, PP391545, PP391549 در همه ژنوم‌های بررسی شده مشاهده گردید. ویژگی‌های تکرارهای پشت سرهم توالی‌ها با استفاده از روش مونتاژ از نو بخش‌هایی از توالی‌های DNA که به‌گونه‌ای همپوشانی دارند، تأیید شدند. تعداد متغیری از نسبت‌ها و کپی‌های ژنوم برای این عناصر ثبت شده است. Aubrieta erubescens در مقایسه با A. scardica که کپی‌های کمتری دارد، تعداد کپی بالایی دارد. ترتیب خوشه‌های تکرار پشت سر هم در ژنوم به جز A. scardica (AscaSAT95 H) که پراکنده شده بود به‌صورت پشت سر هم سازماندهی شد. بنابراین، این ژنوم‌ها را می‌توان از نظر ترکیب، ساختار و روابط تکاملی توضیح داد. تعداد متغیری از نسبت‌ها و کپی‌های ژنوم برای این عناصر ثبت شده است. Aubrieta erubescens در مقایسه با A. scardica که کپی‌های کمتری دارد، تعداد کپی بالایی دارد. ترتیب خوشه‌های توالی‌های تکراری پشت سر هم در ژنوم به‌جز A. scardica (AscaSAT95 H) که پراکنده شده بود، به‌صورت پشت سر هم سازماندهی شد. بنابراین، این ژنوم‌ها را می‌توان از نظر ترکیب، ساختار و روابط تکاملی توضیح داد.

INTRODUCTION

In Angiosperms, the genome size varies from 60 to 150,000 Mb. This variation mainly resulted from the duplication of the whole genome (polyploidization) and the accumulation of repetitive elements (Pellicer & al. 2018). The repetitive DNA has a significant role in genome diversity and composition (Heslop-Harrison 2000). In the eukaryotic genome, there is a great amount of repetitive DNA sequences. These repeated sequences are present in the form of tandem repeats; satellite DNA, dispersed repeats (retroelements), rDNAs, and endogenous pararetrovirus. In addition, transposable elements (TEs) are another type of repeat sequences that shape major components in the genome and can affect their position, function, and evolution. They fall into two categories, class I, which is transposed by RNA (Retrotransposons), and class II, which is transposed by DNA (DNA transposons), (Bourque & al. 2018). Moreover, there is evidence that tandemly repeated sequences are frequently copied and arranged in different sizes of arrays within the genome (Kapustová & al. 2019). Generally, their classification is based on the sequence length as follows: 1) microsatellites; up to 5 base pairs (bp) repeats with array size 10–100 units, 2) minisatellites, usually up to 15 bp long monomer with an array size of 0.5–30 kb, 3) satellites up to several hundred bp in length with up to 100 Mb array size (Sharma and Raina 2005). In the mustard family, there is a distinctive feature which is polyploidy; nearly half of the family members have polyploidy features, which is the main reason for the family evolution (Lysak and Koch 2011). Additionally, Brassicaceae members went through a series of events as known as palaeopolyploidization; Whole Genome Duplication (WGDs) and mesopolyploidy states; whole-genome triplication (Haudry & al. 2013). In Brassicaceae, a quarter of the small-sized genome *Arabidopsis thaliana* (146 Mbp) represents repeat sequences (Li & al. 2004), whereas 64% of the *Hesperis sylvestris* (4264 Mbp) large genome is an accumulation of the repeated DNA sequences (Hlouskova & al. 2019). Brassicaceae (Cruciferae) is one of the major families in the plant kingdom, comprising 51 tribes, 325 genera, and 3740 species in the world including the mountain and sub (alpine) tribe Arabideae (Hohmann & al. 2015). Irano–Turanian region is regarded as the probable origin and center of diversity of Brassicaceae (Franzke & al. 2009; Koch & al. 2017). Boissier (1867) expected that the genus *Aubrieta* (Arabideae) could be easily distinguished from other related genera morphologically by its stamen filaments; winged and tooth-like appendage; and sessile with a cordate base leaf. Similarly, Mattfield in 1937 updated the taxonomic information of the genus mainly on morphological overview. Although there are reasonable morphological characteristics overlap. The fruit morphology (terete siliquae) and indumentum (simple or furcate hairs) are the reliable keys to identifying the genus (Cullen 1965; Jalas & al. 1994). There are 38 binomials for *Aubrieta* listed in IPNI (http://www.uk.ipni.org/; International Plant Names Index). This reflects a serious taxonomic difficulty in identifying the taxa in terms of systematics (Yüzbaşıoğlu & al. 2015; Dönmez & al. 2017 and 2023; Koch & al. 2017; Muhammed 2017).

These plants grow in different mountains ranging from 200-2900 m, a.b.s. and grow on rocks, crevices, and scree slopes, in shaded or fully sunny areas. *Aubrieta* species are mainly distributed from Anatolia through the Balkan Peninsula to Italy, and the Near East (Mattfeld 1939; Ancev and Goranova 2009). The genus *Aubrieta* is more likely to have originated in Turkey, due to the phylogenetic clade's expansion independently from Anatolia to the Levant, Greece, and the Balkan Peninsula. The first diversification period of *Aubrietia* was in Turkey 1.1 million years ago (MYA) toward hot and dry areas, and then one million years later diversified again in Greece to adapt to the mid and

low temperatures (Koch & al. 2017). Although sequencing technologies have been invented and developed, the cost is the main barrier to focusing on the complete genome. Recently, the price dropped, and the next-generation sequencing (NGS) applications were possible to work on. Consequently, the complete genomes including nuclear, mitochondria, and chloroplast genomes were sequenced (Vik and Repkova 2017). There are few works on *Aubrieta* genomes using the NGS applications; Complete chloroplast genome of *A. gracilis* (Muhammed 2017); *A. canescens* repeat elements (Mandáková & al. 2020); Genome organization of *A. canescens* subsp. *canescens* and *A. macrostyla* (Kaya & al. 2022); structure and composition of four *Aubrieta* endogenous pararetroviruses sequences (Alisawi & al. 2022). Interestingly, genome analysis in *Aubrieta* has not been studied except for those mentioned above, thus this work is conducted to figure out tandem repeats in the genomes of four *Aubrieta* species.

MATERIALS AND METHODS

Plant material

Fresh leaves of four *Aubrieta* species; *Aubrieta gracilis*, *A. scardica*, *A. erubescens,* and *A. pinardii* were collected from the wild and cultivated plants in the University of Leicester Botanic Garden as a national collection of *Aubrieta* (Table 1). The species were chosen in the framework of a project to study the genomes of species of the above-mentioned collection. The names of the taxa are checked and compared with the voucher specimens from different herbaria as well as using molecular data from the GenBank (NCBI).

DNA isolation from *Aubrieta* species

One gram of young leaves was wrapped in aluminum foil and immersed in liquid nitrogen for quick freezing, and then DNA was extracted directly as in the next step. The young leaves of the four Aubrieta species were subjected to DNA extraction using the cetyl-trimethylammonium bromide (CTAB) method (Doyle & Doyle 1990), with minor modifications (The incubation in the preheated buffer of CTAB was for 45 min, while the precipitated DNA was centrifuged at 735 g for 5 min). The extracted samples were sent to 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.

Repeat-Explorer and TAREAN programs

RepeatExplorer and TAREAN pipelines were applied to characterize and explore the repetitive DNA sequences in the data of next-generation sequencing (Novak & al. 2013). RepeatExplorer2 clustering was applied (Galaxy Version 2.3.8.1), and Viridplantae version 3.0 was selected for the taxon and protein domain database (Table 2). The generated clusters from the pipelines were extracted, and then the whole contigs of each cluster were submitted to the National Center for Biotechnology Information (Benson & al. 1990), and the Basic Local Alignment Search Tool (Atschul & al. 1990). Further, raw reads were mapped to identify tandem repeats using Geneious assembler at medium sensitivity/fast (Kearse & al. 2012) (http://www.geneious.com/) to get copy number and genome proportion as below: 1. Copy number: number of assembled reads x read length/reference sequence length; 2.Genome proportion: number of assembled reads/numbers of total NGS reads x 100 (Mustafa & al. 2018).

De novo assembly

De novo assembly of 20% of examined raw reads was achieved using the Geneious software assembler (Fig. 1). Before applying the assembler, the reads were paired by Geneious software. Then, 100 contigs with different lengths were constructed from the de novo assembly. Each contig was extracted and checked against every repeat sequence extracted from TAREAN using a dot plot tool to figure out the repeat organization and confirm its tandemly repeated arrangement within the genome.

Table 1. List of wild-collected, herbarium, and cultivated material of four *Aubrieta* species and their sources.

Country	Province	Locality	Taxon Names	Herbarium voucher	Collectors
Greece	Ioannina	Timfi	*A. gracilis*	2015-011	R.J. Gornall & J. Muhammed
Greece	Ioannina	Timfi	*A. scardica*	2015-010	R.J. Gornall & J. Muhammed
Greece	Chalkidiki	Athos	*A. erubescens*	1986-0165-01	ex hort. Copenhagen BG
Turkry	Isparta	Isparta	*A. pinardii*	1994-0153-01	ex hort. Trieste BG

RESULTS AND DISCUSSION

Identification of tandem repeats by graph-based clustering

Aubrieta raw read sequences for the four examined genomes have been analysed through a graph-based clustering pipeline of RepeatExplorer and Tandem Repeat Analyser (TAREAN) to find out the candidate tandem repeat clusters. In total, analysing sequence assembly data resulted in the identification of six candidates of tandemly repeated sequences (Table 3).

Each tandem repeat had a characteristic monomer length and genome proportion. RepeatExplorer has classified these repeats according to their abundance in each genome. Some clusters have not been discovered by the RepeatExplorer/TAREAN pipeline in other genomes and for that, we examined them against other species reads. Based on their abundance and host genome, we named them as one cluster in *A. pinardii* (ApinSAT1L), one in *A. scardica* (AscaSAT95H), two in *A. erubescens* (AeurSAT132L, AeurSAT230L) and two in *A. gracilis* (AgraSAT2L, AgraSAT15H), (Table 3). Other bioinformatics plots using selected approaches were also used to identify tandem-repeat motifs, including dot plots of *de novo* assembled reads to show tandem arrays as a series of lines parallel to the diagonal (Fig. 2). These further analyses confirmed the same families found by the pipelines. We are therefore confident that we exhausted the *Aubrieta* repeateaome and captured all tandem repeat types and families.

Genome organization by raw reads

Analysis of sequences from the *Aubrieta* raw reads by mapping reads to reference motifs for each tandem repeat type identified in the genomes showed each was abundant, although there were substantial differences in relative copy. Copies of all tandem repeat motifs were found in the four whole genome reads; assembled lengths of tandem repeat arrays are also shown. However, the genome size of the four species was unknown. In four *Aubrieta* genomes, *A. scardica* had a total repeat copy number of 30% less but with a higher genome proportion (about 3.94%) while *A. erubescens* had a higher number of repeats; among individual tandem repeat types, some repeats were increased in copy number while some were decreased. *Aubrieta pinardii* had a close number of repeats (relative copy numbers) to *A. gracilis* with the most abundant generated copies (absolute copy numbers) and lower genome proportion (2.96%) (Fig. 2).

Table 2. Numbers of input and analysed reads with the proportion of reads in top clusters of each *Aubrieta* genome in the RepeatExplorer pipeline.

***Aubrieta* species**	**The number of input reads**	**The number of analyzed reads**	**Proportion of reads in top clusters (%)**
A. pinardii	16121154	580596	47
A. gracilis	13441852	588006	39
A. erubescens	15454208	495310	48
A. scardica	8436374	418619	51

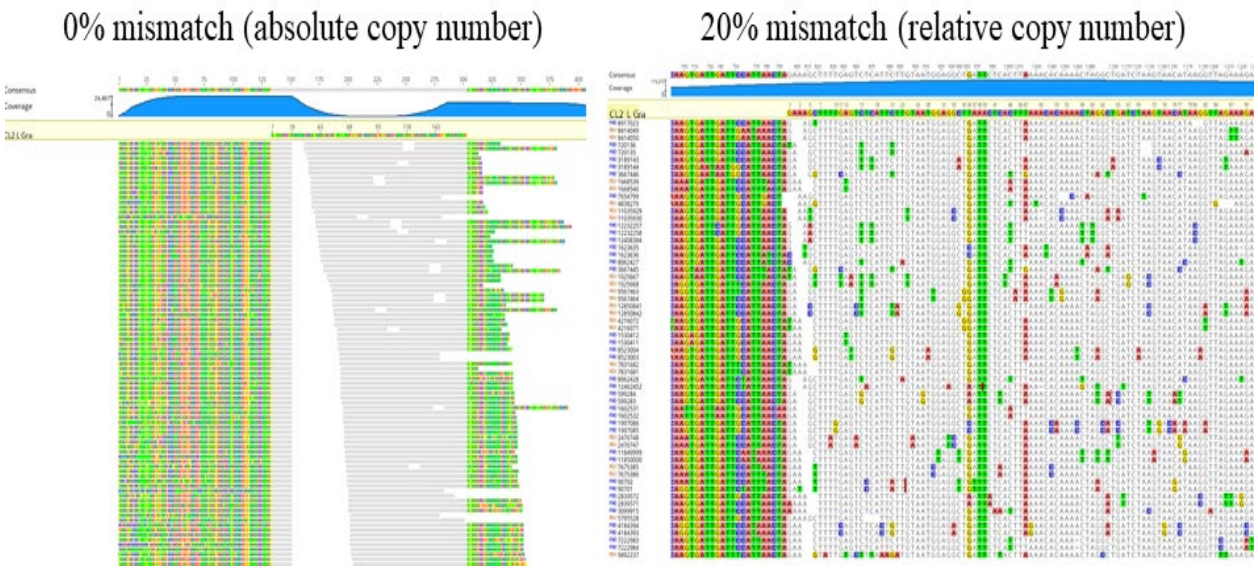

Fig. 1. Absolute and relative copy numbers of mapping read of tandem repeats (CL2LGra as an example) showed assembled reads with 0 and 20% mismatch.

The TAREAN and RepeatExplorer clustering graphs showed a wide range of shapes related to their abundance and genome proportions. ApinSAT1L, AgraSAT2L, and AgraSAT15H graphs were condensed star-like, while AeurSAT132Lshowed irregular shapes. AscaSAT95H and AeurSAT230L showed donut shapes. *De novo* assembly analysis confirmed tandem repeats arrangement within their genomes and most of them showed their tandemly arranged sequences except AscaSAT95 which showed dispersed organization (Fig. 2).

The detailed analysis of repetitive elements from genomes of four *Aubrieta* species used complementary tools (graph-based repeat clustering using unassembled raw reads, raw read mapping, and sequence assemblies). Repetitive elements represented 62-64% of the *Aubrieta* genomes; our analysis showed that 2.96-3.94% of the genome was represented by a comparatively high number of satellite DNA clusters (Table 2). This result is close to the genome percentage of tandem repeats in other species; 2.01% (372 Mb) in *Arabis alpina*, 2.85 % (392 Mb) in *Aubrieta canescens*, 4.36% (313 Mb) in *Draba nemorosa*, 5.81% (382 Mb) in *Arabis cypria,* (Mandáková & al. 2020).

In partially de novo assembled sequences, we found different patterns of tandemly arranged satellite repeats to confirm their repetitive features. We have not seen complete or large portions of arrays due to the partial assembly that we achieved in this work. However, we would expect to see repeat arrays (kbs to Mbs) in case of full data assembly. In whole-genome sequence assembly, the graph-based clustering algorithms were widely applied (Zerbino & al. 2008; Kingsford & al. 2010; Novak & al. 2010).

Here, the RepeatExplorer and TAREAN graphs are used to cluster repetitive elements. The tandem repeats, generally show a graph with a dense center (star) or ring (donut) graph shape which represents multiple similar overlaps reads (Fig. 2). Similarly, the lower number of peripheral variants reads, represents the ends of arrays.

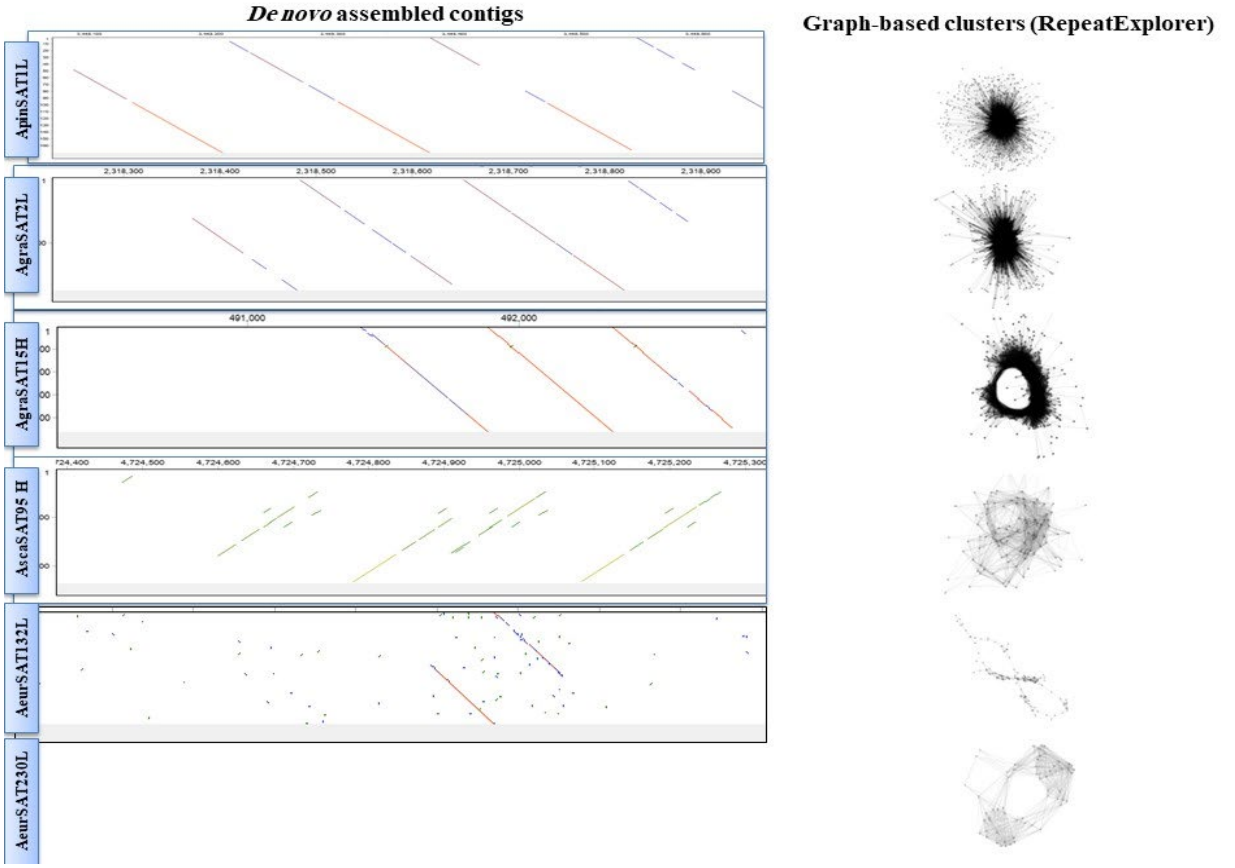

Fig. 2. Dot plots of *de novo* assembled contigs on the left show tandemly arranged clusters of six examples of high-confidence satellites (the first two), and low-confidence satellites (the last four), to confirm the tandem repeat features except the latter one (AeurSAT230L) as we could not find its feature in the *de novo* assembly. The graph-based clusters next to each particular cluster on the right, produced from the TAREAN pipeline.

Table 3. Percentage of mapped tandem repeat sequences to the four examined raw read sequences with 0% mismatch (absolute copy number) over 20% mismatch (relative copy number) and GenBank accession numbers; G.P=Genome proportion

Types	Gen Bank AN	Length(bp)	*A. pinardii*		*A. erubescens*		*A. gracillis*		*A. scardica*	
			Mismatch 0/20%	**G.P**	**Mismatch 0/20%**	**G.P**	**Mismatch 0/20%**	**G.P**	**Mismatch 0/ 20%**	**G.P**
ApinSAT1L	PP391544	170	38 763/ 192 265	1.19	18 698/ 244 113	1.57	18 918/ 158 805	1.18	14 718/ 120 193	1.42
AgraSAT2L	PP391545	170	36 430/ 209 452	1.29	40 896/ 192 348	1.24	21 506/ 179 537	1.33	16 189/ 109 275	1.29
AgraSAT15H	PP391546	48	1 606/ 74 725	0.46	2 072/ 104 962	0.67	42/108 654	0.80	1 711/ 100 856	1.19
AscaSAT95 H	PP391547	250	0/1 435	0.008	1/71	0.00	2/4 944	0.03	1 258/3 947	0.04
AeurSAT132L	PP391548	207	72/4 263	0.02	299/3 682	0.02	40/2 302	0.017	17/170	0.002
AeurSAT230L	PP391549	319	6/11	0.00	1 019/1 822	0.011	6/50	0.00	1/19	0.00
Total			76 877/ 482 151	2.96	62 985/ 546 998	3.51	21 158/ 454 292	3.35	33 894/ 334 460	3.94

However, the graphs were helpful for the identification of tandem repeat candidates, there was not any link showing the relationships between biological features of the repeat (abundance, location, monomer length, array size, or inter-monomer variation) and the graph shape. The tandem repeat portion is comparatively in balance based on other species ratios and has similar numbers to most studied species. Mainly, due to the potential roles in the maintenance of nuclear composition of chromosome structure at metaphase and interphase including centromeric function (Hemleben & al. 2007), roles in large-scale genome organization (Biscotti & al. 2015), chromatin packaging, the maintenance of chromosome stability (Vershinin & Heslop-Harrison 1998), and the environmental factors that effect on the gene expression (Lamb & al. 2007). *Aubrieta* species separated from a common ancestor 2.5 MYA, and the last diversification among them was ~ 0.19 MYA in Greece (Koch & al. 2017). Tandem repeats most likely evolve in copy number by unequal crossing-over at meiosis, or slippage during replication before mitosis, events that our data show have occurred in both recent and older evolutionary times with both amplification and loss. The abundance of at least 12 clusters suggests that these types are ancient. According to the nature of the tandem repeats and availability of straightforward analysis pathways, it will be valuable to resynthesize these species from their parental species, and analyse the copy number changes occurring over a few generations; comparing both sexual and vegetative pathways to distinguish meiotic and mitotic events, examining if these changes are salutatory occurred in bursts soon after hybridization as with transposon activation and potentially related through chromatin remodelling and methylation changes (Grandbastien 1998). Genomic changes have been seen regularly in hybrid species (Alix & al. 2017). Furthermore, resynthesized tetraploid Brassicas species, homoeologous chromosome exchanges, and some specific gene determinations were found (Gaeta & al. 2007; Klima & al. 2019). Thus, it would be valuable to examine the repetitive DNA changes over similar timescales in *Aubrieta*. Similarly, changes in new triticale (wheat x rye) hybrids, and more genomic changes in the rye chromosomes, occurring immediately after hybridization (Ma and Gustafson 2008), major rearrangements of the genome are also seen in hybrids in the grass genus *Brachypodium* (Lusinska & al. 2018). *Aubrieta pinardii* showed that ApinSAT1L was younger than others while *A. erubescens* was the most abundant copies in the amplifying stage. There were very few copies detected at the origin level in AeurSAT230L as well. The present research will consequently use the chromosomal and genetic maps that are planned to be achieved by future genome works through molecular mapping and in situ hybridization to explain how these genomes are differentiated and geographically distributed.

REFERENCES

Alisawi, O., Muhammed, J. & Heslop-Harrison, J. 2022: Endogenous pararetroviruses sequences can be used as markers to differentiate four *Aubrieta* species. -Jilin Daxue Xuebao (Gongxueban) 41: 176-184.

Alix, K., Gérard, P., Schwarzacher, T. & Heslop-Harrison, J. 2017: Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. -Ann. Bot. 120: 183-194.

Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. 1990: Basic local alignment search tool. -J. Mol. Biol. 215: 403-410.

Ančev, M. & Goranova, V. 2009: *Aubrieta* (Brassicaceae) in the Bulgarian flora. -Phytologia Balcanica 15: 43-50.

Benson, D., Boguski, M., Lipman, D.J. & Ostell, J. 1990: The national center for biotechnology information. -Genomics 6: 389-391.

Biscotti, M., Olmo, E. & Heslop-Harrison, J. 2015: Repetitive DNA in eukaryotic genomes. -Chromosome Res. 23: 415-420.

Boissier, E. 1867: Flora Orientalis. 1: 249-254. -Basel: H. Georg.

Bourque, G., Burns, K., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M. Izsvák, Z., Levin, H., Macfarlan, T. & Mager, D. 2018: Ten things you should know about transposable elements. -Genome biol. 19: 1-12.

Cullen, J. 1965: *Aubrieta* Adans., in Davis, P.H., ed. Flora of Turkey and the East Aegean Islands, 1: 444-447. -Edinburgh: Edinburgh University Press.

Dönmez, A., Aydin, Z., Kaya, Y. & Yüzbaşioğlu, İ. 2023: *Aubrieta birolmutlui* (Brassicaceae), a new species from Eastern Turkey with molecular phylogenetic support. -Phytotaxa 579: 278-288.

Dönmez, A., Aydin, Z. & Koch, M. 2017: *Aubrieta alshehbazii* (Brassicaceae), a new species from Central Turkey. -Phytotaxa 299: 103-110.

Doyle, J. 1990: Isolation of DNA from small amounts of plant tissues. -BRL Focus 12: 13-15.

Franzke, A., German, D., Al-Shehbaz, I. & Mummenhoff, K. 2009: *Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae. -Taxon 58: 425-437.

Gaeta, R., Pires, J., Iniguez-Luy, F., Leon, E. & Osborn, T. 2007: Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. -Plant Cell 19: 3403-3417.

Grandbastien, M. A. 1998: Activation of plant retrotransposons under stress conditions Trends. -Plant Science 3: 181-187.

Haudry, A., Platts, A., Vello, E., Hoen, D., Leclercq, M., Williamson, R., Forczek, E., Joly-Lopez, Z., Steffen, J., Hazzouri, K. & Dewar, K. 2013: An atlas of over 90.000 conserved noncoding sequences provides insight into crucifer regulatory regions. -Nat. Genet. 45: 891-898.

Hemleben, V., Kovarik, A., Torres-Ruiz, R., Volkov, R. & Beridze, T. 2007: Plant highly repeated satellite DNA: molecular evolution. distribution and use for the identification of hybrids. -Syst. Biodivers. 277-289.

Heslop-Harrison, J. 2000: Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. -Plant Cell 12: 617-635.

Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P. & Lysak, M. 2019: The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. -Ann. Bot. 124: 103-120.

Hohmann, N., Wolf, E., Lysak, M. & Koch, M. 2015: A time-calibrated road map of Brassicaceae species radiation and evolutionary history. -Plant Cell 27: 2770-2784.

Jalas, J., Suominen, J. & Lampinen, R. 1994: Atlas Florae Europaeae 10 Cruciferae (*Sisymbrium* to *Aubrieta*). -Committee for Mapping the Flora of Europe, Helsinki.

Kapustová, V., Tulpová, Z., Toegelová, H., Novák, P., Macas, J., Karafiátová, M., Hřibová, E., Doležel, J. & Šimková, H. 2019: The dark matter of large cereal genomes: long tandem repeats. -Int. J. Mol. Sci. 20: 2483.

Kaya, Y., Aydın, Z., Cai, X., Wang, X. & Dönmez, A. 2022: Genome-wide characterization of two *Aubrieta* taxa: *Aubrieta canescens* subsp. *canescens* and *Au. macrostyla* (Brassicaceae). -AoB Plants 14: 35.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton S., Cooper, A., Markowitz, S. & Duran, C. 2012: Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. -Bioinformatics 28: 1647-1649.

Kingsford, C., Schatz, M. & Pop, M. 2010: Assembly complexity of prokaryotic genomes using short reads. -BMC Bioinform. 11: 21.

Klíma, M., Jozová, E., Jelínková, I., Kučera, V., Hu, S. & Čurn, V. 2019: Early in vitro selection of winter oilseed rape (*Brassica napus* L.) plants with the fertility restorer gene for CMS Shaan 2A via non-destructive molecular analysis of microspore-derived embryos. -Czech J. Genet. Plant Breed. 55: 162-165.

Koch, M., Karl, R. & German, D. 2017: Underexplored biodiversity of Eastern Mediterranean biota: systematics and evolutionary history of the genus *Aubrieta* (Brassicaceae). -Ann. Bot. 119: 39-57.

Lamb, J., Yu, W., Han, F. & Birchler, J. 2007: Plant chromosomes from end to end: telomeres, heterochromatin and centromeres. -Curr. Opin. Plant Biol. 10: 116-122.

Li, Y., Korol, A., Fahima T. & Nevo, E. 2004: Microsatellites within genes: structure, function,

and evolution. -Mol. Biol. Evol. 21: 991-1007.

Lusinska, J., Majka, J., Betekhtin, A., Susek, K., Wolny, E. & Hasterok, R. 2018: Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon B. stacei* and *B. hybridum*. -Ann. Bot. 122: 445-459.

Lysak, M. & Koch, M. 2011: Phylogeny, genome, and karyotype evolution of crucifers (Brassicaceae). -Genetics and Genomics of the Brassicaceae 1-31. Springer, New York, NY.

Ma, X. & Gustafson, J. 2008: Allopolyploidization-accommodated genomic sequence changes in triticale. -Ann. Bot. 101: 825-832.

Mandáková, T., Hloušková, P., Koch, M. & Lysak, M. 2020: Genome evolution in Arabideae was marked by frequent centromere repositioning. -Plant Cell 32: 650-665.

Mattfeld, J. 1939: The species of the genus *Aubrieta* Adanson. -Quart. Bull. Alp. Gard. Soc. 7: 157-181.

Muhammed, J. 2017: Systematic and Genomic Studies in the Genus *Aubrieta* (Brassicaceae), (Doctoral dissertation. University of Leicester).

Mustafa, S., Schwarzacher, T. & Heslop-Harrison, J. 2018: Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors Mitochondrial DNA. -Part A 1-14.

Novak, P., Neumann, P. & Macas, J. 2010: Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. - BMC Bioinform. 11: 378.

Novak, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. 2013: RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. -Bioinformatics 29: 792-793.

Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. 2018: Genome size diversity and its impact on the evolution of land plants. -Genes 9: 88.

Sharma S. & Raina, S. 2005: Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes Cytogenet. -Genome Res. 109: 15-26.

Vershinin, A. & Heslop-Harrison, J. 1998: Comparative analysis of the nucleosomal structure of rye, wheat and their relatives Plant. -Mol. Biol. 36: 149-161.

Vlk, D. & Řepková, J. 2017: Application of next-generation sequencing in plant breeding. -Czech J. Genet. Plant Breed. 53: 89-96.

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

Zerbino, D. & Birney, E. 2008: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. -Genome Res.18: 821-829.