UNIVERSITÀ DEGLI STUDI DI TORINO

SCUOLA DI DOTTORATO

DOTTORATO IN
SCIENZE AGRARIE, FORESTALI E ALIMENTARI

GENETIC MARKERS IDENTIFICATION FOR MAP DEVELOPMENT AND QTL ANALYSIS IN ORNAMENTALS

Matteo Martina

Docente Guida:
Prof. Ezio Portis

Coordinatore del Ciclo:
Prof. Domenico Bosco

Co-Relatore:
Prof. Sergio Lanteri

YEARS
2020; 2021; 2022; 2023

Contents

Chapter's references:

- ***Chapter I***: Martina, Matteo, et al. "Genome-wide survey and development of the first microsatellite markers database (*AnCorDb*) in *Anemone coronaria* L." *International Journal of Molecular Sciences* 23.6 (2022): 3126.
- ***Chapter II***: Martina, Matteo, et al. "Microsatellite-based identification of double-haploid plants by androgenesis in *Anemone coronaria* L." *Acta Horticulturae* (2023) – *In press.*
- ***Chapter III***: Martina, Matteo, et al. "First genetic maps development and QTL mining in *Ranunculus asiaticus* L. through ddRADseq." *Frontiers in Plant Science* 13 (2022): 1009206.
- ***Chapter IV***: Martina, Matteo, et al. "Diversity analyses in two ornamental and large-genome Ranunculaceae species based on a low-cost Klenow NGS-based protocol." *Frontiers in Plant Science* 14 (2023): 1187205.
- ***Chapter V***: Martina, Matteo, et al. "The first reference genome of *Ranunculus asiaticus* L. reveals a key region related to anthocyanin pigmentation" *Acta Horticulturae* (2023) – *In press.*

General introduction

Ranunculus asiaticus L. and *Anemone coronaria* L. are flowering plants belonging to the Ranunculaceae family. These species are highly valued for their ornamental flowers, which come in a wide range of colours and add beauty and vibrancy to gardens and floral arrangements (Figure 1).

Figure 1. Botanical illustration of *A. coronaria* (left) and *R. asiaticus* (right).

In recent years, there has been an increasing demand for the development of molecular tools to support plant breeding efforts in these ornamentals. Molecular-related techniques, such as DNA markers, genetic mapping, and QTL (*quantitative trait loci*) analysis, can help accelerate the breeding process by enabling the selection of superior varieties, also allowing the

understanding of genetic diversity and population structure within the cultivated species. By increasing the genetic and genomic knowledge on *R. asiaticus* and *A. coronaria*, this research can contribute to the development of new and improved varieties with enhanced traits, such as flower color, size, and disease resistance, ensuring their continued popularity and economic significance in the ornamental industry.

Poppy anemone (*Anemone coronaria* L.)

Poppy anemone has been targeted by breeders, who developed several varieties, cultivated both as garden plants and for cut flower production (Figure 2; [1,2]).

Figure 2. A set of commercial lines, marketed by Biancheri Creazioni. A) Cut varieties; B) Pot varieties.

The genus *Anemone* originated and diversified in the temperate zones of the Southern and Northern hemispheres, with the Mediterranean basin being a significant center of origin for many species that have contributed to the development of cultivated varieties, particularly *A. coronaria*, *A. hortensis*, and *A. pavonina*. Among these, *A. coronaria* stands out as the most extensively grown for cut flower production, with popular varieties including 'Wicabri', 'Tetranemone', 'Jerusalem', 'Cristina', 'Mona Lisa', and 'Mistral'. These varieties are cultivated across Western and Southern Europe, the United States, and Israel (Figure 3; [3]).

Figure 3. Most popular varieties of *A. coronaria* market as cut flower.

Poppy anemone is an herbaceous perennial diploid plant ($2n = 2x = 16$), although some commercial varieties are tetraploid. Plant height ranges from 20 to 40 cm, occasionally reaching up to 60 cm. The plant spreads to about 15-23 cm, forming a basal rosette of a few leaves. The leaves are composed of three leaflets, each deeply lobed. The flowers, which bloom from April to June, are solitary and borne on tall stems with a whorl of small leaves just below the flower (Figure 4; [4]).

Figure 4. *A. coronaria* wildly growing in the northern regions of Italy.

The diameter of the flower ranges from 3 to 8 cm, and the showy petal-like tepals come in shades of red (although white or blue varieties are also found), with a distinctive black center. The dry pollen grains, measuring less than 40 nm in diameter, have an unsculpted exine and are typically deposited within a short distance from their source. The flower structure consists of tightly packed pistils at the center, surrounded by a crown-like ring of stamens, which gives the species its specific epithet "coronaria." The flowers produce an average of 200 to 300 seeds. The plant forms hard black tubers as storage organs [5].

Anemone's breeding efforts

The current status of *A. coronaria* breeding involves careful consideration of its distinct characteristics, such as allogamy, protogyny, inbreeding

depression, a slow generation turnover, sufficient seed production, technical challenges associated with vegetative cloning for phytosanitary purposes, and limited genetic variation for economically relevant traits like disease resistance and specific colors [6]. Additionally, the relatively low economic importance of the species has hindered breeding efforts. Existing cultivars of *A. coronaria* can be grouped into sub-cultivars based on flower color, with additional genetic variability observed within each cluster in terms of flower yield, precocity, and flower quality. This non-uniformity reflects the dual objectives of breeding, which aim to maintain high vigor while providing a uniform end product. Sub-cultivars are created through intentional hand crosses between distantly related families, and these families are maintained through sibling crosses. As a result, individuals within each family are typically heterogeneous and partially heterozygous, leading to non-uniform progeny. Commercial seed lots for each sub-cultivar are generated by combining several crosses. Consequently, sub-cultivars represent plant populations whose genetic composition can vary over time due to the assembly of seed lots from diverse sources [7]. In recent years, several techniques have been developed to support *A. coronaria* conventional breeding. These include *in vitro* propagation of adult plants and seedlings, generation of androgenetic plants, *in vitro* regeneration from somatic tissue, genetic characterization of populations and cultivars using molecular markers, and the discovery of single nucleotide polymorphisms (SNPs) in expressed sequences [7–9]. Researchers have successfully established *in vitro* clones of selected plants from floral stalk portions, overcoming contamination issues associated with tuber-derived explants. By culturing on a medium containing zeatin, the ability to generate bud primordia and *de novo*

meristematic patches capable of developing into adventitious buds has been demonstrated. *In vitro*-formed shoots have been micropropagated and acclimatized, resulting in uniform plants with no observable somaclonal variation. Clones have also been derived from seed explants. Anther cultures have been successfully established in *Anemone* spp., with subsequent regeneration of somatic embryos and plantlets from anthers of elite cultivars [8].

Genetic characterization studies have been conducted using molecular markers such as AFLP (amplified fragment length polymorphism - [7]), with modifications to standard protocols due to *A. coronaria*'s large genome size (9,08-11,93Gb [10,11]) These studies have revealed wide genetic variation within wild populations and lower genetic variation in specific cultivars. The analysis of sub-cultivars within different cultivars has demonstrated high differentiation within sub-cultivars and comparable levels of differentiation among sub-cultivars and cultivars. These advancements in breeding techniques and genetic characterization offer valuable tools for improving the traits of poppy anemone, ultimately enhancing its commercial value as a cut flower and garden plant [12]. Through the integration of conventional and molecular breeding approaches, it becomes possible to achieve increased product uniformity, disease resistance, specific colours, and other economically important traits. By leveraging these tools, breeders can further expand the genetic diversity of *A. coronaria* and meet the evolving demands of the market [8].

Persian buttercup (*Ranunculus asiaticus* L.)

The *Ranunculus* genus comprises nearly 600 botanical species that span both hemispheres [13]. These species are commonly employed in traditional medicinal practices and serve as decorative flora. Persian buttercup (*Ranunculus asiaticus* L.; *2x=2n*=16; expected genome size: 7.6Gb; [14]) is indigenous to the Eastern Mediterranean region and was introduced to Western Europe during the sixteenth century. It gained significant popularity as a garden plant in Europe during the eighteenth century, leading to extensive breeding efforts in nineteenth-century England, resulting in the development of over 500 cultivars (Figure 5).

Figure 5. Selection of the market varieties developed by Biancheri Creazioni.

Subsequently, *R. asiaticus* fell into obscurity; however, in the past three to four decades, it has experienced a resurgence in popularity as a cut flower and potted plant, thanks to the endeavours to develop seed-propagated

lines [13]. This species is a perennial geophyte undergoing a biological cycle characterized by vegetative growth during cool, moist winter seasons and a quiescence phase in hot, dry summer months. Following pollination, dry, brown, cylindrical-shaped seeds are released. Under cool and moist summer climates, sowing seeds in May enables early November flowering. However, this approach is suitable only under Mediterranean conditions with artificially cooled soil at 18 °C. In hot and dry summer conditions, sowing is typically postponed until July or August, or preferably early autumn when the embryos have naturally overcome dormancy and reached maturity [15]. Rapid growth of plants commences approximately two months after sowing, leading to flowering in spring. The commercial production of *Ranunculus* relies on the categorization of tuberous roots into small (2–3 cm) and large (5–7 cm) sizes, each corresponding to different flowering periods. Large tuberous roots yield commercial flower stalks from mid-October, while small-sized roots contribute to flowering in January to February. This practice ensures an extended production season through the utilization of diverse plant material.

R. asiaticus is characterized by its branched stems, reaching heights of 30-45 cm, and adorned with vibrant, large multi-petaled flowers displaying a wide spectrum of colours including white, yellow, pink, gold, orange, red, and mixed hues (Figure 6; [13]).

Figure 6. Greenhouse variability of *R. asiaticus* at different flowering stages.

These inflorescences emerge approximately five months after planting. The flowers exhibit actinomorphic symmetry, featuring ovate sepals and obovate petals, and the number of stamens is highly variable, ranging from well-developed to absent [16]. Interestingly, intermediate floral organs and remnants of anthers on petals have been observed in this plant family, indicating a capacity for conversion between petals and stamens [17]. Pollen grains, which are spheroid and poly-porate, are released from the anthers through a longitudinal slit [18]. The well-developed receptacle bears numerous uniovular carpels that expand during the flowering period and seed formation. The stigma is unicellular, papillated, and dry. The resulting fruitlets, known as achenes, possess wings. When these mature fruitlets are shed, the embryos of *R. asiaticus* remain arrested at the torpedo stage [13].

Buttercup's breeding efforts

The focus of commercial breeding programs for *R. asiaticus*, especially when cultivated as a cut flower, lies primarily in enhancing flower characteristics such as diameter, morphology, and colour. A notable recent addition to the market is the "PON PON" line, developed through an extensive breeding program and micropropagation by Biancheri Creations (Sanremo, Italy) [19]. This line stands out due to its bicolor and jagged petals, offering a unique shape and a diverse range of flower colors. While *Ranunculus* already exhibits a wide range of flower colors, the elusive blue color remains a desirable trait yet to be achieved. Moreover, interspecific crosses, such as *R. asiaticus* x *R. cortusifolius*, have shown promise in improving certain agronomic characteristics, such as larger leaves and the reintroduction of perfumed petals, as well as branched stems with multiple flowers, offering a fresh perspective on the familiar *Ranunculus* aesthetic [13]. Also, in the development of pot varieties achieving uniformity in terms of color and flowering time holds significant value. Notably, a current market trend involves the introduction of compact lines that eliminate the necessity for growth regulators. This approach not only promotes sustainable management practices but also resolves issues arising from the diminishing availability of approved growth regulators in cultivation [16].

Breeders worldwide have made significant efforts to select seed lines capable of achieving commercial production in the first year. This involves sowing from September to October and harvesting flowers in March to April, with tuberous roots discarded at the end of the cultural cycle. Seed propagation allows for healthy propagation since the major virus and fungal diseases affecting *Ranunculus* are not transmitted through seeds.

However, this approach prevents winter production, which is more profitable than spring production [13]. Furthermore, breeding programs for seed propagation are complex, requiring approximately 4-5 years to develop a new line, and population homogeneity is not always guaranteed. Vegetative propagation involves dividing tuberous roots, but the annual multiplication rate is low (2-5), and diseases caused by viruses and fungi pose significant challenges. In vitro propagation provides a solution to these problems and offers the opportunity to stabilize new varieties obtained through sexual reproduction, especially for ornamental characteristics that may be unstable through seed propagation [20]. Biancheri Creations, the largest company producing in vitro *Ranunculus* lines, has observed an increasing impact of micropropagated plant material on the total number of tuberous roots sold. Micropropagation has been shown to produce more vigorous tuberous roots with larger and higher-quality flower buds, as well as better uniformity compared to conventionally obtained tuberous roots [16]. Axillary bud stimulation has been identified as the optimal method for in vitro propagation of *Ranuculus*, providing true-to-type plant material. However, care should be taken to avoid prolonged in vitro culture, as it can lead to reduced flowering percentage and abnormalities in plant vigor, stems, leaves, and flowers. Micropropagated plant material is typically used to produce tuberous roots at the end of the in vivo culture cycle [21]. Successful acclimatization can be achieved under Mediterranean conditions by covering ex vitro plantlets with nonwoven sheets, although caution is advised during warm autumn months. Gradual removal of the tunnel and controlled wetting are necessary during acclimatization, and the use of arbuscular mycorrhizal fungi during the acclimatization stage has shown

varied effects on *Ranunculus* performance when transferred to the field, providing new insights into their role [22].

Molecular Markers in NGS Era

Marker-assisted selection (MAS) represents an effective approach that overcomes certain limitations of conventional breeding methods, enhancing the selection of desirable phenotypes and genes. Classical molecular markers, such as random fragment length polymorphisms (RFLPs), amplified fragment length polymorphism (AFLPs), simple sequence repeats (SSRs), random amplification of polymorphic DNA (RAPD), and others, provide valuable tools for revealing polymorphisms in nucleotide sequences and identifying genetic variation. This approach is not influenced by environmental conditions and can be applied at different stages of plant growth [23]. With the availability of diverse molecular markers and genetic maps, MAS has become feasible for both major crops, contributing to the advancement of breeding strategies. The successful utilization of MAS in breeding programs relies on several factors, including the tight association of molecular markers with quantitative trait loci (QTLs) or major genes of agronomic interest, sufficient recombination between markers and the rest of the genome, and the development of efficient methods for analysing many individuals in a time- and cost-effective manner [24–26].

The advent of next-generation sequencing (NGS) technologies has transformed the field of plant breeding, offering powerful tools for marker discovery and genotyping. These technologies have revolutionized the identification of DNA markers, moving from fragment-based polymorphism identification to sequence-based SNP identification

[27,28]. The development of NGS technologies, coupled with powerful computational pipelines, has dramatically reduced the cost of whole genome sequencing, enabling the discovery, sequencing, and genotyping of thousands of markers in a single step. NGS has emerged as a game-changing tool for next-generation plant breeding, allowing the rapid detection of DNA sequence polymorphisms within a short timeframe [29,30].

Several targeted marker discovery techniques have been devised using NGS platforms, even in the absence of prior whole genome sequence knowledge. These techniques involve the construction of partial genome representation libraries, either through complexity reduction using restriction enzymes or sequence capture without involving restriction digestion. Examples of complexity reduction techniques include reduced-representation libraries, restriction-site associated DNA sequencing (RAD-seq; [31]), genotyping-by-sequencing (GBS;[32]), SLAF-seq [33], and K-seq [34]. These approaches enable the discovery of SNPs and simultaneous genotyping, providing a faster and more cost-effective marker development process. RADseq, short for Restriction-site Associated DNA sequencing, is a powerful molecular biology technique that employs restriction enzymes to selectively target and sequence specific regions of an organism's genome. By cleaving DNA at specific recognition sites, RADseq generates fragments for sequencing, enabling the discovery of single nucleotide polymorphisms (SNPs) and other genetic markers across a wide range of organisms, thus facilitating insights into evolutionary processes and genetic diversity. On the other hand, K-seq, short for Klenow-based Sequencing, is a genotyping methodology that involves two rounds of Klenow amplification with short oligonucleotides,

followed by standard PCR and Illumina sequencing. K-seq has been successfully applied in various species, such as tomato, dog, and wheat, providing genetic distance estimates comparable to whole genome sequencing (WGS) in dogs. In tomato, K-seq demonstrated similar genetic results to genotyping-by-sequencing (GBS) but with the advantage of detecting more single nucleotide polymorphisms (SNPs) from the same quantity of Illumina sequencing reads. The technology exhibits high reproducibility, and it has been proven to be effective in polyploid species, generating specific markers for subgenomes and accurately locating SNPs from diploid ancestors within their expected subgenomes. Notably, K-seq is an accessible, cost-effective, and reliable technology without patent restrictions, making it readily accessible for molecular research applications. The integration of NGS technologies is not without challenges. While whole genome sequencing offers comprehensive information, targeted sequencing approaches like RAD-seq and GBS are more cost-effective for large-scale marker discovery, particularly in genomes that are large and not fully decoded. The cost of genotyping itself may decrease rapidly with technological advancements, but the associated costs of library preparation and bioinformatic analysis may not decrease at the same pace [23]. Therefore, efficient data mining and extraction of usable information are crucial for the successful application of NGS technologies in next-generation plant breeding.

Beyond marker discovery, NGS technologies have additional applications in targeted re-sequencing for identifying domestication-related genes and conducting genome-wide selection studies to predict breeding values of traits. These applications hold great potential for becoming essential tools for next-generation plant breeders in the development of superior cultivars.

The ability to directly explore genome sequences has revolutionized the science of plant breeding in recent years, positioning NGS as a valuable weapon in the breeder's toolkit.

Third Generation Sequencing

Following the emergence of next-generation sequencing (NGS) technologies, a new generation of sequencing methods, known as third-generation sequencing (TGS) technologies, was introduced. TGS approaches enable the sequencing of nucleotide molecules (DNA or RNA) without the need for PCR amplification and provide real-time analysis of the data [35]. In 2011, Pacific Biosciences introduced the PacBio RS sequencer, which marked the advent of single-molecule real-time (SMRT) sequencing technology. Although the initial version had relatively short read lengths (~1.5 kb) and high error rates (~13%), PacBio continuously improved the technology, leading to the development of the Sequel System [36]. Over time, further advancements resulted in the release of the Sequel II System, the Sequel IIe System, and more recently the Revio.

In addition to PacBio's technology, another TGS approach emerged in 2014 with the introduction of nanopore sequencing by Oxford Nanopore Technologies (ONT). Nanopore sequencing utilizes nanopores in a membrane to sequence single-stranded DNA or RNA molecules, a concept proposed in the late 1980s. The original idea suggested that if oligomers were driven through a protein nanopore channel by an electrical current, the current would be disrupted in a manner specific to the base composition of the oligomers. Through the rapid evolution of ONT chemistries, significant improvements in throughput have been achieved, establishing nanopore sequencing as a valuable tool in various research fields. Such

technology offers several distinct advantages for genome sequencing and analysis: one notable benefit is its capacity to generate ultra-long reads of up to 4 megabases (Mb), which is particularly advantageous for assembling large and highly repetitive plant genomes [35]. This capability enables the comprehensive characterization of complex genomic regions that are challenging to resolve using other sequencing platforms. Moreover, ONT facilitates the sequencing of full-length RNA transcripts, allowing for accurate quantification and complete characterization of the transcriptome at the isoform level [37]. By capturing the entire transcript sequence, ONT enables precise identification and analysis of alternative splicing events and isoform diversity. Another noteworthy advantage is the ability to directly sequence native DNA and RNA molecules, enabling simultaneous identification of base modifications, such as DNA methylation or RNA modifications, alongside nucleotide sequence information. This feature provides valuable insights into epigenetic regulation and post-transcriptional modifications. Furthermore, ONT platforms offer innovative methods for portable, scalable, cost-effective, and high-throughput sequencing, which are especially beneficial for resource-limited settings and field applications. These advancements allow sequencing field application, even in remote locations, making it possible to investigate the genome of complex species.

References

1. Laura, M.; Borghi, C.; Bobbio, V.; Allavena, A. The Effect on the Transcriptome of *Anemone cornaria* Following Infection with Rust (Tranzschelia Discolor). *PLOS ONE* **2015**, *10*, e0118565, doi:10.1371/journal.pone.0118565.

2. Rauter, S.; Sun, Y.; Stock, M. Visual Quality, Gas Exchange, and Yield of *Anemone* and *Ranuculus* Irrigated with Saline Water. *HortTechnology* **2021**, *31*, 763–770, doi:10.21273/HORTTECH04930-21.

3. Perevolotsky, A.; Schwartz-Tzachor, R.; Yonathan, R.; Ne'eman, G. Geophytes-Herbivore Interactions: Reproduction and Population Dynamics of *Anemone cornaria* L. *Plant Ecology* **2011**, *212*, 563–571.

4. Yari, V.; Roein, Z.; Sabouri, A. Exogenous 5-AzaCitidine Accelerates Flowering and External GA3 Increases Ornamental Value in Iranian *Anemone* Accessions. *Sci Rep* **2021**, *11*, 7478, doi:10.1038/s41598-021-86940-6.

5. Laura, M.; Allavena, A. *Anemone cornaria* Breeding: Current Status and Perspectives. *European Journal of Horticultural Science* **2007**, *72*, 241–247.

6. Copetta, A.; Dei, F.; Marchioni, I.; Cassetti, A.; Ruffoni, B. Effect of Thermal Shock in the Development of Androgenic Plants of *Anemone coronaria* L.: Influence of Genotype and Flower Parameters. *Plant Cell Tiss Organ Cult* **2018**, *134*, 55–64, doi:10.1007/s11240-018-1399-4.

7. Laura, M.; Allavena, A.; Magurno, F.; Lanteri, S.; Portis, E. Genetic Variation of Commercial *Anemone coronaria* Cultivars Assessed by AFLP. *The Journal of Horticultural Science and Biotechnology* **2006**, *81*, 621–626, doi:10.1080/14620316.2006.11512114.

8. Copetta, A.; Laura, M. The Double-Layer Method to the Genesis of Androgenic Plants in *Anemone coronaria*. *Methods Mol Biol* **2021**, *2264*, 187–196, doi:10.1007/978-1-0716-1201-9_13.

9. Martina, M.; Acquadro, A.; Portis, E.; Barchi, L.; Lanteri, S. Diversity Analyses in Two Ornamental and Large-Genome Ranunculaceae Species Based on a Low-Cost Klenow NGS-Based Protocol. *Frontiers in Plant Science* **2023**, *14*.

10. Wenzel, W.; Hemleben, V. A Comparative Study of Genomes in Angiosperms. *Pl Syst Evol* **1982**, *139*, 209–227, doi:10.1007/BF00989326.

11. Veselý, P.; Bureš, P.; Šmarda, P.; Pavlíček, T. Genome Size and DNA Base Composition of Geophytes: The Mirror of Phenology and Ecology? *Ann Bot* **2012**, *109*, 65–75, doi:10.1093/aob/mcr267.

12. Martina, M.; Acquadro, A.; Barchi, L.; Gulino, D.; Brusco, F.; Rabaglio, M.; Portis, F.; Portis, E.; Lanteri, S. Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in *Anemone coronaria* L. *International Journal of Molecular Sciences* **2022**, *23*, 3126, doi:10.3390/ijms23063126.

13. Beruto, M.; Rabaglio, M.; Viglione, S.; Labeke, M.-C.V.; Dhooghe, E. *Ranuculus*. In *Ornamental Crops: Handbook of Plant Breeding*; Springer, 2018; pp. 649–671.

14. Goepfert, D. *Karyotypes and DNA Values in Ranunculus L. and Related Genera*; Thesis (Ph.D.)--University of Toronto, 1974;

15. Ohkawa, K. Growth and Flowering of *Ranuculus asiaticus*. *Acta Hortic.* **1986**, 165–172, doi:10.17660/ActaHortic.1986.177.22.

16. Beruto, M.; Martini, P.; Viglione, S. *Ranunculus Asiaticus*: From Research to Production. *Acta Hortic.* **2019**, 117–128, doi:10.17660/ActaHortic.2019.1237.16.

17. Zhao, L.; Bachelier, J.B.; Chang, H.; Tian, X.; Ren, Y. Inflorescence and Floral Development in *Ranuculus* and Three Allied Genera in Ranunculeae (Ranunculoideae, Ranunculaceae). *Plant Systematics and Evolution* **2012**, *298*, 1057–1071.

18. Dhooghe, E.; Grunewald, W.; Reheul, D.; Goetghebeur, P.; Van Labeke, M.-C. Floral Characteristics and Gametophyte Development of *Anemone coronaria* L. and *Ranuculus asiaticus* L. (Ranunculaceae). *Scientia Horticulturae* **2012**, *138*, 73–80, doi:10.1016/j.scienta.2011.10.004.

19. Martina, M.; Acquadro, A.; Gulino, D.; Brusco, F.; Rabaglio, M.; Portis, E.; Lanteri, S. First Genetic Maps Development and QTL Mining in *Ranuculus asiaticus* L. through DdRADseq. *Frontiers in Plant Science* **2022**, *13*.

20. Ferrero, F.; Duclos, A.; Ottenwaëlder, L.; Thiebaut, M.J.; Jacob, Y. Use of Homozygosity in the *Ranuculus asiaticus* Breeding Process. *Acta Horticulturae* **2006**, *714*, 119–128, doi:10.17660/ActaHortic.2006.714.14.

21. Beruto, M.; Debergh, P. Micropropagation of *Ranuculus asiaticus*: A Review and Perspectives. *Plant Cell, Tissue and Organ Culture* **2004**, *77*, 221–230, doi:10.1023/B:TICU.0000018416.38569.7b.

22. Borriello, R.; Maccario, D.; Viglione, S.; Bianciotto, V.; Beruto, M. Arbuscular Mycorrhizal Fungi and Micropropagation of *Ranuculus asiaticus* L.: A Useful Alliance? *Acta Horticulturae* **2017**, 549–556, doi:10.17660/ActaHortic.2017.1155.81.

23. Hasan, N.; Choudhary, S.; Naaz, N.; Sharma, N.; Laskar, R.A. Recent Advancements in Molecular Marker-Assisted Selection and Applications in Plant Breeding Programmes. *J Genet Eng Biotechnol* **2021**, *19*, 128, doi:10.1186/s43141-021-00231-1.

24. Gimelfarb, A.; Lande, R. Marker-Assisted Selection and Marker-QTL Associations in Hybrid Populations. *Theor Appl Genet* **1995**, *91*, 522–528, doi:10.1007/BF00222983.

25. Hayes, B.J.; Chamberlain, A.J.; McPartlan, H.; Macleod, I.; Sethuraman, L.; Goddard, M.E. Accuracy of Marker-Assisted Selection with Single Markers and Marker Haplotypes in Cattle. *Genet Res* **2007**, *89*, 215–220, doi:10.1017/S0016672307008865.

26. Kushanov, F.N.; Turaev, O.S.; Ernazarova, D.K.; Gapparov, B.M.; Oripova, B.B.; Kudratova, M.K.; Rafieva, F.U.; Khalikov, K.K.; Erjigitov, D.Sh.; Khidirov, M.T.; et al. Genetic Diversity, QTL Mapping, and Marker-Assisted Selection Technology in Cotton (Gossypium Spp.). *Frontiers in Plant Science* **2021**, *12*.

27. Kumar, S.; Banks, T.W.; Cloutier, S. SNP Discovery through Next-Generation Sequencing and Its Applications. *Int J Plant Genomics* **2012**, *2012*, 831460, doi:10.1155/2012/831460.

28. Bernardo, A.; Wang, S.; Amand, P.S.; Bai, G. Using Next Generation Sequencing for Multiplexed Trait-Linked Markers in Wheat. *PLOS ONE* **2015**, *10*, e0143890, doi:10.1371/journal.pone.0143890.

29. Ray, S.; Satya, P. Next Generation Sequencing Technologies for next Generation Plant Breeding. *Frontiers in Plant Science* **2014**, *5*.

30. Sahu, P.K.; Sao, R.; Mondal, S.; Vishwakarma, G.; Gupta, S.K.; Kumar, V.; Singh, S.; Sharma, D.; Das, B.K. Next Generation Sequencing Based Forward Genetic Approaches for Identification and Mapping of Causal Mutations in Crop Plants: A Comprehensive Review. *Plants* **2020**, *9*, 1355, doi:10.3390/plants9101355.

31. Davey, J.W.; Blaxter, M.L. RADSeq: Next-Generation Population Genetics. *Brief Funct Genomics* **2010**, *9*, 416–423, doi:10.1093/bfgp/elq031.

32. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE* **2011**, *6*, e19379, doi:10.1371/journal.pone.0019379.

33. Sun, X.; Liu, D.; Zhang, X.; Li, W.; Liu, H.; Hong, W.; Jiang, C.; Guan, N.; Ma, C.; Zeng, H.; et al. SLAF-Seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *PLOS ONE* **2013**, *8*, e58700, doi:10.1371/journal.pone.0058700.

34. Ziarsolo, P.; Hasing, T.; Hilario, R.; Garcia-Carpintero, V.; Blanca, J.; Bombarely, A.; Cañizares, J. K-Seq, an Affordable, Reliable, and Open Klenow NGS-Based Genotyping Technology. *Plant Methods* **2021**, *17*, 30, doi:10.1186/s13007-021-00733-6.

35. Athanasopoulou, K.; Boti, M.A.; Adamopoulos, P.G.; Skourou, P.C.; Scorilas, A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. *Life (Basel)* **2021**, *12*, 30, doi:10.3390/life12010030.

36. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* **2015**, *13*, 278–289, doi:10.1016/j.gpb.2015.08.002.

37. Ying, Y.-L.; Hu, Z.-L.; Zhang, S.; Qing, Y.; Fragasso, A.; Maglia, G.; Meller, A.; Bayley, H.; Dekker, C.; Long, Y.-T. Nanopore-Based Technologies beyond DNA Sequencing. *Nat. Nanotechnol.* **2022**, *17*, 1136–1146, doi:10.1038/s41565-022-01193-2.

Aim and thesis structure

Since the anemone and buttercup are two important crops for Italy, especially in the Ligurian region, the purpose of the thesis is to increase the genetic knowledge on these species and make it available for developing strategies of targeted breeding programs. The thesis is divided into 5 chapters: Chapter I and Chapter II are focused on poppy anemone, Chapter III focused on persian buttercup, while in Chapter IV both species are investigated through a novel genotyping protocol, and finally in Chapter V a first version of the genome of *R. asiaticus* is presented.

Chapter I "Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in *Anemone coronaria* L.", presents the development of the first draft genome of *A. coronaria*, a perennial plant with a large genome size, using Illumina sequencing. The genome assembly consisted of 2 million scaffolds, and a database called "*Anemone coronaria* Microsatellite DataBase" (*AnCorDb*) was created to incorporate SSR markers. The markers were validated and applied for cultivars fingerprinting, assessing intra-cultivar variability, and can be utilized in various applications such as Breeding Rights disputes, genetic mapping, marker-assisted breeding (MAS), and phylogenetic studies.
A paper has already been published
(https://doi.org/10.3390/ijms23063126).

Chapter II "Microsatellite-based identification of double-haploid plants by androgenesis in *Anemone coronaria* L.", presents a technique for discriminating between double-haploid (DH) plants and somatic embryogenesis-derived *A. coronaria* plants. By selecting microsatellites with heterozygous alleles in the anthers mother plant, a fast and easy technique was developed to identify DH plants in the early stages of plant development, as they exhibit one allele, while plants from somatic embryogenesis show two alleles. This method helps overcome challenges in producing F_1 hybrids and facilitates the selection of homozygous plants in *A. coronaria* breeding programs.

The chapter has been presented as Poster Presentation at the XXVII International EUCARPIA Symposium Section Ornamentals, and it is currently in press in *Acta Horticolturae*.

Chapter III "First genetic maps development and QTL mining in *Ranunculus asiaticus* L. through ddRADseq", presents the development of molecular-genetic maps in Persian Buttercup using a double digest restriction-site associated DNA (ddRAD) approach. The maps were constructed based on two F_1 mapping populations and revealed eight linkage groups corresponding to the species' haploid chromosome number. The maps were used for quantitative trait loci (QTL) analysis, leading to the identification of major QTLs associated with purple pigmentation.

A paper has already been published (https://doi.org/10.3389/fpls.2022.1009206).

Chapter IV "Diversity analyses in two ornamental and large-genome Ranunculaceae species based on a low-cost Klenow NGS-based protocol", presents the application of the K-seq protocol to generate high-throughput sequencing data and identify genetic polymorphisms in persian buttercup and poppy anemone. Primer design was based on a related species' genome sequence and allowed the identification of thousands of SNPs to assess the genetic diversity of commercial varieties. The study introduces molecular fingerprinting in persian buttercup and demonstrates the efficiency of K-seq for genotyping complex genetic backgrounds, comparing the results with a previously published SSR-based fingerprinting in poppy anemone.

A paper has already been published (https://doi.org/10.3389/fpls.2023.1187205)

Chapter V "The first reference genome of *Ranunculus asiaticus* L. reveals a key region related to anthocyanin pigmentation", presents the successful assembling and annotation of the first version of the *R. asiaticus* genome. Additionally, the alignment of the previously developed genetic maps with the assembled genome allowed the identification of genomic regions influencing anthocyanin pigmentation.

The chapter has been presented as Oral Presentation at the XXVII International EUCARPIA Symposium Section Ornamentals, and it is currently in press in *Acta Horticolturae*.

Chapter I

International Journal of Molecular Sciences

MDPI

Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in *Anemone coronaria* L.

Matteo Martina[1], Alberto Acquadro[1], Lorenzo Barchi[1], Davide Gulino[1], Fabio Brusco[2], Mario Rabaglio[2], Flavio Portis[3], Ezio Portis [1,*] and Sergio Lanteri[1]

1. DISAFA, Plant Genetics and Breeding, University of Turino, Largo P. Braccini 2, 10095 Grugliasco, Italy;
2. Biancheri Creazioni, 18033 Camporosso, Italy; fabio@bianchericreazioni.it (F.B.); mario@bianchericreazioni.it (M.R.)
3. Yebokey, Turin (10100), Italy; flavioportis@gmail.com

Abstract: *Anemone coronaria* L. (*2n=2x*=16) is a perennial, allogamous, highly heterozygous plant marketed as a cut flower or in gardens. Due to its large genome size, limited efforts have been made in order to develop species-specific molecular markers. We obtained the first draft genome of the species by Illumina sequencing an androgenetic haploid plant of the commercial line "MISTRAL® Magenta". The genome assembly was obtained by applying the MEGAHIT pipeline and consisted of 2×10^6 scaffolds. The SciRoKo SSR (Simple Sequence Repeats) search module identified 401.822 perfect and 188.987 imperfect microsatellites motifs. Following, we developed a user-friendly "*Anemone coronaria* Microsatellite DataBase" (*AnCorDb*), which incorporates the Pri-mer3 script, making it possible to design couples of primers for downstream

application of the identified SSR markers. Eight genotypes belonging to eight cultivars were used to validate 62 SSRs and a subset of markers was applied for fingerprinting each cultivar, as well as to assess their intra-cultivar variability. The newly developed microsatellite markers will find application in Breeding Rights disputes, developing genetic maps, marker assisted breeding (MAS) strategies, as well as phylogenetic studies.

Keywords: poppy anemone; SSRs; genome sequencing; fingerprinting;

1. Introduction

Anemone genus belongs to the Ranunculaceae family and the nowadays most cultivated species (*A. coronaria* L., *A. hortensis* L., and *A. pavoniana* Lam.) originated in the Mediterranean basin. *A. coronaria* L., also known as poppy anemone, is an herbaceous, perennial crop cultivated both as a cut flower and garden plant [1]. It is a diploid species characterized by 16 chromosomes (2n = 2x = 16), but some of the commercial varieties are tetraploid. The cultivars exploited as cut flower are early flowering and produce robust stems carrying flowers with large petals and sepals, while garden cultivars produce erect leaves and a higher number of smaller flowers with short petioles [2]. Poppy anemone is allogamous, due to protogyny, and highly heterozygous [3]. Although self pollination is possible, the species is characterized by marked inbreeding depression [4], which precludes the obtainment of pure lines suitable to produce F_1 hybrid seeds. Commercial cultivars are produced by inter-crossing of selected heterozygous plants and show variable levels of internal genetic

variability. Growers plant rhizomes, which are generated after one season of nursery cultivation [5].

In previous studies, DNA markers techniques have been applied and adapted mainly for access intra-cultivars genetic variability or to perform varietal fingerprinting, [5–7] but no examples are reported in literature on the development of DNA species-specific markers.

The advances in next-generation sequencing (NGS) techniques and the progressive reduction of sequencing costs facilitated the obtainment of draft sequence genomes in many plant species. However, due to the large size of *A. coronaria* genome, estimated between 9.08 and 11.93 Gb according to the analyzed genotype [8,9], a reference genomic sequence of the species is not available.

We generated the first draft for genome sequence of *A. coronaria*, and we report on the massive microsatellite loci identification following its genome-wide survey. Microsatellite - alias simple sequence repeats (SSR) - markers, are co-dominantly inherited, ubiquitous, highly polymorphic, and have found large application in plant breeding and phylogenetic studies because of their simple application through conventional PCR protocols [10–16]. Unlike single nucleotide polymorphisms (SNPs), which have become the gold standard among molecular markers, SSRs show the advantage of being multi-allelic and highly informative, characterized by a certain level of transferability between related species [17–20], and are easily and automatically scorable.

Based on the microsatellites identified, we developed a public dynamic database, which also provides need-based primer designing facilities and represents the first on-line SSR loci resource available for the scientific

community and breeders of poppy anemone and related species. Furthermore, a set of the newly developed markers have been validated in commercial cultivars.

2. Results and Discussion

2.1. Draft Genome Assembly and Annotation

Since *A. coronaria* is a highly heterozygous species, the sequence divergence between alleles in a diploid genotype may hinder a reliable contig assembly of its genome sequence [21,22]. In order to overcome this hurdle, we performed DNA sequencing of a haploid androgenetic plant originated through *in vitro* anther culture of a diploid plant of the cultivar MISTRAL® Magenta. Overall, 91.24 Gb of cleaned reads were generated and used as input for genome assembly (Supplementary Materials). The obtained draft assembly consisted of ~4.7×10^6 scaffolds (N50 = 5,046 bp) for a total genome size of 6.94 Gb. By removing scaffolds shorter than 500 bp, their number was reduced to 2×10^6 (N50 = 6,157 bp), for a total genome size of ~6.13 Gb (Supplementary Materials). K-mer analyses of Illumina sequencing data were performed in order to es-timate the genome size of the MISTRAL® Magenta genotype. For the 19-mer frequency distribution, the number of K-mers was 3,100,416,21, with a plot peak around 4 (times each 19-mear occurs - see Supplementary Materials). According to our analysis (see Section 3), the MISTRAL® Magenta genome size was estimated around 7.8 Gb, leading our final assembly to cover ~78.6% of the genotype genome.

After the masking of the draft genome, about 75% of the sequences were classified as repetitive elements. This result is in accordance with what was previously reported in the literature, namely that the expansion of gigantic

genomes has been driven by the proliferation of transposable elements [23,24]. Indeed, also due to the sequencing of short libraries (270 bp), the huge amount of repetitive content hampered the assembly procedures and biased some assembly metrics.

The masked assembled draft genome was structurally annotated with the Maker-P suite, identifying an overall number of 26,260 genes (AED ≤ 0.4) covering ~56.12 Mb (0.92%) of the estimated genome size. Functional annotation performed through InterProScan domain inspection highlighted about 84% of the predicted proteins with at least one IPR domain. Among the top SUPERFAMILY domains, the most abundant (8.62%) was SSF56112 (protein kinase-like domain), which acts on regulatory and signaling processes in the eukaryotic cell. The second most represented superfamily (6.19%) was SSF52540 (P-loop containing nucleoside triphosphate hydrolase) which is involved in several UniPathways, such as chlorophyll or CoA biosynthesis, followed by SSF48264 (4.10%-Cytochrome P450). These superfamilies have been previously re-ported as highly abundant in various genomic backgrounds [25–29].

2.2. The SSR Content of the Poppy Anemone Draft Genome

In the assembled poppy anemone genome, a total of 401,822 perfect SSR motifs (density of 65.52 SSR/Mb), which included 42,111 compound SSRs, and 188,987 imperfect SSR motifs were identified (Table 1).

Table 1. Microsatellite motifs distribution across the assembled genome. Perfect (including compound) and imperfect SSR are reported.

		Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-	Total/Mean
Perfect SSR	Types	2	4	10	32	91	304	443
	Count	13,475	241,693	95,326	27,203	10,805	13,320	401,822
	%	3.4	60.2	23.7	6.8	2.7	3.3	100
	Density (SSR/Mbp)	2.2	39.41	15.54	4.44	1.76	2.17	65.52
	Cumulative (Mbp)	0.05	1.94	1.14	0.43	0.21	0.32	4.11
	Cumulative (%)	0.08%	47.20%	27.74%	10.46%	5.11%	7.79%	100%
	Mean Repeat Number	22.7	11.3	6.9	5.2	5	6.8	57.9
Imperfect SSR	Count	2823	111,281	38,183	10,719	12,920	13,061	188,987
	%	1.49%	58.88%	20.20%	5.67%	6.84%	6.91%	100%
	Density (SSR/Mbp)	0.46	18.14	6.23	2.15	2.11	2.13	31.22

Six classes of perfect SSRs were evaluated (from mono- to hexanucleotide) for their abundance in the assembled genome. Dinucleotides were the most abundant, in accordance with what has been previously reported in literature [30–39], representing 60.2% of the identified SSRs. Trinucleotides were the second most abundant class (23.7%), followed by tetranucleotides (6.8%). Penta-, hexa-, and mononucleotides cov-ered the remaining percentage and showed analogous frequency ranging from 2.7 to 3.4% (Figure 1a). The most represented dinucleotide motifs, AT/AT, AG/CT, and AC/GT, accounted respectively for 72.56%, 16.05%, and 11.39% (Figure 1b), while CG/GC motifs were approximately absent (0.009%). The high abundance of AT/AT motifs was in line with a number of previously reported genome surveys, confirming these microsatellites as the most represented dinucleotide motifs in higher plants. Within the trinucleotide repeat motifs, the most abundant were AAG/CTT, accounting for 36.84%, AAT/ATT for 16.55%, and ATC/GAT for 15.75% (Figure 1c).

The variation of perfect microsatellites repeats was investigated in all SRR classes (Supplementary Materials). As previously reported, longer repeats (>25) tend to be less abundant in the genome [37,38,40,41]. As can be observed in Figure 2, the tri-, tetra-, penta-, and hexanucleotides relative distribution was higher between one and 10 motif repeats, while mononucleotides distribution increased from 14 motif repeats onward and dinucleotides showed higher abundance between 8 and 19 motif repeats.

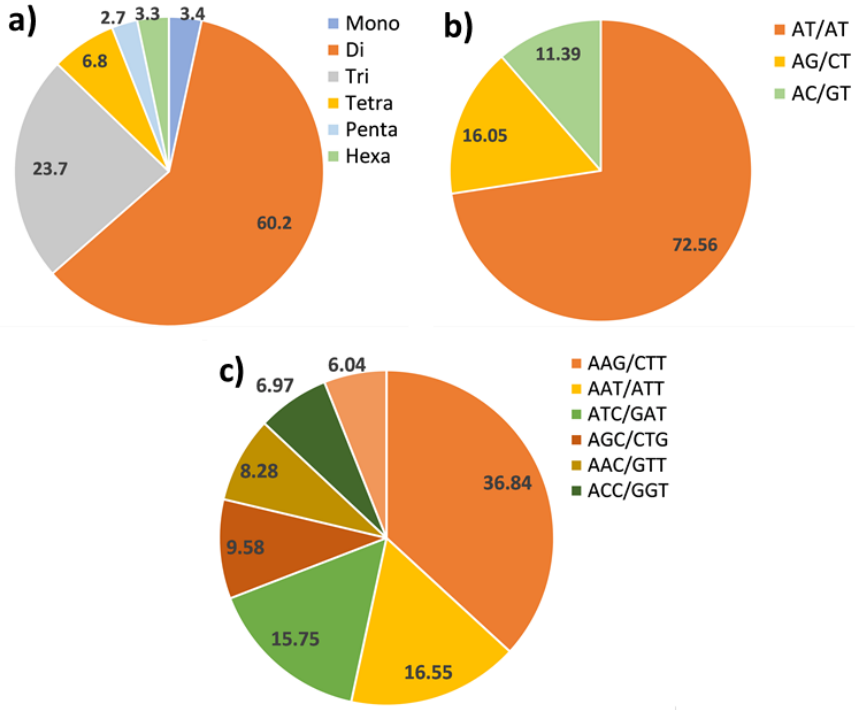


Figure 1. Microsatellites distribution in the poppy anemone genome. (a) Percentage distribution of the most frequent classes of SSRs; (b) dinucleotides motifs e and (c) main trinucleotides motifs identified by SciRoKo.

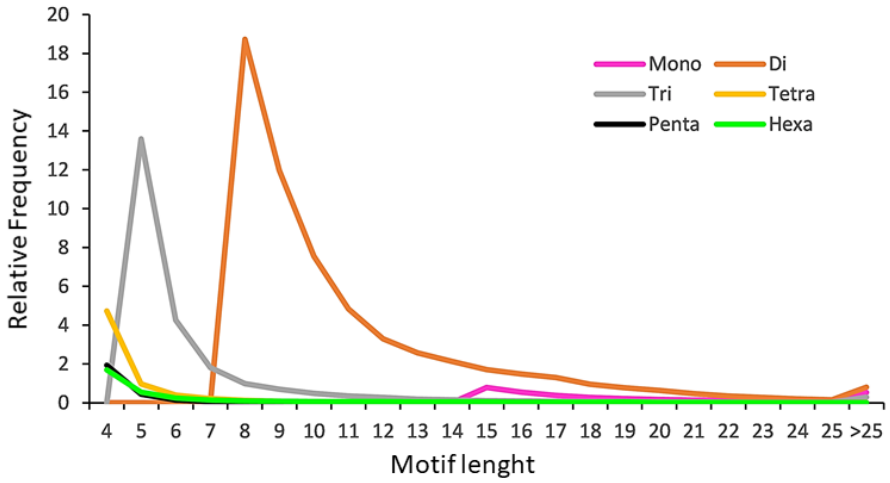
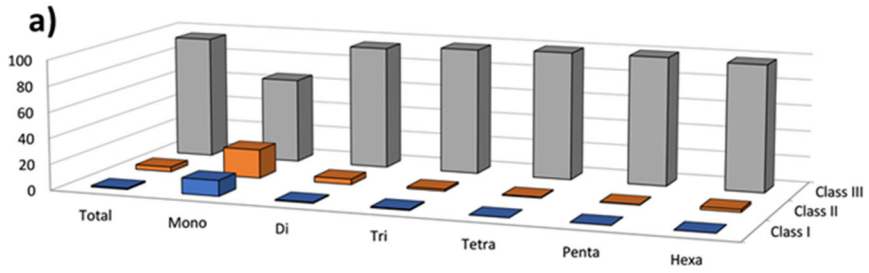


Figure 2. The relative frequency of SSR motifs with different lengths, classified by the number of repeats.

Based on the number of motif repeats, 0.95% of SSRs were classified within the hypervariable class I (≥30 motif repeats), 3.71% were assigned to the potentially varia-ble class II (20–30 motif repeats) types, while the remaining 95.34% were included in the variable class III (<20 motif repeats) types (Figure 3a).



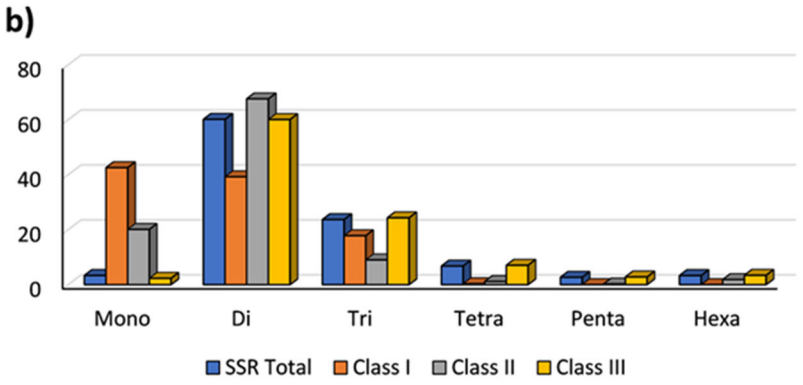


Figure 3. (a) The frequency of repeat classes (class I > 30 motif repeats, class II 20–30 motif repeats, class III < 20 motif repeats; (b) the distribution of motif type within each class.

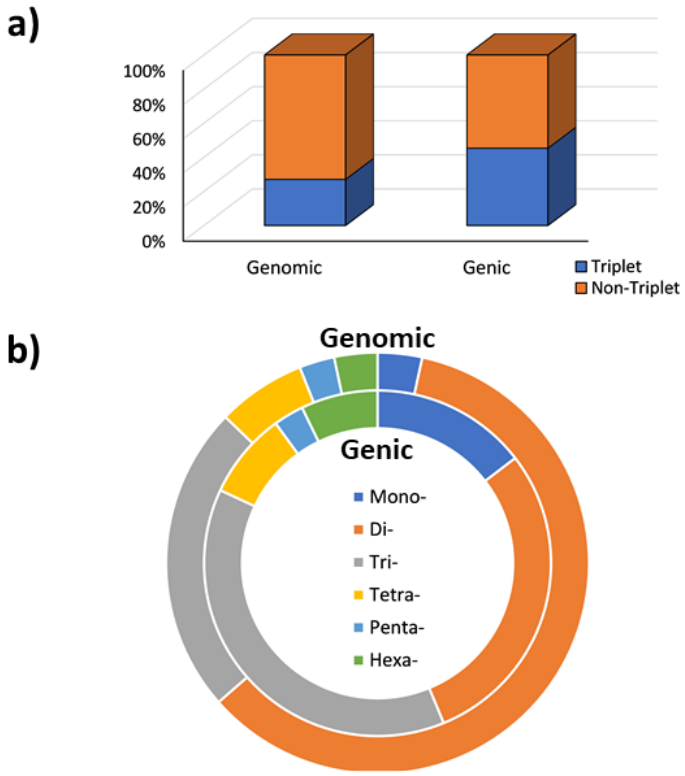
Compared with our previously published data [37,38], in which SSR classification was based on microsatellite length (nt), the present classification reports a lower number of SSRs belonging to class I and II. The choice of shifting from a microsatellite length (nt) classification to a repeat number-based one was performed to maximize the polymorphism discrimination power and informativeness of the Class I and Class II markers (Figure 3b).

2.3. Gene Context of SSRs

The obtained gene annotation made it possible to investigate the distribution of microsatellites across the gene space. Overall, 3,223 perfect (0.80% of the total) and 1,261 imperfect SSRs (0.67%) were associated with 3,223 and 1,261 genes respectively, representing 0.23% of the gene space. These SSRs were estimated to cover a total of 134 Kb, values which

translates to a density across the gene space of 57.48 and 22.52 SSRs/Mbp for perfect and imperfect motifs, respectively.

We also investigated the perfect SSR motifs detected in the global set of genomic and genic SSRs. The microsatellites were classified in non-triplet repeats (mono-, di-, tetra- and pentanucleotides), and triplet repeats (tri- and hexanucleotides), and a fair balance between the two classes (45.33% triplets; 54.67% non-triplets; Figure 4a) was detected in genic SSRs, while in the whole genomic set the triplets were just about 27%.



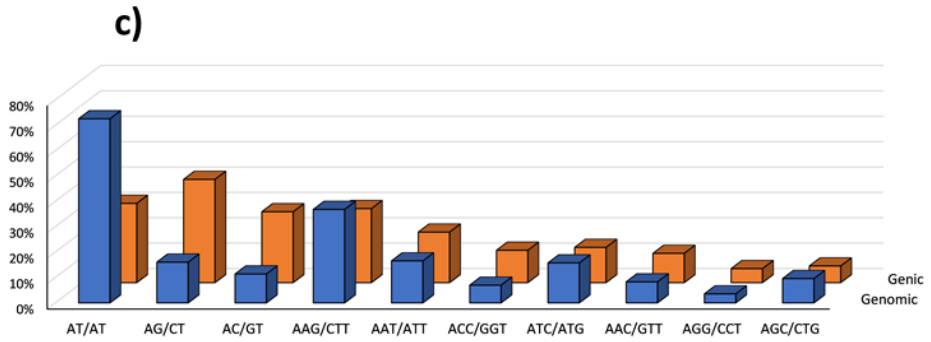


Figure 4. (a) Non-triplet SSR vs. triplets SSR; (b) distribution of repeat types within perfect and imperfect SSR motifs in both the genomic and genic regions; (c) Comparison between di- and tri-nucleotide repeats in both full genomic regions and gene space.

Trinucleotides were the most common class among the genic perfect microsatellites (38.2%), and the second most common class were the dinucleotides (29.0%; Figure 4b). The predominance of trinucleotides in the gene space has been widely reported in literature as a direct effect of negative selection against frameshift mutations in coding regions [38,42–45]. Furthermore, the increase of trinucleotides frequency in genomic coding regions might be due to a positive selection for specific single amino acids [46,47]. For this reason, the most frequent trinucleotides genic perfect SSR motif types were investigated (Figure 4c), identifying AAG/CTT, coding for lysine, as the most represented motif (11.14%), followed by AAT/ATT (7.60%), ATC/GAT (5.31%), and ACC/GGT (4.87%) coding for aspargine, isoleucine, and threonine respectively. In the genic regions, the most common dinucleotides were AG/CT (11.88%) followed by AT/AT (9.12%). The predominance of AG/CT motif in gene

sequences has been widely reported in literature, as well as the higher frequency of AT/AT in the non-transcribed regions. Being present in transcripts, genic SSRs have been reported as an important class of "functional markers" (DNA markers derived from functionally characterized sequence motifs [48]) playing a crucial role in gene expression in both mammals and plants [49–53]. Furthermore, genic microsatellite markers have been reported to possess higher portability among related species, making it possible to use them as anchor markers in comparative genetics [54].

The GO categorisation of the genic SSR highlighted 1,113 sub-categories of three main GO categories-Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Thirteen sub-GO categories represented ~33% of the identified entries (Figure 5).

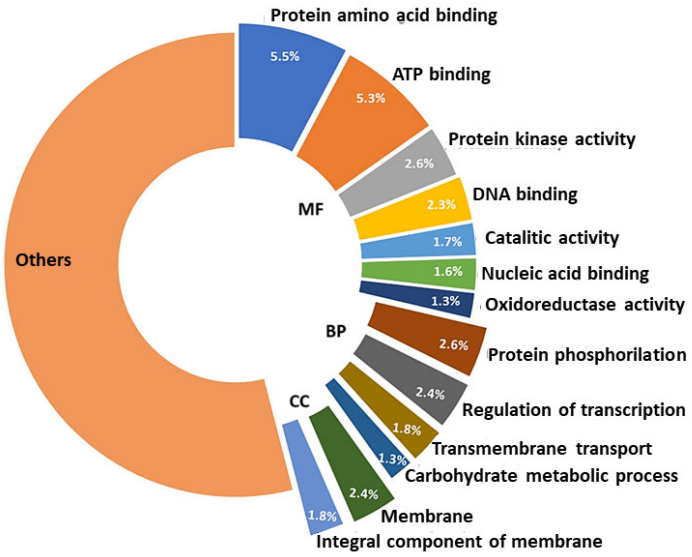


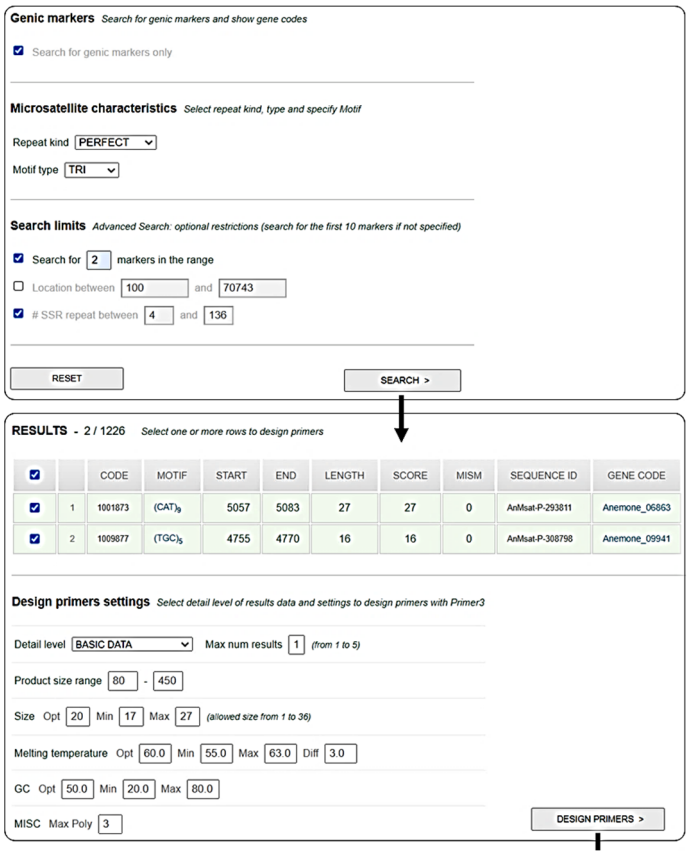
Figure 5. Thirteen main sub-GO categories of genes containing SSRs.

The MF sub-categories “protein amino acid binding” (GO:0005515) and “ATP binding” (GO:0005524) represented more than the 10% of the overall identified accessions. The occurrence of SSRs within specific gene functions has been previously observed [22,37,55–57], as well as the presence of SSRs in binding-associated genes, specifically in the 5′-UTR regions [49]. Unexpectedly, only 2.4% of the SSRs identified fell among the “regulation of transcription” (GO:0006355) sub-GO category (BP) as the accumulation of microsatellites in transcription factors, and more in general in transcription regulation loci has been repeatedly reported in literature [22,37,55–57].

2.4. *AnCorDB* Construction, System Architecture, Features and Utility

A public and searchable database of the microsatellites data reported in this paper was developed (*AnCorDB* - available at www.anemone.unito.it, accessed on 09/03/2022). It offers similar features to the CyMSatDB [37] and the EgMiDB databases [38] and it can be used to retrieve SSRs based on either simple or complex searches. The database provides browsable access to all the SSRs identified in the poppy anemone genome. SSRs can be retrieved on the basis of simple characteristics, such as “SSR feature” (whole genomic or only genic SSR), “repeat kind” (perfect vs imperfect), or advanced characteristics, such as “motif type” (mono- to hexanucleotide), “specific motif sequence”, “repeat number”. Multiple parameters can be also combined to search for a specific set of SSRs as per user requirement, as researchers can limit the search via motif repetition

and number of markers required (1–99). Scaffold position can be changed through a dedicated query. The output lists a wide range of information (SSR identifiers, scaffold number, motif type and length, genomic location - start and end position, SSR length - including an optional download of the flanking sequences. Primer3 tool is implemented in the database, allowing primers de-sign through the "Design Primers" button which directs the use to a list of up to five possible primer pairs, with their melting temperatures (Tm), their GC content, and the expected length of the amplicon. The obtained pairs of primers can be downloaded in MS-Excel format (Figure 6).



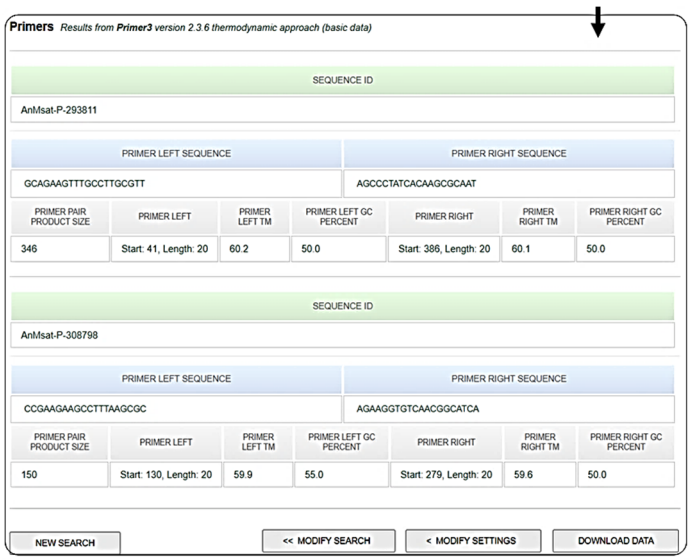


Figure 6. Example of SSR search and primer design at *AnCorDb*.

2.5. SSR Validation and Varietal Fingerprinting

A set of 150 microsatellite loci was selected as representative of the overall ge-nome distribution of every class and motif, their primer pairs were designed, and they were PCR-validated. On the basis of the amplicon quality, 62 SSRs were selected for varietal fingerprinting of poppy anemone cultivars (Supplementary Materials). In some cases, we detected a low efficiency of the primer design which could be attributed to the low coverage of our draft genome, leading to misassembly in the repetitive ele-ments regions [58,59]. Nevertheless, the percentage of suitable primer pairs resulted in line with the one detected in previous SSR mining reports based on draft genome se-quence obtained at low-coverage [60–62].

We tested the selected 62 SSR markers on eight commercial cultivars of which six were diploids and two tetraploids (see Section 3), representative

of the phenotypic variability of the varieties marketed by Biancheri Creazioni. A total of 203 alleles were generated, with a mean of 3 (range 1–8) alleles per locus. The largest range in amplicon length detected was 199–604 bp, resulting from the amplification of *Ancor33*, a dinucleotide AT motif. In the 25.8% of the loci, the assay generated the amplicon predicted length, while in the 48.4% the amplicon was longer than expected and in the 25.8% shorter. Only three SSRs were monomorphic across the evaluated genotypes and thirty markers were nullallelic for at least one genotype. The polymorphism information content (PIC) values of the polymorphic SSRs varied from 0.13 to 0.85 (mean 0.52 ± 0.025). AnCor49 had the highest PIC, and *AnCor71* the lowest (Supplementary Materi-als).

The scored allele peaks were used to elaborate a UPGMA-based dendrogram (Fig-ure 7a) which allowed the fingerprinting of each cultivar.

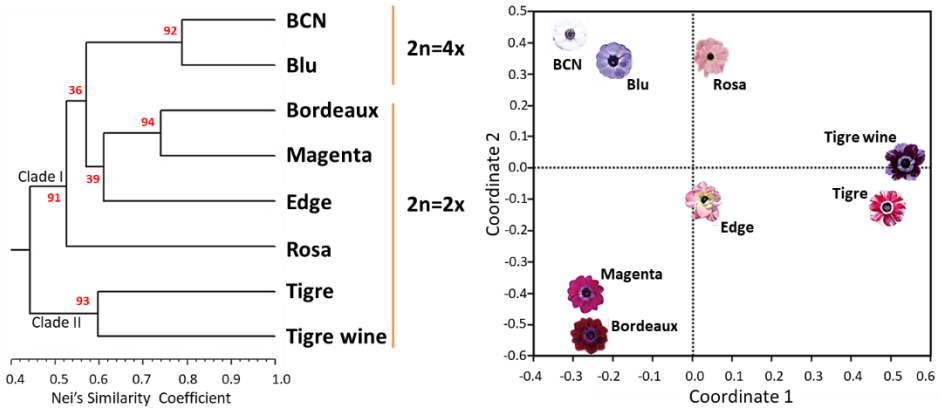


Figure 7. UPGMA dendrogram (left) and PCoA analysis of the eight varieties based on 62 mi-crosatellite loci (right). Bootstrap values (%) are reported in red.

The detected genetic relationships among varieties were in accordance with their breeding origin (Biancheri, personal communication). Two major clades were identified and supported with bootstrap values higher than 90. In Clade I, the two tetraploid cultivars ("BCN" and "Blu") clustered with an average genetic similarity of 78% and a bootstrap probability of 92%, while among the other three cultivars, "Edge" resulted more genetically differentiated from "Bordeaux" and "Magenta", which showed a genetic similarity of 74% and clustered with a bootstrap probability of 94%. In Clade II, "Tigre" and "Tigre Wine", which in turn resulted highly differentiated, showed an average genetic similarity of 60%.

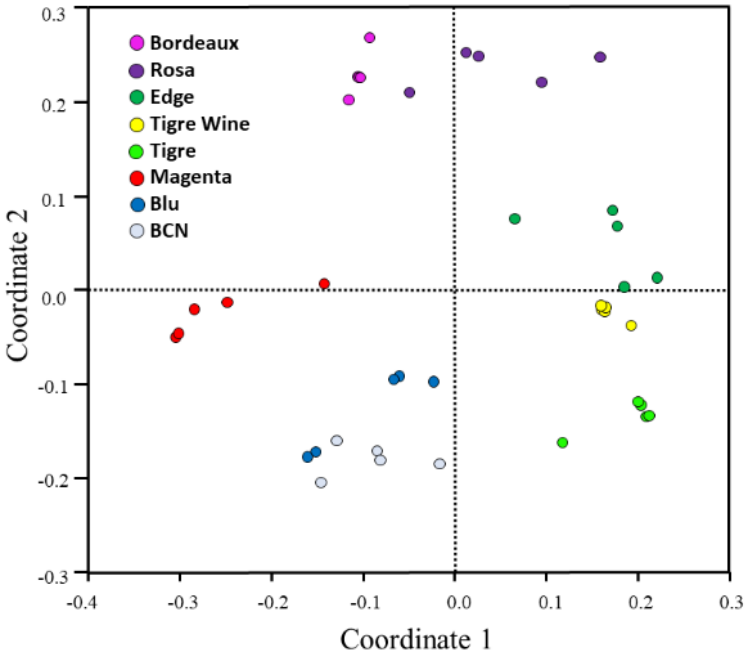
The first two axes of the PCoA scatter plot (Figure 7b) explained 42 and 31% of the overall genetic variation respectively confirming the genetic relationships between the cultivars. Interestingly, the cultivars "Edge" and "Rosa", although resulted genetically differenciated following UPGMA analyses, showed a common value for the first main coordinate of the PCoA.

Aiming at developing a fingerprint protocol for poppy anemone and identifing the minimum number of SSR loci needed to fully discriminate between the cultivars in study, we selected six SSRs. Five of them were selected based on their PIC values, namely the dinucleotide SSRs *Ancor33*, *-36*, *-49*, *-59*, and *-83*, applied togheter with the tetranucleotide *Ancor177* (detailed information in Supplementary Materials). On the basis of the 6 SSR markers, we created a similarity matrix and correlation between this matrix and the one obtained using the whole data set indicated a good fit of the genetic relationships (r = 0.92) and made it possible to fingerprint

each cultivar. This suggests their possible application as a valuable tool for varietal identification in the species.

2.6. Intra-Cultivar Variability Assessment

To assess the intra-cultivar variability among the 8 cultivars in study and furtherly validate the newly developed SSRs, five plants per cultivar were genotyped using the previously described set of six microsatellites. Fixation index (FIS) values ranged from −0.68 to 0.79. As expected from selected genotypes obtained within breeding programs, most the loci showed significant deviation from HWE, with only one marker (Ancor89) showing no significant difference between expected (HE) and observed (HO) heterozygosity values (Supplementary Materials). The principal coordinate analysis and UPGMA dendrogram illustrate the genetic relationships between members of this extended germplasm panel (Figure 8).



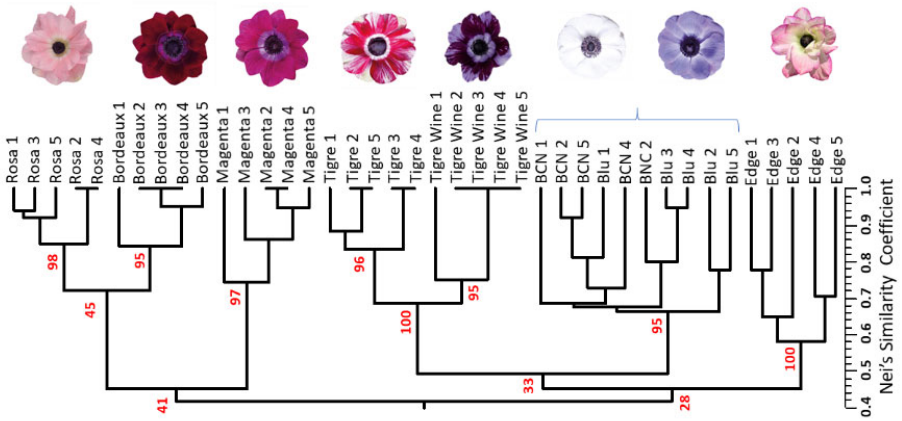


Figure 8. Dendrogram and PCoA obtained from UPGMA cluster analysis of five plants for each of the eight cultivars, based on six microsatellites (47 alleles). Bootstrap values (%) for the main nodes are reported in red.

PCoA axes 1 and 2 accounted for ~74% of the overall genetic variation, the former contributing ~48%, and the latter ~26%. As expected, the cultivar "Edge" and "Rosa" shared positive (or slightly negative) values for the first coordinate, together with "Tigre" and "Tigre Wine". "Edge" showed the highest intra-cultivar variability, while "Bordeaux" the lowest (Figure 7). The UPGMA based on 62 and six SSRs in some cases provided different clustering among the cultivars under study. This is the case of the cultivars "Rosa" and "Bordeaux", which appeared more genetically distant on the basis of 62 microsatellites (Figure 7), while closer on the basis of six SSRs (Figure 8) and with a bootstrap value as low as 45.

Each plant of the cultivars "Edge", "Blu", and "BCN" showed a unique fingerprinting, while in the other five, some plants shared common alleles. Despite the observed intra-cultivar genetic variability, the application of only six SSRs made it possible to clearly discriminate the diploid cultivars,

each of them clustered with bootstrap support from 95 to 100, while no clear genetic differentiation between the tetraploid cultivars “BCN” and “Blu” was detectable, suggesting the application of additional specific markers for their fingerprinting. For this purpose, the sixty-two amplified markers were investigated, leading to the identification of five microsatellites each of which might be individually applied for “BCN” and “Blu” discrimination (Table 2).

Table 2. List and primer sequences of the six candidate markers for cultivar discrimination analyses between BCN and Blu.

				Alleles (bp)	
SSR	SSR type	Motif	N° of Repeats	BCN	BLU
AnCor49	Di	AT	26	450; 468	435
AnCor74	Di	GT	30	501; 530	529; 539
AnCor87	Di	TC	29	271	262; 275
AnCor115	Tri	AAC	28	518	528
AnCor132	Tri	AAC	27	533; 589	536; 572
AnCor139	Tri	AAG	26	501	504; 512

3. Materials and Methods

3.1. Draft Genome Sequencing, Assembly, and Annotation

Leaves of a haploid plant originated from the commercial line MISTRAL® MAGENTA obtained through “in vitro” androgenesis by applying the regeneration protocol adapted by [5], were provided by Biancheri Creazioni (Camporosso (IM), Italy). Plant DNA Kit (E.Z.N.A.®) was used for the genomic DNA extraction following the manufacturer’s instructions. DNA quality was assessed through the NanoDrop™ 2000 spectrophotometer and the Qubit® 2.0 Fluorometer was used for DNA quantification. One microgram of DNA was used for the construction of a 270 bp

insertion library (Novogene, Hong Kong), which was sequenced using a NovaSeq Illumina platform (Illumina Inc., San Diego, CA, USA) with paired-end chemistry (2 × 150 bp). Raw reads were cleaned with Scythe (v0.994, https://github.com/vsbuffalo/scythe, accessed on 02/01/2022) for removing contaminant residual adapters and Sickle (v1.33, https://github.com/najoshi/sickle, accessed on 02/01/2022)) for removing reads with poor quality ends (Q < 30). De novo assembly was performed with standard parameters using the MEGAHIT assembler ([63]; https://github.com/voutcn/megahit, accessed on 02/01/2022)), an ultra-fast and memory efficient NGS assembler based on succinct de Bruijn graphs that can be applied both for metagenomics and single genome assembly. The quality of the genome assembly (e.g., N50, scaffolds/scaffolds number/size/length, genome length) was assessed using the perl script Assemblathon_stats.pl ([64]; https://github.com/ucdavis-bioinformatics/assemblathon2-analysis, accessed on 02/01/2022)). Cleaned reads were then used for kmer-based genome size estimation using the jelly-bean software and applying the formula Genome Size = 19-mers count/peak position of the number of times each 19-mer occurs (see Supple-mentary Materials). The assembled draft genome was pre-masked using RepeatMasker v4.1.0 [65] with a de novo approach. A species-specific repeats library was constructed following the Repeat Library Construction Advanced pipeline ([66]-http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction-Advanced, accessed on 02/01/2022)) which requires the use of mite hunter, LTRdigest, LTR_harvest (available in genome tools, v1.5.10), and Repeatmodeler v1.0.11. The new library was

then combined with Repbase-viridiplantae to identify transposable elements (TEs). TEs were classified into two main classes: Class I (retrotransposon elements) and Class II (DNA transposons). Gene prediction was performed using Maker-P v2.31.08. Augustus v3.3.2 ([67]) Hidden Markov Models, and SNAP ([68]) gene prediction algorithms were combined with transcripts and protein alignments as evidence to support the prediction. All predicted gene models were filtered and only the ones with an AED $\leq$ 0.4 were maintained. AED measures the concordance of a gene predicted with aligned transcripts, mRNA-seq, and protein homology data. AED scores range from 0 and 1, where 0 indicates perfect concordance between evidence and gene prediction, while 1 absence of concordance. To measure the quality and completeness of the predicted proteomes, a quantitative assessment was carried out based on evolutionary informed expectations of gene content known as Benchmarking Universal Single-Copy Orthologs (BUSCO v3.0.2., Embryophyta odb 10 - [69]). The sequences of the predicted proteins were also noted using InterproScan5 ([70]) compared to all the available databases (ProSitePro les-20.119 - [71], PAN-THER-10.0 - [72], Coils-2.2.1 - [73], PIRSF-3.01 - [74], Hamap-201511.02—[75], Pfam-29.0—[76], ProSitePatterns - 20.119- [71], SUPERFAMILY-1.75 - [77], ProDom-2006.1 - [78], SMART-7.1 - [79], Gene3D-3.5.0 - [80], and TIGR-FAM-15.0 - [81]). Then, GOfeat ([82]) was used to identify the enrichment of GO terms for specific gene clusters.

3.2. SSR-Mining

The un-masked draft assembly of the *A. coronaria* L. genome was used for SSR mining. Scaffolds were chopped into manageable pieces using SciRoKo tool ([83]—v3.4; https://kofler.or.at/bioinformatics/SciRoKo, accessed on 02/01/2022)), and perfect, compound, and imperfect SSRs were identified in silico using SciRoKo and the misa.pl pipeline (https://github.com/cfljam/SSR_marker_design, accessed on 02/01/2022)). A minimum of four repetitions together with a minimum length of 15 nt were requested. Any sequence was considered as a perfect SSR when a motif was repeated at least fifteen times (1 nt motif), eight times (2 nt), five times (3 nt), or four times (4–6 nt), allowing for only one mismatch. For compound repeats, the maximum default interruption (spacer) length was set at 100 bp. The coordinates (start/end position) of each SSR were matched with those of the gene space using Bedtools intersect (using the default parameters) with -loj (left outer join) option: where the overlap comprised at least 1 nt, the repeat was designated as a genic SSR. A GO categorization of the three main GO categories - "biological processes" (BP), "molecular functions" (MF), and "cellular components" (CC) - were applied to genes carrying at least one SSR.

3.3. *AnCorDB*, an SSR Database for Poppy Anemone

The *Anemone coronaria* Microsatellite DataBase (*AnCorDB*; www.anemone.unito.it, accessed on 09/03/2022)) was developed to provide browsable access to the SSR data. This web application, based on a LAMP stack, comprises a client tier (client browser), a middle tier (Apache web server with PHP interpreter), and a database tier (MySQL DBMS). A user-friendly interface was developed using PHP, which is an open-source

server-side scripting language. The set of in silico detected SSRs were stored in the MySQL database, using PHP scripts to parse the text file from SciRoKo. User need-based customized queries can be generated from the web interface and allow users to search the microsatellite marker information in MySQL database. A stand-alone version of Primer3 has been also provided to design primer pairs for any given SSR: its output lists alternative sets of primer pairs, and the characteristics of the expected amplicon.

3.4. Marker Validation

One hundred and fifty microsatellites were selected among the ones with a number of repetitions between 20 and 30, in line with the overall genome representation of every class and motif. Specifically, di- and trinucleotides were selected in the interval ranging from 25 and 30 motif repetitions, while this threshold was lowered to the interval between 20 and 30 motif repetitions for the other classes of microsatellites. These parameters were applied with the aim of obtaining the highest potential polymorphism rate of the selected markers. The primer pairs obtained from the database were used for the DNA amplification of eight *A. coronaria* cultivars representative of the phenotypic variability of the ones marketed by Biancheri Creazioni. Among them, six cultivars were diploid ("Bordeaux", "Edge", "Magenta", "Rosa", "Tigre", and "Tigre Wine"), while two were tetraploid ("BCN" and "BLU"). The following touchdown PCR protocol was applied: 94 °C for 5 min followed by 13 touchdown cycles with denaturation step at 94 °C for 30 s, a step at 60 °C for 30 s decreasing the annealing temperature of 0.38 °C every cycle, and lastly extension step at 72 °C for 30 s. At last, 35 cycles at 94 °C for 30 s (denaturation), 55 °C for 30 s

(annealing), and 72 °C for 30 s (extension), and a final extension cycle at 72 °C for 5 min. PCR products were separated using a 2% agarose gel to check their occurred amplification.

3.5. SSR Fingerprinting and Intra-Cultivar Variation Assessment

A subset of 62 SSRs (Supplementary Materials) were further analyzed through capillary sequencing (ABI PRISM® 310, Applied Biosystems™). M13-Tailed Forward primers were designed for each microsatellite and applied in a three-primers unbalanced PCR reaction with a fluorescent-labelled M13 primer ([84]). PCR was carried out in a final volume of 20 µL containing: 4 µL of 5× GoTaq Colorless Buffer (GoTaq® DNA Polymerase, Promega), 1 µL of MgCl_2 (25 mM), 0.4 µL of dNTPs (10 mM), 3 µL of DNA template (5 ng/µL), 1 µL of Reverse and M13-labeled primer (10 µM), 0.2 µL of Forward-M13 Tailed primer (10 µM), and 9.2 µL of ultrapure water. In each reaction, 1 µL of amplification product was pooled with other three products labelled with different fluorophores (FAM, VIC, NED, and PET) and purified using the PEG-precipitation method described by [85]. Multiplex genotyping reactions were carried out in ABI PRISM® 310 according to the GeneScan® Reference Guide (Applied Biosystems™). Results were visualized using Peak Scanner™ Software v1.0 (Applied Bi-osystems™) and for each microsatellite the amplicons' length was scored. A binary matrix was generated by scoring the band presence (1) and absence (0), which was used to compute pairwise similarity coefficients [86] and then to construct a UPGMA-based dendrogram [87] with 1000 bootstraps. Principal coordinate analysis (PCoA) was also performed for displaying the multi-dimensional relationship between genotypes, and

the two axes were plotted graphically, according to the extracted eigenvectors. All analyses were performed using the NTSYS software package v2.10 [88] and Past 4.09 software [89]. The polymorphic information content (PIC) was calculated for each locus as described by [90] and used for selecting the most informative SSRs and identify the lowest number of loci needed for fingerprinting each of the cultivar in study. Mantel test [91] was performed to establish correlations between the similarity matrices generated by the most informative SSRs with the one generated from the complete data set. An intra-cultivar variability assessment was also performed by applying the most informative SSR loci on five plants belonging to each of the eight cultivars. PCR reactions, capillary sequencing, UPGMA-based dendrogram, and PCoA analysis were performed as described above. Calculations of observed (HO) and expected (HE) heterozygosity and Wright's fixation index (FIS) were estimated with the program IDENTITY 1.0 [92]. Exact tests of Hardy–Weinberg equilibrium (HWE) were made by means of the software GENEPOP 3.4 [93].

4. Conclusions

The development of a draft genome assembly of *Anemone coronaria* L. represents the first step toward genomic studies in poppy anemone. Its availability made it possible to identify a wide set of SSR markers and release the comprehensive microsatellite database *AnCorDb* (www.anemone.unito.it). The latter contains a full set of information regarding both genic and non-genic, perfect and imperfect SSR loci. Its intuitive web interface and its customized primer design offer a highly flexible tool to the scientific community and breeders, exploitable for genetic as well as phylogenetic studies. Our results also demonstrated that the application of a

limited number of SSRs might be suitable for varietal discrimination and may contribute to solve Breeding Rights disputes.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/article/10.3390/ijms23063126/s1.

Author Contributions: Conceptualization: S.L., E.P., A.A., M.R., F.B., and M.M.; methodology: S.L., E.P., A.A., M.R., and M.M.; software: A.A., F.P., and M.M.; validation, M.M., D.G., and E.P.; formal analysis: M.M., D.G., E.P., A.A., and L.B.; investigation: E.P., M.M., A.A., and L.B.; resources: M.R., F.B., and D.G.; data curation: M.M., L.B., E.P., and A.A.; writing—original draft preparation: M.M., A.A., and E.P.; writing—review and editing: L.B. and S.L.; visualization: M.R., F.B., D.G., and F.P.; supervision, E.P. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by Biancheri Creazioni (Camporosso, Italy).

Data Availability Statement: Sequencing data used in this study are openly available in the NCBI database (PRJNA808392).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the analyses, or interpretation of data and approved the publication of the results.

References

1. Laura, M.; Borghi, C.; Bobbio, V.; Allavena, A. The Effect on the Transcriptome of *Anemone coronaria* following Infection with Rust (Tranzschelia Discolor). PLoS ONE 2015, 10, e0118565. https://doi.org/10.1371/journal.pone.0118565.

2. Laura, M.; Allavena, A. *Anemone coronaria* Breeding: Current Status and Perspectives. Eur. J. Hortic. Sci. 2007, 72, 241–247.

3. Horovitz, A. The pollination syndrome of *Anemone coronaria* L. ; an insect-biased mutualism. Acta Hortic. 1991, 283–287. https://doi.org/10.17660/ActaHortic.1991.288.44.

4. Horovitz, A.; Galil, J.; Zohary, D. Biological Flora of Israel. 6. *Anemone coronaria* L. Isr. J. Bot. 1975.

5. Laura, M.; Safaverdi, G.; Allavena, A. Androgenetic Plants of *Anemone coronaria* Derived through Anther Culture. Plant Breed. 2006, 125, 629–634. https://doi.org/10.1111/j.1439-0523.2006.01302.x.

6. Nissim, Y.; Jinggui, F.; Arik, S.; Neta, P.; Uri, L.; Avner, C. Phenotypic and Genotypic Analysis of a Commercial Cultivar and Wild Populations of *Anemone coronaria*. Euphytica 2004, 136, 51–62. https://doi.org/10.1023/B:EUPH.0000019520.19707.59.

7. Shamay, A.; Fang, J.; Pollak, N.; Cohen, A.; Yonash, N.; Lavi, U. Discovery of C-SNPs in *Anemone coronaria* L. and Assessment of Genetic Variation. Genet. Resour. Crop Evol. 2006, 53, 821–829. https://doi.org/10.1007/s10722-004-6377-5.

8. Wenzel, W.; Hemleben, V. A Comparative Study of Genomes in Angiosperms. Plant Syst. Evol. 1982, 139, 209–227.

9. Veselý, P.; Bureš, P.; Šmarda, P.; Pavlíček, T. Genome Size and DNA Base Composition of Geophytes: The Mirror of Phenology and Ecology? Ann. Bot. 2012, 109, 65–75. https://doi.org/10.1093/aob/mcr267.

10. Acquadro, A.; Magurno, F.; Portis, E.; Lanteri, S. DbEST-Derived Microsatellite Markers in Celery (Apium graveolens L. Var. Dulce). Mol. Ecol. Notes 2006, 6, 1080–1082. https://doi.org/10.1111/j.1471-8286.2006.01440.x.

11. Barchi, L.; Lanteri, S.; Portis, E.; Acquadro, A.; Valè, G.; Toppino, L.; Rotino, G.L. Identification of SNP and SSR Markers in Eggplant Using RAD Tag Sequencing. BMC Genom. 2011, 12, 304. https://doi.org/10.1186/1471-2164-12-304.

12. Lanteri, S.; Portis, E.; Acquadro, A.; Mauro, R.; Mauromicale, G. Morphology and SSR Fingerprinting of Newly Developed Cynara cardunculus Genotypes Exploitable as Ornamentals. Euphytica 2012, 184, 311–321. https://doi.org/10.1007/s10681-011-0509-8.

13. Gharsallah, C.; Ben Abdelkrim, A.; Fakhfakh, H.; Salhi-Hannachi, A.; Gorsane, F. SSR Marker-Assisted Screening of Commercial Tomato Genotypes under Salt Stress. Breed. Sci. 2016, 66, 823–830. https://doi.org/10.1270/jsbbs.16112.

14. Yang, Y.; He, R.; Zheng, J.; Hu, Z.; Wu, J.; Leng, P. Development of EST-SSR Markers and Association Mapping with Floral Traits in Syringa Oblata. BMC Plant Biol. 2020, 20, 436. https://doi.org/10.1186/s12870-020-02652-5.

15. Huang, C.-W.; Chu, P.-Y.; Wu, Y.-F.; Chan, W.-R.; Wang, Y.-H. Identification of Functional SSR Markers in Freshwater Ornamental Shrimps Neocaridina Denticulata Using Transcriptome Sequencing. Mar. Biotechnol. 2020, 22, 772–785. https://doi.org/10.1007/s10126-020-09979-y.

16. Li, Q.; Su, X.; Ma, H.; Du, K.; Yang, M.; Chen, B.; Fu, S.; Fu, T.; Xiang, C.; Zhao, Q.; et al. Development of Genic SSR Marker Resources from RNA-Seq Data in Camellia japonica and Their Application in the Genus Camellia. Sci. Rep. 2021, 11, 9919. https://doi.org/10.1038/s41598-021-89350-w.

17. Acquadro, A.; Portis, E.; Lee, D.; Donini, P.; Lanteri, S. Development and Characterization of Microsatellite Markers in Cynara cardunculus L. Genome 2005, 48, 217–225. https://doi.org/10.1139/g04-111.

18. Portis, E.; Scaglione, D.; Acquadro, A.; Mauromicale, G.; Mauro, R.; Knapp, S.J.; Lanteri, S. Genetic Mapping and Identification of QTL for Earliness in the Globe Artichoke/Cultivated Cardoon Complex. BMC Res. Notes 2012, 5, 252. https://doi.org/10.1186/1756-0500-5-252.

19. Feng, S.; He, R.; Lu, J.; Jiang, M.; Shen, X.; Jiang, Y.; Wang, Z.; Wang, H. Development of SSR Markers and Assessment of Genetic Diversity in Medicinal Chrysanthemum morifolium Cultivars. Front. Genet. 2016, 7, 113.

20. Aiello, D.; Ferradini, N.; Torelli, L.; Volpi, C.; Lambalk, J.; Russi, L.; Albertini, E. Evaluation of Cross-Species Transferability of SSR Markers in Foeniculum vulgare. Plants 2020, 9, 175. https://doi.org/10.3390/plants9020175.

21. Jaillon, O.; Aury, J.-M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The Grapevine Genome Sequence Suggests Ancestral Hexaploidization in Major Angiosperm Phyla. Nature 2007, 449, 463–467. https://doi.org/10.1038/nature06148.

22. Scaglione, D.; Reyes-Chin-Wo, S.; Acquadro, A.; Froenicke, L.; Portis, E.; Beitel, C.; Tirone, M.; Mauro, R.; Lo Monaco, A.; Mauromicale, G.; et al. The Genome Sequence of the Outbreeding Globe Artichoke Constructed de Novo Incorporating a Phase-Aware Low-Pass Sequencing Strategy of F1 Progeny. Sci. Rep. 2016, 6, 19427. https://doi.org/10.1038/srep19427.

23. Barchi, L.; Pietrella, M.; Venturini, L.; Minio, A.; Toppino, L.; Acquadro, A.; Andolfo, G.; Aprea, G.; Avanzato, C.; Bassolino, L.; et al. A Chromosome-Anchored Eggplant Genome Sequence Reveals Key Events in Solanaceae Evolution. Sci. Rep. 2019, 9, 11769. https://doi.org/10.1038/s41598-019-47985-w.

24. Wang, J.; Itgen, M.W.; Wang, H.; Gong, Y.; Jiang, J.; Li, J.; Sun, C.; Sessions, S.K.; Mueller, R.L. Gigantic Genomes Provide Empirical Tests of Transposable Element Dynamics Models. Genom. Proteom. Bioinform. 2021, 19, 123–139. https://doi.org/10.1016/j.gpb.2020.11.005.

25. Scheeff, E.D.; Bourne, P.E. Structural Evolution of the Protein Kinase-like Superfamily. PLoS Comput. Biol. 2005, 1, e49. https://doi.org/10.1371/journal.pcbi.0010049.

26. Shalaeva, D.N.; Cherepanov, D.A.; Galperin, M.Y.; Golovin, A.V.; Mulkidjanian, A.Y. Evolution of Cation Binding in the Active Sites of P-Loop Nucleoside Triphosphatases in Relation to the Basic Catalytic Mechanism. eLife 2018, 7, e37373. https://doi.org/10.7554/eLife.37373.

27. Di Nardo, G.; Gilardi, G. Natural Compounds as Pharmaceuticals: The Key Role of Cytochromes P450 Reactivity. Trends Biochem. Sci. 2020, 45, 511–525. https://doi.org/10.1016/j.tibs.2020.03.004.

28. Acquadro, A.; Barchi, L.; Portis, E.; Nourdine, M.; Carli, C.; Monge, S.; Valentino, D.; Lanteri, S. Whole Genome Resequencing of Four Italian Sweet Pepper Landraces Provides Insights on Sequence Variation in Genes of Agronomic Value. Sci. Rep. 2020, 10, 9189. https://doi.org/10.1038/s41598-020-66053-2.

29. Pavese, V.; Cavalet-Giorsa, E.; Barchi, L.; Acquadro, A.; Torello Marinoni, D.; Portis, E.; Lucas, S.J.; Botta, R. Whole-Genome Assembly of Corylus avellana cv "Tonda Gentile Delle Langhe" Using Linked-Reads (10X Genomics). G3 Genes|Genomes|Genetics 2021, 11, jkab152. https://doi.org/10.1093/g3journal/jkab152.

30. Hamarsheh, O.; Amro, A. Characterization of Simple Sequence Repeats (SSRs) from Phlebotomus papatasi (Diptera: Psychodidae) Expressed Sequence Tags (ESTs). Parasit Vectors 2011, 4, 189. https://doi.org/10.1186/1756-3305-4-189.

31. Liu, G.; Xie, Y.; Zhang, D.; Chen, H. Analysis of SSR Loci and Development of SSR Primers in Eucalyptus. J. For. Res. 2018, 29, 273–282. https://doi.org/10.1007/s11676-017-0434-3.

32. Manee, M.M.; Al-Shomrani, B.M.; Al-Fageeh, M.B. Genome-Wide Characterization of Simple Sequence Repeats in Palmae Genomes. Genes Genom. 2020, 42, 597–608. https://doi.org/10.1007/s13258-020-00924-w.

33. Ding, S.; Wang, S.; He, K.; Jiang, M.; Li, F. Large-Scale Analysis Reveals That the Genome Features of Simple Sequence Repeats Are Generally Conserved at the Family Level in Insects. BMC Genom. 2017, 18, 848. https://doi.org/10.1186/s12864-017-4234-0.

34. Chadha, S.; Gopalakrishna, T. Informativeness of Dinucleotide Repeat-Based Primers in Fungal Pathogen of Rice Magnaporthe grisea. Microbiol. Res. 2009, 164, 276–281. https://doi.org/10.1016/j.micres.2006.11.019.

35. Patil, P.G.; Singh, N.V.; Bohra, A.; Raghavendra, K.P.; Mane, R.; Mundewadikar, D.M.; Babu, K.D.; Sharma, J. Comprehensive Characterization and Validation of Chromosome-Specific Highly Polymorphic SSR Markers From Pomegranate (Punica granatum L.) cv. Tunisia Genome. Front. Plant Sci. 2021, 12, 337. https://doi.org/10.3389/fpls.2021.645055.

36. Sahu, K.K.; Chattopadhyay, D. Genome-Wide Sequence Variations between Wild and Cultivated Tomato Species Revisited by Whole Genome Sequence Mapping. BMC Genom. 2017, 18, 430. https://doi.org/10.1186/s12864-017-3822-3.

37. Portis, E.; Portis, F.; Valente, L.; Moglia, A.; Barchi, L.; Lanteri, S.; Acquadro, A. A Genome-Wide Survey of the Microsatellite Content of the Globe Artichoke Genome and the Development of a Web-Based Database. PLoS ONE 2016, 11, e0162841. https://doi.org/10.1371/journal.pone.0162841.

38. Portis, E.; Lanteri, S.; Barchi, L.; Portis, F.; Valente, L.; Toppino, L.; Rotino, G.L.; Acquadro, A. Comprehensive Characterization of Simple Sequence Repeats in Eggplant (Solanum melongena L.) Genome and Construction of a Web Resource. Front. Plant Sci. 2018, 9, 401. https://doi.org/10.3389/fpls.2018.00401.

39. An, J.; Yin, M.; Zhang, Q.; Gong, D.; Jia, X.; Guan, Y.; Hu, J. Genome Survey Sequencing of Luffa cylindrica L. and Microsatellite High Resolution Melting (SSR-HRM) Analysis for Genetic Relationship of Luffa Genotypes. Int. J. Mol. Sci. 2017, 18, 1942. https://doi.org/10.3390/ijms18091942.

40. Shi, J.; Huang, S.; Fu, D.; Yu, J.; Wang, X.; Hua, W.; Liu, S.; Liu, G.; Wang, H. Evolutionary Dynamics of Microsatellite Distribution in Plants: Insight from the Comparison of Sequenced Brassica, Arabidopsis and Other Angiosperm Species. PLoS ONE 2013, 8, e59988. https://doi.org/10.1371/journal.pone.0059988.

41. Cheng, J.; Zhao, Z.; Li, B.; Qin, C.; Wu, Z.; Trejo-Saavedra, D.L.; Luo, X.; Cui, J.; Rivera-Bustamante, R.F.; Li, S.; et al. A Comprehensive Characterization of Simple Sequence Repeats in Pepper Genomes Provides Valuable Resources for Marker Development in Capsicum. Sci. Rep. 2016, 6, 18919. https://doi.org/10.1038/srep18919.

42. Tóth, G.; Gáspári, Z.; Jurka, J. Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. Genome Res. 2000, 10, 967–981. https://doi.org/10.1101/gr.10.7.967.

43. Mun, J.H.; Kim, D.J.; Choi, H.K.; Gish, J.; Debellé, F.; Mudge, J.; Denny, R.; Endré, G.; Saurat, O.; Dudez, A.M.; et al. Distribution of Microsatellites in the Genome of Medicago Truncatula: A Resource of Genetic Markers That Integrate Genetic and Physical Maps. Genetics 2006, 172, 2541–2555.

44. Scaglione, D.; Acquadro, A.; Portis, E.; Taylor, C.A.; Lanteri, S.; Knapp, S.J. Ontology and Diversity of Transcript-Associated Microsatellites Mined from a Globe Artichoke EST Database. BMC Genom. 2009, 10, 454. https://doi.org/10.1186/1471-2164-10-454.

45. Cavagnaro, P.F.; Senalik, D.A.; Yang, L.; Simon, P.W.; Harkins, T.T.; Kodira, C.D.; Huang, S.; Weng, Y. Genome-Wide Characterization of Simple Sequence Repeats in Cucumber (Cucumis sativus L.). BMC Genom. 2010, 11, 569. https://doi.org/10.1186/1471-2164-11-569.

46. Morgante, M.; Hanafey, M.; Powell, W. Microsatellites Are Preferentially Associated with Nonrepetitive DNA in Plant Genomes. Nat. Genet. 2002, 30, 194–200. https://doi.org/10.1038/ng822.

47. Subramanian, S.; Mishra, R.K.; Singh, L. Genome-Wide Analysis of Microsatellite Repeats in Humans: Their Abundance and Density in Specific Genomic Regions. Genome Biol. 2003, 4, R13. https://doi.org/10.1186/gb-2003-4-2-r13.

48. Andersen, J.R.; Lübberstedt, T. Functional Markers in Plants. Trends Plant Sci. 2003, 8, 554–560. https://doi.org/10.1016/j.tplants.2003.09.010.

49. Li, Y.-C.; Korol, A.B.; Fahima, T.; Nevo, E. Microsatellites within Genes: Structure, Function, and Evolution. Mol. Biol. Evol. 2004, 21, 991–1007. https://doi.org/10.1093/molbev/msh073.

50. Brouwer, J.R.; Willemsen, R.; Oostra, B.A. Microsatellite Repeat Instability and Neurological Disease. Bioessays 2009, 31, 71–83. https://doi.org/10.1002/bies.080122.

51. Golubov, A.; Yao, Y.; Maheshwari, P.; Bilichak, A.; Boyko, A.; Belzile, F.; Kovalchuk, I. Microsatellite Instability in Arabidopsis Increases with Plant Development1[W][OA]. Plant Physiol. 2010, 154, 1415–1427. https://doi.org/10.1104/pp.110.162933.

52. Nelson, D.L.; Orr, H.T.; Warren, S.T. The Unstable Repeats—Three Evolving Faces of Neurological Disease. Neuron 2013, 77, 825–843. https://doi.org/10.1016/j.neuron.2013.02.022.

53. Vieira, D.D.S.S.; Emiliani, G.; Michelozzi, M.; Centritto, M.; Luro, F.; Morillon, R.; Loreto, F.; Gesteira, A.; Maserti, B. Polyploidization Alters Constitutive Content of Volatile Organic Compounds (VOC) and Improves Membrane Stability under Water Deficit in Volkamer Lemon (Citrus limonia Osb.) Leaves. Environ. Exp. Bot. 2016, 126, 1–9. https://doi.org/10.1016/j.envexpbot.2016.02.010.

54. Varshney, R.K.; Graner, A.; Sorrells, M.E. Genic Microsatellite Markers in Plants: Features and Applications. Trends Biotechnol. 2005, 23, 48–55. https://doi.org/10.1016/j.tibtech.2004.11.005.

55. Yu, J.-K.; Paik, H.; Choi, J.P.; Han, J.-H.; Choe, J.-K.; Hur, C.-G. Functional Domain Marker (FDM): An In Silico Demonstration in Solanaceae Using Simple Sequence Repeats (SSRs). Plant Mol. Biol. Rep. 2010, 28, 352–356. https://doi.org/10.1007/s11105-009-0154-8.

56. Kujur, A.; Bajaj, D.; Saxena, M.S.; Tripathi, S.; Upadhyaya, H.D.; Gowda, C.L.L.; Singh, S.; Jain, M.; Tyagi, A.K.; Parida, S.K. Functionally Relevant Microsatellite Markers from Chickpea Transcription Factor Genes for Efficient Genotyping Applications and Trait Association Mapping. DNA Res. 2013, 20, 355–374. https://doi.org/10.1093/dnares/dst015.

57. Liu, W.; Jia, X.; Liu, Z.; Zhang, Z.; Wang, Y.; Liu, Z.; Xie, W. Development and Characterization of Transcription Factor Gene-Derived Microsatellite (TFGM) Markers in Medicago truncatula and Their Transferability in Leguminous and Non-Leguminous Species. Molecules 2015, 20, 8759–8771. https://doi.org/10.3390/molecules20058759.

58. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and Next-Generation Sequencing: Computational Challenges and Solutions. Nat. Rev. Genet. 2011, 13, 36–46. https://doi.org/10.1038/nrg3117.

59. Wang, H.; Yang, B.; Wang, H.; Xiao, H. Impact of Different Numbers of Microsatellite Markers on Population Genetic Results Using SLAF-Seq Data for Rhododendron Species. Sci. Rep. 2021, 11, 8597. https://doi.org/10.1038/s41598-021-87945-x.

60. Stoll, A.; Harpke, D.; Schütte, C.; Stefanczyk, N.; Brandt, R.; Blattner, F.R.; Quandt, D. Development of Microsatellite Markers and Assembly of the Plastid Genome in Cistanthe long-iscapa (Montiaceae) Based on Low-Coverage Whole Genome Sequencing. PLoS ONE 2017, 12, e0178402. https://doi.org/10.1371/journal.pone.0178402.

61. Huang, Y.; Yin, Q.; Do, V.T.; Meng, K.; Chen, S.; Liao, B.; Fan, Q. Development and Characterization of Genomic Microsatellite Markers in the Tree Species, Rhodoleia championii, R. parvipetala, and R. forrestii (Hamamelidaceae). Mol. Biol. Rep. 2019, 46, 6547–6556. https://doi.org/10.1007/s11033-019-05106-w.

62. Li, D.; Long, C.; Pang, X.; Ning, D.; Wu, T.; Dong, M.; Han, X.; Guo, H. The Newly Developed Genomic-SSR Markers Uncover the Genetic Characteristics and Relationships of Olive Accessions. PeerJ 2020, 8, e8573. https://doi.org/10.7717/peerj.8573.

63. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. Bioinformatics 2015, 31, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

64. Bradnam, K.R.; Fass, J.N.; Alexandrov, A.; Baranay, P.; Bechner, M.; Birol, I.; Boisvert, S.; Chapman, J.A.; Chapuis, G.; Chikhi, R.; et al. Assemblathon 2: Evaluating de Novo Methods of Genome Assembly in Three Vertebrate Species. GigaScience 2013, 2, 2047-217X-2–10. https://doi.org/10.1186/2047-217X-2-10.

65. Smit, S.A.F.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013.

66. Campbell, M.S.; Law, M.; Holt, C.; Stein, J.C.; Moghe, G.D.; Hufnagel, D.E.; Lei, J.; Achawanantakun, R.; Jiao, D.; Lawrence, C.J.; et al. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiol. 2014, 164, 513–524. https://doi.org/10.1104/pp.113.230144.

67. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. Nucleic Acids Res. 2006, 34, W435–W439. https://doi.org/10.1093/nar/gkl200.

68. Bromberg, Y.; Rost, B. SNAP: Predict Effect of Non-Synonymous Polymorphisms on Function. Nucleic Acids Res. 2007, 35, 3823–3835. https://doi.org/10.1093/nar/gkm238.

69. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. Bioinformatics 2015, 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

70. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. Bioinformatics 2014, 30, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

71. Sigrist, C.J.A.; de Castro, E.; Cerutti, L.; Cuche, B.A.; Hulo, N.; Bridge, A.; Bougueleret, L.; Xenarios, I. New and Continuing Developments at PROSITE. Nucleic Acids Res. 2013, 41, D344–D347. https://doi.org/10.1093/nar/gks1067.

72. Mi, H.; Muruganujan, A.; Thomas, P.D. PANTHER in 2013: Modeling the Evolution of Gene Function, and Other Gene Attributes, in the Context of Phylogenetic Trees. Nucleic Acids Res. 2013, 41, D377–D386. https://doi.org/10.1093/nar/gks1118.

73. Lupas, A.; Van Dyke, M.; Stock, J. Predicting Coiled Coils from Protein Sequences. Science 1991, 252, 1162–1164. https://doi.org/10.1126/science.252.5009.1162.

74. Wu, C.H.; Nikolskaya, A.; Huang, H.; Yeh, L.L.; Natale, D.A.; Vinayaka, C.R.; Hu, Z.; Mazumder, R.; Kumar, S.; Kourtesis, P.; et al. PIRSF: Family Classification System at the Protein Information Resource. Nucleic Acids Res. 2004, 32, D112–D114. https://doi.org/10.1093/nar/gkh097.

75. Lima, T.; Auchincloss, A.H.; Coudert, E.; Keller, G.; Michoud, K.; Rivoire, C.; Bulliard, V.; de Castro, E.; Lachaize, C.; Baratin, D.; et al. HAMAP: A Database of Completely Sequenced Microbial Proteome Sets and Manually Curated Microbial Protein Families in UniProtKB/Swiss-Prot. Nucleic Acids Res. 2009, 37, D471–D478. https://doi.org/10.1093/nar/gkn661.

76. Punta, M.; Coggill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; et al. The Pfam Protein Families Database. Nucleic Acids Res. 2012, 40, D290–D301. https://doi.org/10.1093/nar/gkr1065.

77. de Lima Morais, D.A.; Fang, H.; Rackham, O.J.L.; Wilson, D.; Pethica, R.; Chothia, C.; Gough, J. SUPERFAMILY 1.75 including a Domain-Centric Gene Ontology Method. Nucleic Acids Res. 2011, 39, D427–D434. https://doi.org/10.1093/nar/gkq1130.

78. Bru, C.; Courcelle, E.; Carrère, S.; Beausse, Y.; Dalmar, S.; Kahn, D. The ProDom Database of Protein Domain Families: More Emphasis on 3D. Nucleic Acids Res. 2005, 33, D212–D215. https://doi.org/10.1093/nar/gki034.

79. Letunic, I.; Doerks, T.; Bork, P. SMART 7: Recent Updates to the Protein Domain Annotation Resource. Nucleic Acids Res. 2012, 40, D302–D305. https://doi.org/10.1093/nar/gkr931.

80. Lees, J.; Yeats, C.; Perkins, J.; Sillitoe, I.; Rentzsch, R.; Dessailly, B.H.; Orengo, C. Gene3D: A Domain-Based Resource for Comparative Genomics, Functional Annotation and Protein Network Analysis. Nucleic Acids Res. 2012, 40, D465–D471. https://doi.org/10.1093/nar/gkr1181.

81. Haft, D.H.; Selengut, J.D.; Richter, R.A.; Harkins, D.; Basu, M.K.; Beck, E. TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res. 2013, 41, D387–D395. https://doi.org/10.1093/nar/gks1234.

82. Araujo, F.A.; Barh, D.; Silva, A.; Guimarães, L.; Ramos, R.T.J. GO FEAT: A Rapid Web-Based Functional Annotation Tool for Genomic and Transcriptomic Data. Sci. Rep. 2018, 8, 1794. https://doi.org/10.1038/s41598-018-20211-9.

83. Kofler, R.; Schlötterer, C.; Lelley, T. SciRoKo: A New Tool for Whole Genome Microsatellite Search and Investigation. Bioinformatics 2007, 23, 1683–1685. https://doi.org/10.1093/bioinformatics/btm157.

84. Barkley, N.A.; Dean, R.E.; Pittman, R.N.; Wang, M.L.; Holbrook, C.C.; Pederson, G.A. Genetic Diversity of Cultivated and Wild-Type Peanuts Evaluated with M13-Tailed SSR Markers and Sequencing. Genet. Res. 2007, 89, 93–106. https://doi.org/10.1017/S0016672307008695.

85. Rosenthal, A.; Coutelle, O.; Craxton, M. Large-Scale Production of DNA Sequencing Templates by Microtitre Format PCR. Nucleic Acids Res. 1993, 21, 173–174.

86. Nei, M.; Li, W.H. Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. Proc. Natl. Acad. Sci. USA 1979, 76, 5269–5273.

87. Sneath, P.H.A.; Sokal, R.R. Numerical Taxonomy: The Principles and Practice of Numerical Classification; W. H. Freeman and Co.: 1973.

88. Rohlf, F.J. NTSYS-Pc: Numerical Taxonomy and Multivariate Analysis System; Exeter Software: 1988; ISBN 978-0-925031-00-6.

89. Hammer, O.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontol. Electron. 2001, 4, 9.

90. Anderson, J.A.; Churchill, G.A.; Autrique, J.E.; Tanksley, S.D.; Sorrells, M.E. Optimizing Parental Selection for Genetic Linkage Maps. Genome 1993, 36, 181–186. https://doi.org/10.1139/g93-024.

91. Mantel, N. The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res. 1967, 27, 209–220.

92. Wagner, H.W.; Sefc, K.M. IDENTITY 4.0. Centre for Applied Genetics; University of Agricultural Sciences: Vienna, Austria, 1999. -References-Scientific Research Publishing. Available online: https://www.scirp.org/%28S%28vtj3fa45qm1ean45vvffcz55%29%29/reference/referencespapers.aspx?referenceid=564391 (accessed on 8 March 2022).

93. Raymond, M.; Rousset, F. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. J. Hered. 1995, 86, 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573.

Chapter II

Microsatellite-based identification of double-haploid plants by androgenesis in *Anemone coronaria* L.

Matteo Martina[1], Lorenzo Barchi[1], Davide Gulino[1], Fabio Brusco[2], Mario Rabaglio[2], Alberto Acquadro[1], Ezio Portis[1,*] and Sergio Lanteri[1]

1. DISAFA, Plant Genetics, University of Turin, Italy
2. Biancheri Creazioni, 18033 Camporosso, Italy

Abstract

Anemone coronaria L. (2n=2x=16) is a perennial, allogamous, and highly heterozygous plant belonging to the Ranunculaceae family, marketed both as cut flower and garden plant. The production of F1 hybrids obtained by crossing highly homozygous inbred lines is hampered by the marked inbreeding depression characterizing the species. In order to overcome this limit, protocols have been developed to obtain double-haploid homozygous plants following diploidization of haploid plants obtained through in vitro androgenesis. However, during the androgenesis process a high frequency of spontaneous diploidization has been observed. This makes it difficult to discriminate between doubled haploid plants (DH) and plants originated from somatic anther tissues, as both have the same number of chromosomes (2x), and are thus not discriminable on the basis of their

DNA content through the flow-cytometric technique or following chromosome count.

Following a genome-wide survey, we previously developed a Microsatellite Markers Database (*AnCorDb*) in the species. By selecting microsatellites showing two alleles (heterozygous) in the anthers mother plant, we developed a fast and easy to apply technique suitable for selecting the DH plants in an early stage of plant development. DHs are characterized by the presence of one allele, whereas plants originated by somatic embryogenesis are characterized by the presence of two alleles.

Introduction

In recent years, in vitro androgenesis technology has grown as a powerful tool in plant biotechnology (Dwivedi et al., 2015; Ferrie & Möllers, 2011; Germanà, 2011). By exploiting the haploid state of the microspore and inducing genome doubling, genetically-stable DH tissues can be generated, avoiding several cycles of selfing and enabling the obtainment of homozygous lines in species suffering of inbreeding depression. These DHs have numerous applications, including, but not limited to, reducing undesirable events during breeding, such as self-incompatibility and hemizygosity, facilitating transgenesis and genome editing, and aiding in linkage map construction (Niazian & Shariatpanahi, 2020). The lines generated through this technique have potential applications as parental or test-cross lines in cross-pollinated species, as well as in recurrent selection schemes (Murovec & Bohanec, 2012; Rudolf Pilih et al., 2019). Additionally, in recent years, DHs have emerged as a promising breeding material for enhancing the hybrid breeding process through reverse breeding-based combining ability tests (Dezfouli et al., 2019). Indeed, hybrid breeding systems

have emerged as the primary approach for developing new crop varieties with economic significance. Doubled haploids are particularly important in producing genetically distinct and pure inbred lines, especially when working with genetically distant parent plants or species characterized by marked inbreeding depression hampering the production of inbred lines. Therefore, the progress of DH platforms has become a critical focus in hybrid breeding systems for many desirable crops (Begheyn et al., 2016; Chaikam et al., 2019; Liu et al., 2023).

Anemone coronaria L. (2n=2x =16, Ranunculaceae family), commonly known as poppy anemone, is a perennial plant that is both marketed as a garden plant and as a cut flower. The plant is allogamous and highly heterozygous. Selfing results in a marked inbreeding depression, which makes it challenging to develop highly homozygous inbred lines crossable for the production of F_1 hybrids. As a result, commercial cultivars are generated by selecting desirable traits from segregating populations obtained by crossing heterozygous genotypes. These cultivars are then propagated and marketed by rhizomes after a first growing season performed starting from seed material. At the same time, the development of successful protocols for obtaining androgenetic haploid plants has been achieved in the species under study, aiming at obtaining homozygous parental lines (Copetta & Laura, 2021). However, both flow cytometric analyses and chromosome counts in the meristematic cells of root apexes have revealed a high frequency of diploid plants, that might be originated both from spontaneous diploidization of haploid plants and from regeneration of the diploid anther tissues.

Recently, we generated the first draft for genome sequence of *A. coronaria*, and we reported on the massive microsatellite loci identification following its genome-wide survey (Martina et al., 2022). Based on the loci identified (401,822 perfect and 188,987 imperfect microsatellites motifs) we developed the "*Anemone coronaria* Microsatellite DataBase" (*AnCorDb*, www.anemone.unito.it) which represents the first on-line SSR loci resource available for the scientific community and breeders of poppy anemone and related species. Two of these SSR markers, which originated from heterozygous alleles in the cultivars from which the anthers were obtained, were then chosen. When applied to in vitro seedlings, these markers enable rapid and easy differentiation between diploid plants resulting from the spontaneous diploidization of a haploid embryo (one allele) and those originating from somatic embryogenesis of anther wall cells (two alleles). Here we report the first application of SSR markers for androgenetic origin discrimination in *A. coronaria* L., as a powerful tool to speed up the obtainment of F_1 hybrids in the species.

Materials and Methods

Plant Material

Androgenetic plants originated from two plants of *A. coronaria* cv. 'Magenta'(M4A) and cv. 'Rosa' (R2A) according to Copetta & Laura (2021), both obtained after several cycles of classical breeding selection, were acclimated in a cold greenhouse in Liguria (Italy), according to cultivation practices. After ploidy level identification (data not shown), the diploid 'Magenta' plant CP104 and the diploid 'Rosa' plant R17 were crossed, and a phenotypically highly stable and uniform F_1 population was obtained. Three individuals from the F_1 population were selected as representative

and were here screened for the assessment of the genetic uniformity of the population (Figure 1).

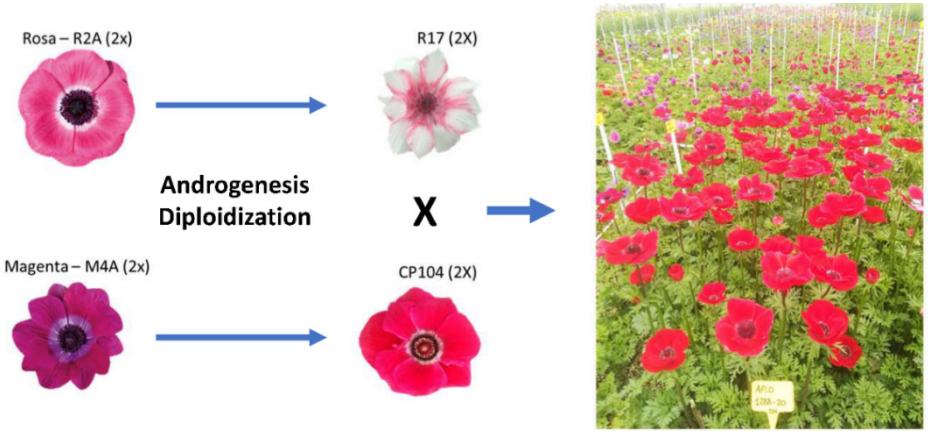


Figure 1. Schematic representation of the plant material. The two market varieties 'Rosa' and 'Magenta' were subjected to androgenesis and produced two diploid individuals (R17 and CP104), hypothetically originated through diploidization. The two plants were crossed, obtaining a uniform population that was molecularly screened for the assessment of its F1 hybrid nature.

Microsatellite screening

A set of nine microsatellites, previously applied for varietal fingerprinting by Martina et al. (2022), were applied (*AnCor33*, *AnCor36*, *AnCor59*, *Ancor60*, *Ancor68*, *Ancor83*, *Ancor102*, *Ancor107*, *Ancor177*). Genomic DNA was extracted by M4A, R2A, CP104, R17, and from three representatives of the F1 population from frozen leaves with the Plant DNA Kit (E.Z.N.A.®) following the manufacturer's instructions. DNA quality was assessed through the NanoDrop™ 2000 spectrophotometer, and the Qubit® 2.0 Fluorometer was used for DNA quantification. M13-Tailed Forward primers were designed for each microsatellite and applied in a three-primers unbalanced PCR reaction with a fluorescent-labelled M13

primer (Barkley et al., 2007). PCR was carried out in a final volume of 20 µL containing: 4 µL of 5× GoTaq Colorless Buffer (GoTaq® DNA Polymerase, Promega), 1 µL of MgCl_2 (25 mM), 0.4 µL of dNTPs (10 mM), 3 µL of DNA template (5 ng/µL), 1 µL of Reverse and M13-labeled primer (10 µM), 0.2 µL of Forward-M13 Tailed primer (10 µM), and 9.2 µL of ultrapure water. In each reaction, 1 µL of amplification product was pooled with other three products labelled with different fluorophores (FAM, VIC, NED, and PET) and purified using the PEG-precipitation method described by Rosenthal et al. (1993). Multiplex genotyping reactions were carried out in ABI PRISM® 310 according to the GeneScan® Reference Guide (Applied Biosystems™). Results were visualized using Peak Scanner™ Software v1.0 (Applied Biosystems™) and for each microsatellite the amplicons' length was scored.

Results and Discussion

The amplification of nine microsatellites was carried out in the material, highlighting differences between the genetic material tested by Martina et al., (2022) and the 'Magenta' and 'Rosa' accessions accessed in the present study (Table 1). As a certain degree of variability in the heterozygous level of the applied marker might be expected within the investigated cultivars, due to the segregation within crossing populations originating commercial varieties, such preliminary analysis is required to select the best markers to be used in the material under investigation.

Table 1. Differences between the expected alleles previously scored in the material fingerprinted by Martina et al. (2022) and the alleles scored in the present material (R2A and M4A). The markers that were evaluated as suitable for the screening of the androgenetic material are reported in bold.

	Expected (from Martina et al., 2022)			
	'Magenta'	'Rosa'	M4A	R2A
AnCor33	262	199-215	-	-
AnCor36	140	140-164	140	164
AnCor59	510-535	498-527	510	527
Ancor60	**404-435**	**515-553**	**404-435**	**553**
Ancor68	**404-435**	**515-553**	**435**	**515-553**
Ancor83	510-553	510-553	510	510
Ancor102	264-267	267	264-267	-
Ancor107	522	525-538	-	525-538
Ancor177	345-350	345-350	345	350

The applied nine SSR markers were selected as potentially providing amplification of two alleles in at least one of the genotypes. However, in both R2A and M4A *Ancor33* did not generate any amplicon, while *Ancor102* did not provide any amplicon in M4A, but two amplicons (alleles) in R2A (264bp and 267bp). A similar situation was observed in Ancor107, where no amplification was obtained in R2A, while two alleles were genotyped in M4A (at 525bp and 538bp). One allele of different size was detected in R2A and M4A following amplification of *AnCor36*, *AnCor59*, *Ancor83*, and *Ancor177*. Differently *Ancor60* generated one allele in M4A of 553bp but two of 494 and 435bp in R2A, while *Ancor68* two alleles of 515 and 553bp in M4A and on of 435bp in R2A. For the last two markers, the resulting amplicon sizes agreed with what previously reported by Martina et al. (2022) for the varieties ‘Magenta’ and ‘Rosa’. (Figure 2). The lower

level of heterozygosity detected in the plants derived from the two cultivars is attributable to selection performed by breeders.

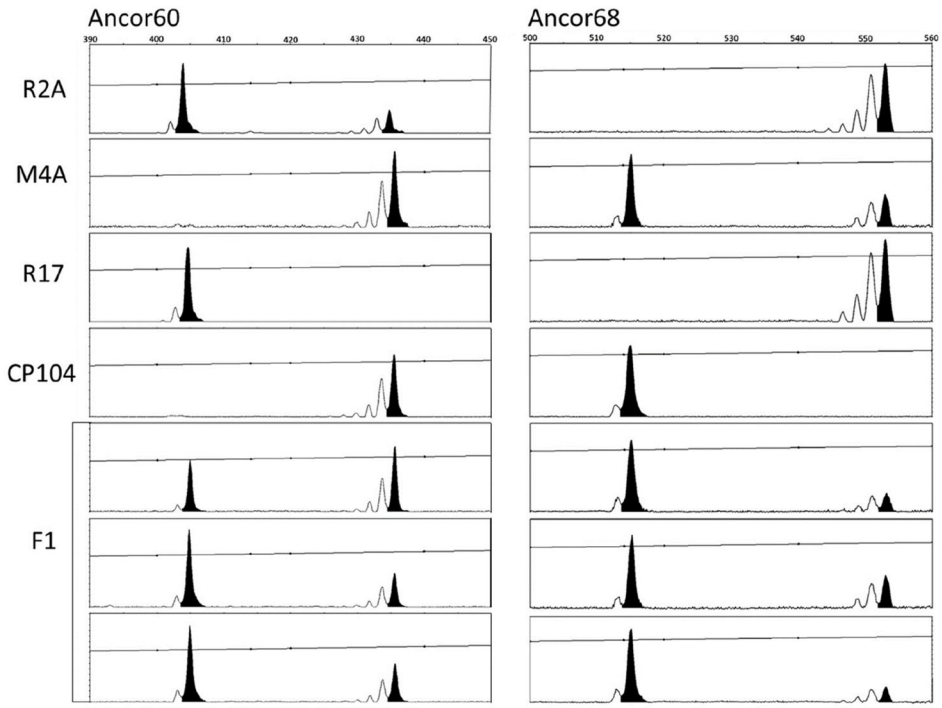


Figure 2. PeakScanner chromatograms obtained by analysing the data obtained by GeneScan elettrophoresis. The individuals' names are reported on the left side of the picture. Two peaks (alleles) are visible in both the applied markers (Ancor60 and Ancor68).

Based on the present results, *Ancor60* was chosen for the assessment of the androgenetic origin of the R17 diploids, showing the presence of just one allele of 404 or 435 bp, while Ancor68 for the androgenetic origin of the M4A diploids showing the presence of just one allele of 515 or 553bp. The SSR-assay on the F1 population highlighted a fixed heterozygous state in all the three analyzed samples, thus providing the first molecular assessment of the obtainment of F_1 hybrids in *A. coronaria* L. (Figure 2).

Conclusions

The present study demonstrates the successful application of two SSR markers for androgenetic origin discrimination in *A. coronaria*. The application of SSR markers offers a reliable and rapid method for identifying diploid plants that have arisen from spontaneous diploidization of a haploid embryo or from somatic embryogenesis of anther tissues. The ability to differentiate between these two origins of diploid plants has important implications for the breeding of F1 hybrids in *A. coronaria* L., a process that is difficult due to the marked inbreeding depression resulting from selfing. Furthermore, the molecular confirmation of the obtainment of the first F1 hybrid in poppy anemone based on SSR markers demonstrates the potential of this technique for accelerating the development of new cultivars in the species. Overall, this study highlights the potential applications and advantages of androgenesis technology and SSR markers in plant biotechnology and plant breeding.

References

Barkley, N. A., Dean, R. E., Pittman, R. N., Wang, M. L., Holbrook, C. C., & Pederson, G. A. (2007). Genetic diversity of cultivated and wild-type peanuts evaluated with M13-tailed SSR markers and sequencing. Genetical Research, 89(2), 93–106. https://doi.org/10.1017/S0016672307008695

Begheyn, R. F., Lübberstedt, T., & Studer, B. (2016). Haploid and Doubled Haploid Techniques in Perennial Ryegrass (Lolium perenne L.) to Advance Research and Breeding. Agronomy, 6(4), Article 4. https://doi.org/10.3390/agronomy6040060

Chaikam, V., Molenaar, W., Melchinger, A. E., & Boddupalli, P. M. (2019). Doubled haploid technology for line development in maize: Technical advances and prospects. Theoretical and Applied Genetics, 132(12), 3227–3243. https://doi.org/10.1007/s00122-019-03433-x

Copetta, A., & Laura, M. (2021). The Double-Layer Method to the Genesis of Androgenic Plants in *Anemone coronaria*. Methods in Molecular Biology (Clifton, N.J.), 2264, 187–196. https://doi.org/10.1007/978-1-0716-1201-9_13

Dwivedi, S. L., Britt, A. B., Tripathi, L., Sharma, S., Upadhyaya, H. D., & Ortiz, R. (2015). Haploids: Constraints and opportunities in plant breeding. Biotechnology Advances, 33(6 Pt 1), 812–829. https://doi.org/10.1016/j.biotechadv.2015.07.001

Ferrie, A. M. R., & Möllers, C. (2011). Haploids and doubled haploids in Brassica spp. For genetic and genomic research. Plant Cell, Tissue and Organ Culture (PCTOC), 104(3), 375–386. https://doi.org/10.1007/s11240-010-9831-4

Germanà, M. A. (2011). Anther culture for haploid and doubled haploid production. Plant Cell, Tissue and Organ Culture (PCTOC), 104(3), 283–300. https://doi.org/10.1007/s11240-010-9852-z

Liu, C., Wang, S., Liu, Y., Wang, M., Fan, E., Liu, C., Zhang, S., Yang, C., Wang, J., Sederoff, H. W., You, X., Chiang, V. L., Chen, S., Sederoff, R. R., & Qu, G. (2023). Exceptionally high genetic variance of the doubled haploid (DH) population of poplar. Journal of Forestry Research. https://doi.org/10.1007/s11676-023-01612-7

M. Dezfouli, P., Sedghi, M., E. Shariatpanahi, M., Niazian, M., & Alizadeh, B. (2019). Assessment of general and specific combining abilities in doubled haploid lines of rapeseed (Brassica napus L.). Industrial Crops and Products, 141, 111754. https://doi.org/10.1016/j.indcrop.2019.111754

Martina, M., Acquadro, A., Barchi, L., Gulino, D., Brusco, F., Rabaglio, M., Portis, F., Portis, E., & Lanteri, S. (2022). Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in *Anemone coronaria* L. International Journal of Molecular Sciences, 23(6), Article 6. https://doi.org/10.3390/ijms23063126

Murovec, J., & Bohanec, B. (2012). Haploids and Doubled Haploids in Plant Breeding. In I. Abdurakhmonov (Ed.), Plant Breeding. InTech. https://doi.org/10.5772/29982

Niazian, M., & Shariatpanahi, M. E. (2020). In vitro-based doubled haploid production: Recent improvements. Euphytica, 216(5), 69. https://doi.org/10.1007/s10681-020-02609-7

Rosenthal, A., Coutelle, O., & Craxton, M. (1993). Large-scale production of DNA sequencing templates by microtitre format PCR. Nucleic Acids Research, 21(1), 173–174.

Rudolf Pilih, K., Petkovsek, M., Jakše, J., Stajner, N., Murovec, J., & Bohanec, B. (2019). Proposal of a New Hybrid Breeding Method Based on Genotyping, Inter-Pollination, Phenotyping and Paternity Testing of Selected Elite F1 Hybrids. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.01111

Chapter III

frontiers | Frontiers in Plant Science

First genetic maps development and QTL mining in *Ranunculus asiaticus* L. through ddRADseq

Matteo Martina[1], Alberto Acquadro[1], Davide Gulino[1], Fabio Brusco[2], Mario Rabaglio[2], Ezio Portis[1,*] and Sergio Lanteri[1]

1. DISAFA, Plant Genetics, University of Turin, Italy
2. Biancheri Creazioni, 18033 Camporosso, Italy

Keywords: Ornamentals[1], ddRADseq[2], linkage maps[3], genetic markers[4], anthocyanins[5], QTLs[6]

Abstract

Persian Buttercup (*Ranunculus asiaticus* L.; 2x=2n=16; estimated genome size: 7.6Gb) is an ornamental and perennial crop native of Asia Minor and Mediterranean basin, marketed both as cut flower or potted plant. Currently new varieties are developed by selecting plants carrying desirable traits in segregating progenies obtained by controlled mating, which are propagated through rhizomes or micro-propagated in vitro. In order to escalate selection efficiency and respond to market requests, more knowledge of buttercup genetics would facilitate the identification of markers associated with loci and genes controlling key ornamental traits, opening the way for molecular assisted breeding programs. Reduced-representation sequencing (RRS) represents a powerful tool for plant genotyping, especially in case of large genomes such as the one of buttercup, and have been applied for the development of high-density genetic maps

in several species. We report on the development of the first molecular-genetic maps in *R. asiaticus* based on of a two-way pseudo-testcross strategy. A double digest restriction-site associated DNA (ddRAD) approach was applied for genotyping two F_1 mapping populations, whose female parents were a genotype of a so called 'ponpon' and of a 'double flower' varieties, while the common male parental ('Cipro') was a genotype producing a simple flower. The ddRAD generated a total of ~2Gb demultiplexed reads, resulting in an average of 8,3M reads per line. The sstacks pipeline was applied for the construction of a mock reference genome based on sequencing data, and SNP markers segregating in only one of the parents were retained for map construction by treating the F1 population as a backcross. The four parental maps (two of the female parents and two of the common male parent) were aligned with 106 common markers and 8 linkage groups were identified, corresponding to the haploid chromosome number of the species. An average of 586 markers were associated with each parental map, with a marker density ranging from 1 marker/cM to 4.4 markers/cM. The developed maps were used for QTL analysis for flower color, leading to the identification of major QTLs for purple pigmentation. These results contribute to dissect on the genetics of Persian buttercup, enabling the development of new approaches for future varietal development.

1 Introduction

Persian Buttercup (*Ranunculus asiaticus* L.; 2x=2n=16; estimated genome size: 7.6Gb - Goepfert, 1974) is an outcrossing ornamental and perennial crop native of Iran, Turkey and Greece. The species is marketed both as cut flowers and potted plant (De Hertogh, 1996). As reported by Berruto et al. (2019), *Ranuculus* counted for the 0,4% of the total turnover of cut flowers and foliage, with the highest production in Italy (132 million of stem) and 300-350 ha of cultivated surface. Due to its high level of heterozygosity and self-incompatibility, the crossing of selected genotypes originates highly segregant progenies. The development of new varieties is based on the selection within the progenies of plants carrying traits of interested, which are propagated through rhizomes or micro-propagated in vitro from meristematic apexes (Beruto et al., 2018). In order to escalate breeding efficiency and respond to market demand, some more knowledge of buttercup genetics is needed in order to enable the development of molecular breeding approaches.

High throughput DNA sequencing methodology (next generation sequencing; NGS) has rapidly evolved over the past 20 years (Slatko et al., 2018), and novel sequencing-associated protocols allow the access to thousands of genomic regions across the genome (Barchi et al., 2012; Víquez-Zamora et al., 2013; Scaglione et al., 2016; Vukosavljev et al., 2016; van Geest et al., 2017). The reduced-representation sequencing (RRS) represents one of most powerful tools for plant genotyping, especially in case of large genomes such as the one *R. asiaticus*. Indeed, RRS has been applied in the development of high-density genetic maps in many species

(Acquadro et al., 2017; Feng et al., 2018; Jin et al., 2019; Song et al., 2020; Toppino et al., 2020; Wang et al., 2020; Valentini et al., 2021).

Here we report on the application of a double digest restriction-site associated DNA (ddRAD) approach for genotyping two F1 mapping populations and on development of the first molecular-genetic maps in *R. asiaticus*. The strategy adopted was the two-way pseudo-testcross, previously exploited in a number of out-breeding species (Acquadro et al., 2009; Wu et al., 2010; Portis et al., 2012; Gartner et al., 2013; Torello Marinoni et al., 2018; Mariotti et al., 2020; Wang et al., 2022). The two F1 mapping populations shared a common male parental line, and four parental maps were developed (two of the female parents and two of the common male parent), then aligned on the basis of common SNPs markers. The resulting consensus map included eight linkage groups, corresponding to the haploid chromosome number of the species, and represents a background for future mapping of genes and QTLs and application of marker assisted breeding in the species. Besides, we performed QTL analysis for flower color, identifying a major locus affecting flower purple pigmentation.

2 Materials and methods

2.1 Mapping populations and DNA extraction

Two F_1 segregating populations were obtained by crossing the male parental genotype ‘Cipro’, producing white flowers with one row of petals typical of the wild-type (called ‘single’ flowers), with two female parental genotypes, of which one producing violet-greenish flowers with wavy margin (called ‘pon pon’ flowers) and one producing violet flowers with

multiple rows of petals (called ‘double’ flowers). The progenies were respectively named PON-PON and DOUBLE (Figure 1a).

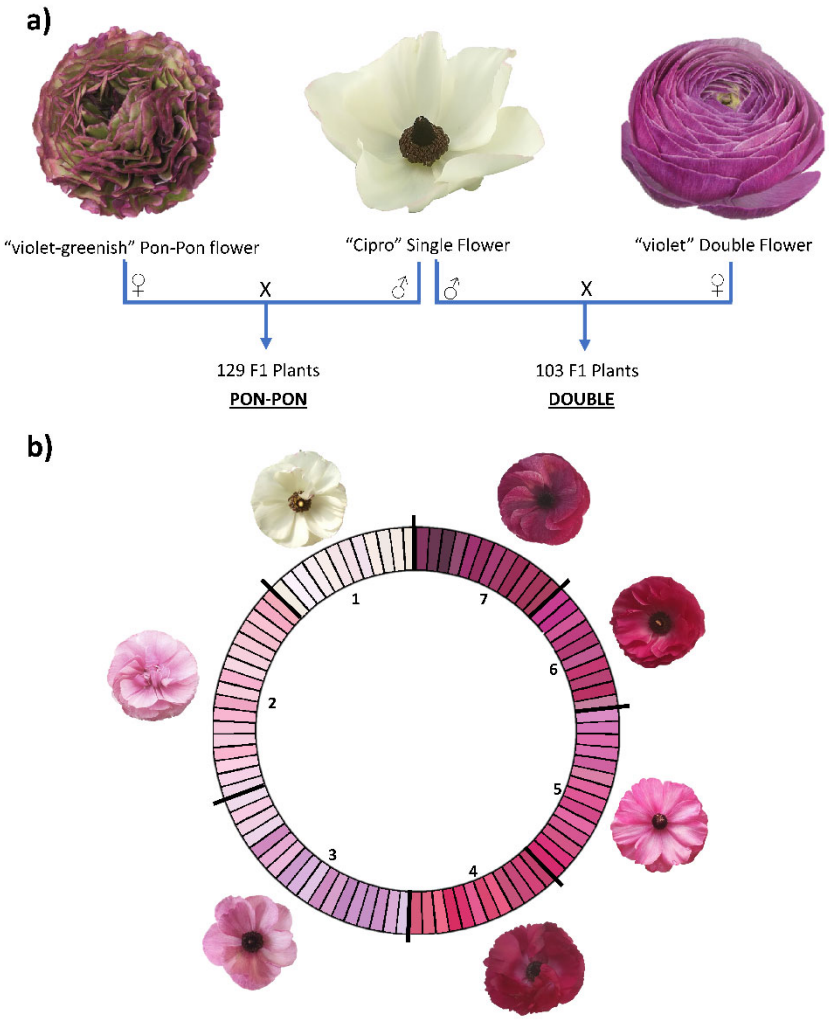


Figure 1. (A) Crossing scheme for the development of the two mapping populations (PON-PON and DOUBLE); (B) RHS-based color scale used for flower phenotyping. RHS classes were coded according to their intensity in 1-7 scale. 1: white; 2: light blue pink; 3: light violet; 4: medium purple red; 5: dark blue pink; 6: medium purple; 7: dark purple.

The parental genotypes were selected by Biancheri Creazioni (IM, Italy). The two F_1 progenies included 129 and 103 plants for the PON-PON and DOUBLE population respectively and were grown in a green-house located in Camporosso (43.794, 7.632, IM – Italy) for two seasons (2020 and 2021), adopting the cultivation techniques in use for commercial production. In 2020 plants originated from seeds, while in the 2021 were vegetatively propagated through rhizomes obtained at the end of the first season. Genomic DNA was extracted from frozen leaves with the Plant DNA Kit (E.Z.N.A.®) following the manufacturer's instructions. DNA quality was assessed through the NanoDrop™ 2000 spectrophotometer, and the Qubit® 2.0 Fluorometer was used for DNA quantification.

2.2 ddRAD-seq library preparation and sequencing

The ddRAD (double digest restriction-site associated DNA) libraries were produced using a custom protocol with minor modifications proposed by Peterson et al. (2012). Enzyme combination was selected by in silico analyzing the draft genome of the related species *Anemone coronaria* L. (Martina et al., 2022), with the aim of selecting a combination predicted to produce ~30k fragments across the provided genome. After fluorometric quantification, the genomic DNA was normalized for uniforming its concentration and 375ng (10ul, ~37ng/ul) were double digested with 2.4U of both PstI and EcoRI endonucleases (New England BioLabs) in 30µL reaction, supplemented with CutSmart Buffer and incubated at 37°C for 90', followed by 20' at 65°C'. Fragmented DNA was purified with 1.5 volumes of AMPureXP beads (Agencourt) and then ligated with 200U of T4 DNA ligase (New England BioLabs) to 2.5pmol of overhang barcoded adapter for rare cut sites and to 5pmol of overhang barcoded adapter for frequent

cut sites. Reactions were performed in 50µL volume at 23°C for 60' and at 20°C for 60', followed by 20' at 65°C. Samples were then pooled on multiplexing batches and bead purified as above. For each pool, targeted fragments distribution was collected on the BluePippin instrument (Sage Science Inc.), setting the range of 380 bp – 500 bp. Gel eluted fraction was then PCR amplified with indexed primers using Phusion High-Fidelity PCR Master Mix (New England BioLabs) in a final volume of 50µL, and subjected to the following thermal protocol: [95°C, 3'] - [95°C, 30" - 60°C, 30" - 72°C, 45"] x 12 cycles - [72°C, 2']. Products were purified with 1 volume of AMPureXP beads. The resulting libraries were checked with both Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) and Bioanalyzer DNA assay (Agilent technologies, Santa Clara, CA). Libraries were then PE sequenced with 150 cycles on NovaSeq 6000 instrument following the manufacturer's instructions (Illumina, San Diego, CA). A low coverage sequencing, and a draft assembly of the male parent ('Cipro') was obtained as reference for investigating data obtained with ddRADseq.

2.3 Sequence Analysis and Linkage Maps Development

Raw reads were demultiplexed using the process_radtags pipeline included in Stacks v2.53 (Catchen et al., 2013). Assembly of the short reads of each sample into matching stacks was performed using the ustacks utility included in Stacks v2.53. A set of consensus loci from all the analyzed samples (catalog) was identified by the cstacks pipeline, and each sample was matched against the catalog using sstack and tsv2bam utilities. Sequences for each genotype were mapped to the catalog file using the Burrows-Wheeler Aligner (BWA, v0.7.17) program and the 'mem' command with the default parameters (Li and Durbin, 2009). BAM files

were processed and used for SNP calling using Samtools mpileup (v1.6 - Danecek et al., 2021) with default parameters except for the minimum mapping quality (Q = 20). Markers were named according to the catalog sequence in which they were identified. Each name starts with an S, followed by the catalog sequence number and the position of the SNP in the sequence (i.e S0004445_269).

Independent framework linkage maps were constructed for each parent (male and female) from each progeny-based dataset (PON-PON and DOUBLE), using the double pseudo-testcross mapping strategy (Plomion and Durel, 1996) and JoinMap v4.0 (Van Ooijen, 2006). Four separate datasets were assembled: (i) 'Violet-greenish' - 'PP -'; (ii) Cipro I - 'C-I' -; (iii) 'Violet' - 'Db' - and (iv) Cipro II - 'C-II'. Only SNP markers in testcross configuration (expected segregation ratio of 1:1) were included in the datasets: maternal testcross markers segregating only in 'PP' or 'Db', and paternal testcross markers segregating only in 'C-I' or 'C-II'.

Genotyping data were quality filtered by removing not segregating markers, markers with >20% missing values, and skewed markers. The similarity of the loci option of JoinMap was used to identify perfectly identical markers (similarity value = 1.000), expected to map to exactly the same position. To reduce the load of calculation effort, only one representative of each group of identical loci was used for mapping. Goodness-of-fit between observed and expected segregation ratios was assessed using the $\chi 2$ test. Markers fitting a Mendelian pattern closely associated with a $\chi 2$ value $\chi 2\ \alpha = 0.1$ or with only a minor deviation ($\chi 2\ \alpha = 0.1 < \chi 2\ \chi 2\ \alpha = 0.01$) were used for map construction, provided that their

inclusion did not alter the local marker order. Loci suffering from significant segregation distortion ($\chi2 > \chi2\ \alpha = 0.01$) were excluded.

For all maps, LGs were established based on a threshold logarithm of odds (LOD) ratio >4. To determine marker order within a linkage group (LG), the JoinMap parameters were set at Rec = 0.40, LOD = 1.0 and Jump = 5. Map distances were converted to centiMorgans (cM) using the Kosambi mapping function (Kosambi, 1943). Linkage maps were drawn using MapChart 2.2 software (Voorrips, 2002), and markers deviating in their segregation only marginally from the expected Mendelian ratio are presented with one ($\chi2\ \alpha = 0.1 < \chi2\ \chi2\ \alpha = 0.05$) or two ($\chi2\ \alpha = 0.05 < \chi2\ \chi2\ \alpha = 0.01$) asterisks. LG of the female and male maps were respectively named PP_01 to PP_08 and CI_01 to CI_08 for the ones present in the PON-PON dataset, and Db_01 to Db_08 and CII_01 to CII_08 for the ones present in the DOUBLE dataset (Supplemental Figure 1).

2.4 Flower color assessment

The first blooming flower of market quality of each genotype was phenotyped in both seasons (2020 and 2021), as representative of the genetic potential of each individual. The main color of the internal (Main Interior Color - MiC) and external (Main Exterior Color - MeC) layer of the petals were scored following RHS classification (Figure 1b).

2.5 Statistical Analyses and QTL Detection

Statistical analyses were performed with R software (Team, 2016). A conventional analysis of variance was applied to estimate genotype and environment effects based on the linear model $Yij = \mu + gi + bj + eij$, where μ, g, b and e represent the mean, the genotypic effect, the block effect and the error respectively. Correlations between traits were estimated using the

Spearman coefficient, and normality, kurtosis and skewness were assessed with the Shapiro–Wilks test ($\alpha = 0.05$). QTL detection was performed by considering each season independently and was based on the newly developed map using MQM mapping, as implemented in MapQTL v5 software (Van Ooijen, 2004). QTLs were initially identified using interval mapping. LOD thresholds for declaring a QTL to be significant at the 5% genome-wide probability level were established empirically by applying 1000 permutations per trait (Doerge and Churchill, 1996). After permutation analysis, linked marker per putative QTL was treated as a co-factor in the approximate multiple QTL model. Co-factor selection and MQM analysis were repeated until no new QTL could be identified. Additive effect, as well as the percentage of the phenotypic variation (PVE) explained by each QTL, were obtained from the final multiple QTL model. Individual QTLs were prefixed by the parental line abbreviation ('PP', 'C-I', 'C-II', or 'Db'), followed by a trait abbreviation, the relevant chromosome designation, while the season was followed by a letter when there were more than one QTL for the season. Confidence interval of the QTL was calculated by considering 0.5Mb upstream and downstream the marker identified at the QTL.

3 Results and Discussion

3.1 Sequencing and Linkage Map Construction

The ddRAD approach generated a total of ~2Gb demultiplexed reads, and an average of 8,3M reads per genotype (detailed info about the raw data per individual can be found in Supplemental Table 1). Cleaned reads with quality scores >30 were mapped against the catalog obtained through the stacks pipeline (see Materials and Methods). Before filtering, ~578k

polimorphisms where identified. However, after filtering at DP>15, this number was lowered to ~22k SNPs and 790 indels. Only markers heterozygous in a single parent were retained for map construction, by treating the F_1 population as a backcross. For each of the parental lines, skewed markers, showing highly significant distortion from the expected 1:1 ratio, or markers with identical segregation patterns were excluded from further map construction steps (Supplemental Table 1).

Table 1. Statistics of the four linkage maps.

		LG 1	LG 2	LG 3	LG 4	LG 5	LG 6	LG 7	LG 8	Average	Total
PonPon (PP)	*Size*	112.0	130.7	97.4	113.4	102.9	80.5	110.8	82.2	103.7	829.8
	N° of markers	93	83	63	82	74	58	37	78	71	568
	Marker Density	1.2	1.6	1.5	1.4	1.4	1.4	3.0	1.1	1.6	
	Gaps (> 5cM)	3	3	4	1	2	2	8	0	3	23
Cipro I (CI)	*Size*	199.1	201.4	189.5	173.6	156.1	123.7	147.8	114.6	163.2	1305.8
	N° of markers	102	75	64	72	74	55	49	65	70	556
	Marker Density	2.0	2.7	3.0	2.4	2.1	2.2	3.0	1.8	2.4	
	Gaps (> 5cM)	5	9	13	6	4	9	11	3	8	60
Cipro II (CII)	*Size*	226.7	160.2	116.6	132.6	151.4	149.0	127.0	118.8	147.8	1182.2
	N° of markers	152	98	113	71	49	73	50	74	85	680
	Marker Density	1.5	1.6	1.0	1.9	3.1	2.0	2.5	1.6	1.9	
	Gaps (> 5cM)	6	4	3	3	9	6	9	2	5	42
Double (Db)	*Size*	169.1	145.6	102.5	161.4	151.1	106.9	105.4	117.6	132.5	1059.7
	N° of markers	68	33	65	112	68	74	46	74	68	540
	Marker Density	2.5	4.4	1.6	1.4	2.2	1.4	2.3	1.6	2.2	
	Gaps (> 5cM)	9	12	4	4	9	2	5	6	6	51

Overall, 3,376 single nucleotide polymorphisms (SNPs) were identified as segregating in the PONPON population, of which 1,607 and 1,769 were used as input in JoinMap for the development of Cipro (C-I) and PonPon (PP) maps, respectively. In C-I, 732 markers were excluded by the JoinMap pipeline being co-segregating (i.e. markers showing the same segregation pattern), leading to an overall number of 875 markers used for map construction. Analogously, in PP, 796 SNPs were co-segregating, while 973 were used for map construction. A further stringent selection

was applied by considering only markers grouped at LOD >4, and a set of 568 SNPs in PP, and 556 in C-I were identified as suitable for map development. The generated PP map spanned 829 cM over 8 LGs (corresponding to the haploid chromosome number of the species) ranging from 80.5 to 130.7 cM, with a marker density between 1.1 and 3 cM, and 23 gaps bigger than 5 cM. The LG7 harbored the highest number of gaps (8), while only one gap was present in LG4. The C-I map spanned 1305.8 cM across 8 LGs, ranging between 123.7 and 201.4cM and with a marker density between 1.8 and 3 cM. In this map a higher number of gaps (>5 cM) was present (60), with LG3 and LG7 including 13 and 11 gaps respectively, and LG8 harboring the lower number of gaps (3). The LGs of the two maps were aligned on the basis of 35 common markers, as shown in Figure 2.

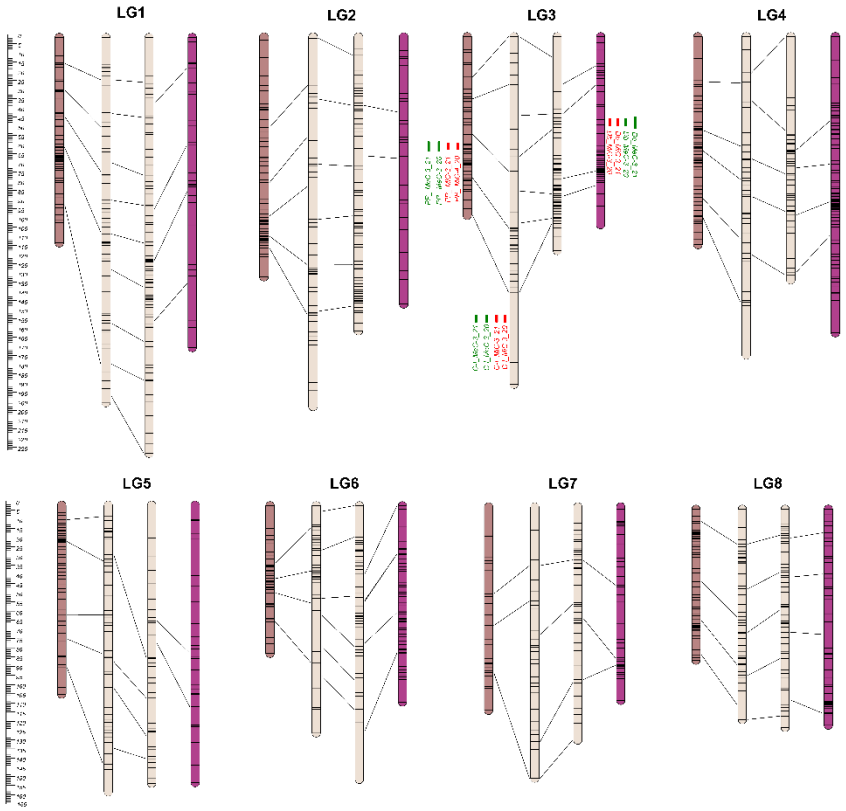


Figure 2. From the left to the right: PP, C-I, C-II, and Db Linkage Groups. Eight LGs (corresponding to the haploid chromosome number of the species) were identified within each parental map. C-I map was the bigger one, covering 1,305.8 cM, while the PP map was the smaller one with 829.8 cM (Table 1). QTL regions are reported as green (MeC) and red (MiC) bars on the right side of the associated LGs.

Overall, 4,098 SNPs were identified in the DOUBLE population, of which 2,344 and 1,754 segregating in Cipro (C-II) and Double (Db) maps respectively. In C-II, 1,211 co-segregating markers were discarded, and 1,113 SNPs used for map construction. For Db, 919 SNPs were co-segregating, leading to 835 mappable markers. A further stringent selection was applied by considering only markers grouped at LOD >4,

identifying 680 loci for C-II, and 540 for Db maps development. A genetic map with an overall dimension of 829 cM was generated for Db, identifying 8 LGs ranging from 102.5 and 169.1 cM with a marker density between 1.4 and 4.4, and 51 gaps bigger than 5 cM. The highest number of gaps (12) was scored on LG2, while only two gaps were present on LG6. The 8 LGs developed for the C-II map covered 1182.2 cM, ranging between 116.6 and 226.7 cM with a marker density between 1 and 3.1 cM. Forty-two gaps were identified across LGs, with LG5 and LG7 presenting 9 gaps, and LG8 harboring only 2 gaps. Each LGs of the two maps were aligned using 27 common markers, as shown in Figure 2 and Table 1.
The 8 LGs of the two maps of the male parental genotype (C-I and C-II) were aligned by means of 44 markers. Furthermore, the four maps were aligned on the basis of 106 common markers of which 7 markers were common across PP, C-I and C-II, while one was common among the four maps. A clustering of markers was observed in some LGs, and in most cases was located in the central part of a LG. This clustering might correspond to the centromeric regions (Keim et al., 1990; Reiter et al., 1992; Tanksley et al., 1992; Vallejos et al., 1992; Ott et al., 2011), in which a reduced recombination usually takes place (Vincenten et al., 2015; Zafar et al., 2017; Acquadro et al., 2020).

An average number of 586 markers was included in each parental map, with a marker density ranging from 1 marker/cM to 4.4 markers/cM. Assuming an expected genome size of 7.6Gb (Goepfert, 1974), the average mapping length of ~1093 cM was used for estimating the physical dimension of a cM, which corresponds to a physical distance of about 7Mb. The average LG dimension ranged from 829.8 cM (PP) to 1305.8

cM (C-I). Both the male parent maps (C-I and C-II) were larger than the ones of the female parents (PP and Db). As previously reported (Lenormand and Dutheil, 2005), differences between paternal and maternal linkage maps could be associated with heterochiasmy (i.e., the presence of different crossover frequencies in male and female meiosis). Differential recombination rates between the sexes have been reported in several plant species and attributed to processes such as sexually antagonistic selection acting on coding and regulatory genetic elements, female meiotic drive as well as selection during the haploid phase of the life cycle (Sardell and Kirkpatrick, 2020). The latter has been associated to the reproductive system (selfing or outcrossing) characterizing the species (Giraut et al., 2011; Mariotti et al., 2020), although even in closely related species sharing the same reproduction pattern, such as *Sinapis alba* (Nelson et al., 2013), *Brassica oleracea* (Kearsey et al., 1996) and i (Kelly et al., 1997), contrasting results have been reported.

3.2 Phenotypic Variation, traits Correlation and QTL identification

Table 2 reports a summary of the phenotypic performance for main color of the internal (MiC) and external (MeC) layer of the petals in the parental genotypes and F1 population, together with skewness and kurtosis of the traits. The male parent ‘Cipro’ did not present anthocyanin content in both petal layers. Conversely, the ‘Double’ genotype produced violet flowers, and PonPon violet- greenish petals. While the MiC trait appears to be stable in the two seasons in the parental lines, MeC showed a certain level of variability in the PP parent. In the F1 populations transgressive

segregation was observed in respect to both maternal parents, except for MeC in the DOUBLE population.

Table 2. List of the traits analyzed and their code, means, standard deviations (SD), overall statistics, coefficients of variation (cv), and transgressive genotypes.

Population	PON-PON			
Trait	*MiC_20*	*MiC_21*	*MeC_20*	*MeC_21*
Trait code	*PP_MiC_20*	*PP_MiC_21*	*PP_MeC_20*	*PP_MeC_21*
PP mean	6	6	7	4
C mean	1	1	1	1
F1 mean	2.7	3.4	3.0	3.7
± SD	1.7	2.0	1.9	2.2
cv	0.6	0.6	0.6	0.6
Shapiro-Wilks	0.802	0.807	0.796	0.783
Skewness	-1.131	-1.610	-1.419	-1.686
SE	0.431	0.427	0.431	0.427
Kurtosis	0.482	-0.112	0.189	-0.217
SE	0.217	0.215	0.217	0.215
Transgressive respect PP	4	45	0	58
Transgressive respect C	0	0	0	0

Population	DOUBLE			
Trait	*MiC_20*	*MiC_21*	*MeC_20*	*MeC_21*
Trait code	*Db_MiC_20*	*Db_MiC_21*	*Db_MeC_20*	*Db_MeC_21*
PP mean	5	5	7	7
C mean	1	1	1	1
F1 mean	3.43	3.65	4.00	4.18
± SD	1.74	1.75	2.01	2.09
cv	0.51	0.48	0.50	0.50
Shapiro-Wilks	0.85	0.85	0.79	0.77
Skewness	-0.18	-0.42	-0.53	-0.63
SE	0.24	0.24	0.24	0.24
Kurtosis	-1.20	-1.01	-1.32	-1.26
SE	0.48	0.47	0.48	0.47
Transgressive respect PP	17	15	0	0
Transgressive respect C	0	0	0	0

Significant inter-trait correlations were detected within and across growing seasons (Table 3), albeit detecting a certain level of variability introduced by the environment. The least correlated traits were DB_MiC_20 and DB_MeC_21 (+0.70), while the most highly correlated were PP_MiC_21 and PP_MeC_21 (+0.94).

Table 3. Inter-trait Spearman correlations assessed in the mapping populations.

	PP_MiC_20	PP_MiC_21	PP_MeC_20	PP_MeC_21
PP_MiC_20	-			
PP_MiC_21	0.76	-		
PP_MeC_20	0.90	0.81	-	
PP_MeC_21	0.77	0.94	0.84	-

	Db_MiC_20	Db_MiC_21	Db_MeC_20	Db_MeC_21
Db_MiC_20	-			
Db_MiC_21	0.82	-		
Db_MeC_20	0.88	0.81	-	
Db_MeC_21	0.70	0.88	0.86	-

In Season 2020, the plants originated from seeds were less vigorous in respect to plants originated from rhizomes in season 2021. The flower color was limitedly affected by the environment traits in the two seasons. However, in Season 2020 the PP population distribution appeared to be leptokurtic, while in the Season 2021 the distribution was platykurtic. Such differences did not appear in the DOUBLE population. Transgressive genotypes always deviate towards the more pigmented parent (PP and Db), with petals exceeding the parental lines in pigment accumulation. Furthermore, in some cases the petal color appeared to be not uniform and presented mottling, suggesting the involvement of regulatory elements and

epigenetic mechanisms, as reported in literature (Koseki et al., 2005; Le Maitre et al., 2019; Wang et al., 2019; Zheng et al., 2021).

The genetic basis of variation in quantitative traits can traditionally be resolved by the QTL approach, which partitions the variation into distinct genomic regions defined by a linkage map (Paterson et al., 1988). One of the most important issues to determine is the stability of the trait and how stable the expression of the various loci is by repeating the phenotypic evaluation over time and/or space. QTL detection was performed considering each season independently and was based on the newly developed map using MQM mapping (see Materials and Methods). LOD score, percentage of variance explained (PVE), and confidence interval (CI) related to QTLs, are described in Table 3. QTL analyses on all traits and environments yielded a total of 12 major (PVE values >10), located on LG3 (see Figure 4 and Table 4). In both female parents (PP and Db), 8 major QTLs were located in the same scaffold (S0004445) for both MiC and MeC. On the other hand, QTLs were identified only in the male parental map from the PON-PON population (C-I), both for MiC and MeC. Our results highlight a very high correlation between the external (MeC) and internal (MiC) pigmentation of the flower (Table 3), suggesting that the same genomic regions affect both traits. Thus, MiC and MeC appear to be controlled by common major QTLs located on LG3, justifying a PVE ranging from 31.7 to 84%, with a LOD ranging from 11.23 to 50.6.

Table 4. QTL detected in the mapping population. For each trait the location (LG and position), the associated marker, the LOD, confidence interval (CI), percentage of explained variation (PVE), additive effect (Add.), and the genome-wide thresholds (GW) at p = 0.05 (as determined from 1000 permutations) are indicated.

Map	Trait	Year	Name	Group	Pos.	Locus	LOD	CI	PVE	Add.	GW
PP	MiC	20	*PP_MiC-3_20*	LG3	60.8	S0004445_269	21.74	58.8-61.3	48.3	2.5	18
		21	*PP_MiC-3_21*	LG3	60.8	S0004445_269	38.66	58.8-61.3	67.4	3.4	23
	MeC	20	*PP_MeC-3_20*	LG3	60.8	S0004445_269	29.2	58.8-61.3	49.4	2.7	19.4
		21	*PP_MeC-3_21*	LG3	60.8	S0004445_269	46.1	58.8-61.3	71.9	3.9	25.8
C-I	MiC	20	*C-I_MiC-3_20*	LG3	155.9	S0038600_63	23.32	155.38-156.38	57.9	-2.7	18
		21	*C-I_MiC-3_21*	LG3	155.9	S0038600_63	43.45	155.38-156.38	79.3	-3.7	23.1
	MeC	20	*C-I_MeC-3_20*	LG3	155.9	S0038600_63	28.81	155.38-156.38	65.7	-3.1	19.6
		21	*C-I_MeC-3_21*	LG3	155.9	S0038600_63	50.6	155.38-156.38	84	-4.1	26
Db	MiC	20	*Db_MiC-3_20*	LG3	48.3	S0004445_68	11.23	47.565-48.796	31.7	-2.0	9.3
		21	*Db_MiC-3_21*	LG3	48.3	S0004445_68	15.62	47.565-48.796	45.1	-2.4	10.1
	MeC	20	*Db_MeC-3_20*	LG3	48.3	S0004445_68	22.37	47.565-48.796	59	-3.2	14.8
		21	*Db_MeC-3_21*	LG3	48.3	S0004445_68	21.41	47.565-48.796	58.9	-3.3	18.7

To the best of our knowledge, this is the first report on QTLs detection in Persian buttercup. The identified QTLsshould be further on investigated and once validated on a large collection of breeding material, the developed molecular markers might be applied for molecular assisted breeding.

4 Conclusions

Genome size varies greatly across the flowering plants and has played an important role in shaping their evolution. *Ranunculus asiaticus* harbors a very large genome size of about 7Gb. Mainly as a consequence of its large genome size, the development of genomic tools in the species has been to date very limited. We have applied for the first time in Persian buttercup, a reduced -representation sequencing (RRS) approach with the goal to develop its first genetic maps. Based on the two-way pseudo test cross strategy in two F_1 populations, sharing the same male parent, we

constructed maps which were used for positioning the first QTLs affecting the flower anthocyanin pigmentation. Our results lay the foundations for future genetic and genomic studies and provide a framework for implementing more targeted breeding programs in *R. asiaticus*.

5 Data Availability Statement

Sequencing data used in this study are openly available in the NCBI database (PRJNA876570).

6 Author Contributions

Conceptualization: S.L., E.P., A.A., M.R., F.B. and M.M.; methodology: S.L., E.P., A.A., M.R. and M.M.; software: A.A., E.P. and M.M.; validation, M.M., D.G. and E.P.; formal analysis: M.M., D.G., E.P. and A.A.; investigation: E.P., M.M., A.A.; resources: M.R., F.B. and D.G.; data curation: M.M., and E.P.; writing—original draft preparation: M.M., A.A. and E.P.; writing—review and editing: S.L.; visualization: M.R., F.B. and D.G..; supervision, E.P. and S.L. All authors have read and agreed to the published version of the manuscript.

7 Funding

This research was partially funded by Biancheri Creazioni (Camporosso, Italy).

8 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9 Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

10 Supplementary materials

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1009206/full#supplementary-material

Supplementary Figure 1 | Linkage Groups obtained after analysis, together with markers name.

Supplementary Figure 2 | Distribution of the investigated traits in the two populations.

11 References

A. A. De Hertogh (1996). Holland bulb forcer's guide. 5th ed. Alkemade Printing Lisse, The Netherlands.

Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., et al. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. PLoS ONE 12, e0180774. doi: 10.1371/journal.pone.0180774.

Acquadro, A., Barchi, L., Portis, E., Nourdine, M., Carli, C., Monge, S., et al. (2020). Whole genome resequencing of four Italian sweet pepper landraces provides insights on sequence variation in genes of agronomic value. Sci Rep 10, 9189. doi: 10.1038/s41598-020-66053-2.

Acquadro, A., Lanteri, S., Scaglione, D., Arens, P., Vosman, B., and Portis, E. (2009). Genetic mapping and annotation of genomic microsatellites isolated from globe artichoke. Theor Appl Genet 118, 1573–1587. doi: 10.1007/s00122-009-1005-6.

Barchi, L., Lanteri, S., Portis, E., Valè, G., Volante, A., Pulcini, L., et al. (2012). A RAD Tag Derived Marker Based Eggplant Linkage Map and the Location of QTLs Determining Anthocyanin Pigmentation. PLoS ONE 7, e43740. doi: 10.1371/journal.pone.0043740.

Beruto, M., Martini, P. and Viglione, S. (2019). *Ranuculus asiaticus*: from research to production. Acta Hortic. 1237, 117-128 doi: 10.17660/ActaHortic.2019.1237.16

Beruto, M., Rabaglio, M., Viglione, S., Van Labeke, M.-C., and Dhooghe, E. (2018). "*Ranuculus*," in Ornamental Crops Handbook of Plant Breeding., ed. J. Van Huylenbroeck (Cham: Springer International Publishing), 649–671. doi: 10.1007/978-3-319-90698-0_25.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology 22, 3124–3140. doi: 10.1111/mec.12354.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience 10, giab008. doi: 10.1093/gigascience/giab008.

Doerge, R. W., and Churchill, G. A. (1996). Permutation tests for multiple loci affecting a quantitative character. Genetics 142, 285–294. doi: 10.1093/genetics/142.1.285.

Feng, X., Yu, X., Fu, B., Wang, X., Liu, H., Pang, M., et al. (2018). A high-resolution genetic linkage map and QTL fine mapping for growth-related traits and sex in the Yangtze River common carp (Cyprinus carpio haematopterus). BMC Genomics 19, 230. doi: 10.1186/s12864-018-4613-1.

Gartner, G. A. L., Mccouch, S., and Moncada, M. D. P. (2013). A genetic map of an interspecific diploid pseudo testcross population of coffee. Euphytica. doi: 10.1007/s10681-013-0926-y.

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O. C., and Mézard, C. (2011). Genome-Wide Crossover Distribution in Arabidopsis thaliana Meiosis Reveals Sex-Specific Patterns along Chromosomes. PLOS Genetics 7, e1002354. doi: 10.1371/journal.pgen.1002354.

Goepfert, D. (1974). Karyotypes and DNA content in species of *Ranuculus* L. and related genera. Botaniska notiser. Available at: https://scholar.google.com/scholar_lookup?title=Karyotypes+and+DNA+content+in+species+of+Ranunculus+L.+and+related+genera&author=Goepfert%2C+D.&publication_year=1974 [Accessed August 1, 2022].

Jin, Y., Zhao, W., Nie, S., Liu, S.-S., El-Kassaby, Y. A., Wang, X.-R., et al. (2019). Genome-Wide Variant Identification and High-Density Genetic Map Construction Using RADseq for Platycladus

orientalis (Cupressaceae). G3: Genes, Genomes, Genetics 9, 3663–3672. doi: 10.1534/g3.119.400684.

Kearsey, M. J., Ramsay, L. D., Jennings, D. E., Lydiate, D. J., Bohuon, E. J. R., and Marshall, D. F. (1996). Higher recombination frequencies in female compared to male meisoses in Brassica oleracea. Theoret. Appl. Genetics 92, 363–367. doi: 10.1007/BF00223680.

Keim, P., Diers, B. W., Olson, T. C., and Shoemaker, R. C. (1990). RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126, 735–742. doi: 10.1093/genetics/126.3.735.

Kelly, A. L., Sharpe, A. G., Nixon, J. H., Lydiate, D. J., and Evans, E. J. (1997). Indistinguishable patterns of recombination resulting from male and female meioses in Brassica napus (oilseed rape). Genome 40, 49–56. doi: 10.1139/g97-007.

Kosambi, D. D. (1943). The Estimation of Map Distances from Recombination Values. Annals of Eugenics 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x.

Koseki, M., Goto, K., Masuta, C., and Kanazawa, A. (2005). The Star-type Color Pattern in Petunia hybrida 'Red Star' Flowers is Induced by Sequence-Specific Degradation of Chalcone Synthase RNA. Plant and Cell Physiology 46, 1879–1883. doi: 10.1093/pcp/pci192.

Le Maitre, N. C., Pirie, M. D., and Bellstedt, D. U. (2019). Floral Color, Anthocyanin Synthesis Gene Expression and Control in Cape Erica Species. Frontiers in Plant Science 10. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01565 [Accessed August 1, 2022].

Lenormand, T., and Dutheil, J. (2005). Recombination Difference between Sexes: A Role for Haploid Selection. PLOS Biology 3, e63. doi: 10.1371/journal.pbio.0030063.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi: 10.1093/bioinformatics/btp324.

Mariotti, R., Fornasiero, A., Mousavi, S., Cultrera, N. G. M., Brizioli, F., Pandolfi, S., et al. (2020). Genetic Mapping of the Incompatibility Locus in Olive and Development of a Linked Sequence-Tagged Site Marker. Frontiers in Plant Science 10. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01760 [Accessed August 1, 2022].

Martina, M., Acquadro, A., Barchi, L., Gulino, D., Brusco, F., Rabaglio, M., et al. (2022). Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in

Anemone coronaria L. International Journal of Molecular Sciences 23, 3126. doi: 10.3390/ijms23063126.

Nelson, D. L., Orr, H. T., and Warren, S. T. (2013). The Unstable Repeats - Three Evolving Faces of Neurological Disease. Neuron 77, 825–843. doi: 10.1016/j.neuron.2013.02.022.

Ott, A., Trautschold, B., and Sandhu, D. (2011). Using Microsatellites to Understand the Physical Distribution of Recombination on Soybean Chromosomes. PLOS ONE 6, e22306. doi: 10.1371/journal.pone.0022306.

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., and Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335, 721–726. doi: 10.1038/335721a0.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLOS ONE 7, e37135. doi: 10.1371/journal.pone.0037135.

Plomion, C., and Durel, C. (1996). Estimation of the average effects of specific alleles detected by the pseudo-testcross QTL mapping strategy. Genetics Selection Evolution 28, 223. doi: 10.1186/1297-9686-28-3-223.

Portis, E., Scaglione, D., Acquadro, A., Mauromicale, G., Mauro, R., Knapp, S. J., et al. (2012). Genetic mapping and identification of QTL for earliness in the globe artichoke/cultivated cardoon complex. BMC Res Notes 5, 252. doi: 10.1186/1756-0500-5-252.

Reiter, R. S., Williams, J. G., Feldmann, K. A., Rafalski, J. A., Tingey, S. V., and Scolnik, P. A. (1992). Global and local genome mapping in Arabidopsis thaliana by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci U S A 89, 1477–1481.

Sardell, J. M., and Kirkpatrick, M. (2020). Sex Differences in the Recombination Landscape. Am Nat 195, 361–379. doi: 10.1086/704943.

Scaglione, D., Reyes-Chin-Wo, S., Acquadro, A., Froenicke, L., Portis, E., Beitel, C., et al. (2016). The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. Sci Rep 6, 19427. doi: 10.1038/srep19427.

Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. Current Protocols in Molecular Biology 122, e59. doi: 10.1002/cpmb.59.

Song, X., Xu, Y., Gao, K., Fan, G., Zhang, F., Deng, C., et al. (2020). High-density genetic map construction and identification of loci controlling flower-type traits in Chrysanthemum (Chrysanthemum × morifolium Ramat.). Horticulture Research 7, 108. doi: 10.1038/s41438-020-0333-1.

Tanksley, S. D., Ganal, M. W., Prince, J. P., de Vicente, M. C., Bonierbale, M. W., Broun, P., et al. (1992). High density molecular linkage maps of the tomato and potato genomes. Genetics 132, 1141–1160. doi: 10.1093/genetics/132.4.1141.

Team, R. C. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. Available at: https://cir.nii.ac.jp/crid/1574231874043578752 [Accessed August 1, 2022].

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., et al. (2020). A New Intra-Specific and High-Resolution Genetic Map of Eggplant Based on a RIL Population, and Location of QTLs Related to Plant Anthocyanin Pigmentation and Seed Vigour. Genes 11, 745. doi: 10.3390/genes11070745.

Torello Marinoni, D., Valentini, N., Portis, E., Acquadro, A., Beltramo, C., and Botta, R. (2018). Construction of a high-density genetic linkage map and QTL analysis for hazelnut breeding. Acta Hortic., 25–30. doi: 10.17660/ActaHortic.2018.1226.3.

Valentini, N., Portis, E., Botta, R., Acquadro, A., Pavese, V., Cavalet Giorsa, E., et al. (2021). Mapping the Genetic Regions Responsible for Key Phenology-Related Traits in the European Hazelnut. Front Plant Sci 12, 749394. doi: 10.3389/fpls.2021.749394.

Vallejos, C. E., Sakiyama, N. S., and Chase, C. D. (1992). A molecular marker-based linkage map of Phaseolus vulgaris L. Genetics 131, 733–740. doi: 10.1093/genetics/131.3.733.

van Geest, G., Bourke, P. M., Voorrips, R. E., Marasek-Ciolakowska, A., Liao, Y., Post, A., et al. (2017). An ultra-dense integrated linkage map for hexaploid chrysanthemum enables multi-allelic QTL analysis. Theor Appl Genet 130, 2527–2541. doi: 10.1007/s00122-017-2974-5.

Van Ooijen, J. W. (2004). MapQTL® 5, Software for the mapping of quantitative trait loci in experimental populations.

Van Ooijen, J. W. (2006). JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations – ScienceOpen. Available at: https://www.scienceopen.com/document?vid=baa76c8c-fb55-4c13-a6ca-24c71002ab5a [Accessed August 1, 2022].

Vincenten, N., Kuhl, L.-M., Lam, I., Oke, A., Kerr, A. R., Hochwagen, A., et al. (2015). The kinetochore prevents centromere-proximal crossover recombination during meiosis. Elife 4, e10850. doi: 10.7554/eLife.10850.

Víquez-Zamora, M., Vosman, B., van de Geest, H., Bovy, A., Visser, R. G., Finkers, R., et al. (2013). Tomato breeding in the genomics era: insights from a SNP array. BMC Genomics 14, 354. doi: 10.1186/1471-2164-14-354.

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93, 77–78. doi: 10.1093/jhered/93.1.77.

Vukosavljev, M., Arens, P., Voorrips, R. E., van 't Westende, W. P., Esselink, G. D., Bourke, P. M., et al. (2016). High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68k rose SNP array. Hortic Res 3, 1–8. doi: 10.1038/hortres.2016.52.

Wang, Z., Lu, G., Wu, Q., Li, A., Que, Y., and Xu, L. (2022). Isolating QTL controlling sugarcane leaf blight resistance using a two-way pseudo-testcross strategy. The Crop Journal 10, 1131–1140. doi: 10.1016/j.cj.2021.11.009.

Wang, Z., Song, M., Li, Y., Chen, S., and Ma, H. (2019). Differential color development and response to light deprivation of fig (Ficus carica L.) syconia peel and female flower tissues: transcriptome elucidation. BMC Plant Biology 19, 217. doi: 10.1186/s12870-019-1816-9.

Wang, Z., Wei, X., Yang, J., Li, H., Ma, B., Zhang, K., et al. (2020). Heterologous expression of the apple hexose transporter Md HT 2.2 altered sugar concentration with increasing cell wall invertase activity in tomato fruit. Plant Biotechnol J 18, 540–552. doi: 10.1111/pbi.13222.

Wu, S., Yang, J., Huang, Y., Li, Y., Yin, T., Wullschleger, S. D., et al. (2010). An Improved Approach for Mapping Quantitative Trait Loci in a Pseudo-Testcross: Revisiting a Poplar Mapping Study. Bioinform Biol Insights 4, BBI.S4153. doi: 10.4137/BBI.S4153.

Zafar, F., Okita, A. K., Onaka, A. T., Su, J., Katahira, Y., Nakayama, J., et al. (2017). Regulation of mitotic recombination between DNA repeats in centromeres. Nucleic Acids Research 45, 11222–11235. doi: 10.1093/nar/gkx763.

Zheng, X., Om, K., Stanton, K. A., Thomas, D., Cheng, P. A., Eggert, A., et al. (2021). The regulatory network for petal anthocyanin pigmentation is shaped by the MYB5a/NEGAN transcription factor in Mimulus. Genetics 217, iyaa036. doi: 10.1093/genetics/iyaa036

Chapter IV

frontiers | Frontiers in Plant Science

Diversity analyses in two ornamental and large-genome Ranunculaceae species based on a low-cost Klenow NGS-based protocol (K-seq)

Matteo Martina, Alberto Acquadro, Ezio Portis, Lorenzo Barchi*, Sergio Lanteri

DISAFA, Plant Genetics, University of Turin, Italy

Keywords: K-seq; genotyping; ornamentals; *Ranuculus asiaticus*; *Anemone coronaria*; fingerprinting; NGS.

Abstract

Persian buttercup (*Ranunculus asiaticus* L.) and poppy anemone (*Anemone coronaria* L.) are ornamental, outcrossing, perennial species belonging to the Ranunculaceae family, characterized by large and highly repetitive genomes. We applied K-seq protocol in both species to generate high-throughput sequencing data and produce a large number of genetic polymorphisms. The technique entails the application of Klenow polymerase-based PCR using short primers designed by analyzing k-mer sets in the genome sequence. To date the genome sequence of both species has not been released, thus we designed primer sets based on the reference the genome sequence of the related species Aquilegia oxysepala var. kansuensis (Brühl). A whole of 11,542 SNPs were selected for assessing genetic diversity of eighteen commercial varieties of *R. asiaticus*, while 1,752 SNPs

for assessing genetic diversity in six cultivars of *A. coronaria*. UPGMA dendrograms were constructed and in *R. asiaticus* integrated in with PCA analysis. This study reports the first molecular fingerprinting within Persian buttercup, while the results obtained in poppy anemone were compared with a previously published SSR-based fingerprinting, proving K-seq to be an efficient protocol for the genotyping of complex genetic backgrounds.

1 Introduction

The Ranunculaceae family is classified as a basal eudicotyledonous group and includes approximately 2500 species distributed across 53 genera (Jiang et al., 2017). Over the centuries, this botanical family has been the subject of systematic investigations due to its intriguing position within flowering plants and its notable variation in vegetative and reproductive structures. Breeding efforts have been mainly focused on the exploitation of the ornamental potential of two species: *Anemone coronaria* L., commonly known as poppy anemone (2n=2x=16; estimated genome size: 9.08Gb; Martina et al., 2022a) and *Ranunculus asiaticus* L., commonly known as Persian buttercup (2n=2x=16; estimated genome size: 7.6Gb; Martina et al., 2022b). Both species are outcrossing, perennial crops marketed as cut flowers and potted or garden plants. In poppy anemone modern cultivars have been primarily selected for the production of cut flowers, and breeding efforts has been focusing on product uniformity, obtainment of new flower colors, adaptation to long photoperiods, precocity, and longevity of vase life. Cut flower varieties must bloom early, possess sturdy stems and large multi-sepaled flowers, while garden and potted plant types must exhibit compact growth habits, upright leaves

with short petioles and abundant and coordinated flowering. In Persian buttercup, selection for cut flowers has been mainly focusing on flower diameter, morphology, and color. The Pon-Pon® varieties, created by the Company 'Biancheri Creations' through a long breeding and micropropagation process, have become increasingly important in the market due to their unique bicolor and jagged petal shape. Other innovative *Ranuculus* varieties, such as those with a "green center" are also gaining popularity. Furthermore, in both species a key breeding objective has been also the resistance/tolerance to biotic and abiotic stresses, and since the commercial varieties are vegetatively propagated through rhizomes, the easy handling of roots at harvest and rhizome storability are highly desirable for commercial profitable production. In this context the availability of molecular markers and the application of novel genomics tools may play an essential role to accelerate breeding programs.

The advent of high-throughput next-generation sequencing (NGS) methods has resulted in a shift from fragment-based polymorphism identification to sequence-based single nucleotide polymorphism (SNP) detection, thereby augmenting the number of markers available for population genetics research and expediting the development of high-density genetic maps and the discovery of genes and quantitative trait loci (QTLs) (Sahu et al., 2020). The sustained reduction in sequencing costs has facilitated the formation of genomic consortia for the release of reference genomes in a diverse range of plant species (Jaillon et al., 2007; Sato et al., 2012; Hibrand Saint-Oyant et al., 2018; Raymond et al., 2018; Acquadro et al., 2020; Barchi et al., 2021; Nakano et al., 2021; Nashima et al., 2021), making it possible to design genotyping assays that offer

potent tools for next-generation plant breeding (Sim et al., 2012; Víquez-Zamora et al., 2013; Barabaschi et al., 2016; Vukosavljev et al., 2016; van Geest et al., 2017; Barchi et al., 2019). Nevertheless, the efficacy of these assays is dependent on the availability of prior genomic information and their application to "orphan" species possessing large, heterozygous, and highly repetitive genomes may prove challenging, and even high-depth sequencing may not guarantee the quality of assembly for such species (Unamba et al., 2015). An alternative approach is represented by reduced-representation sequencing (RRS) techniques, which can generate genome-wide high-throughput sequencing data and obtain a significant number of genetic polymorphisms, representative of the entire genome information of the species (Wright et al., 2019). RRS based on Restriction enzymes-based techniques are generally applicable to any genetic background (Davey and Blaxter, 2010; Barchi et al., 2012; Peterson et al., 2012; Acquadro et al., 2017; Jordon-Thaden et al., 2020; Sunde et al., 2020; Toppino et al., 2020), mainly requiring the selection of the appropriate enzyme, making them a common approach in non-model species and small to medium-scale projects. However, such techniques but require a license which increasing the genotyping costs (Barchi et al., 2019). To overcome these limitations, non-restriction based RRS methodologies have been developed, such as rAmpSeq (Buckler et al., 2016), which is based on the amplification and sequencing of repetitive regions, have been developed. Building on this concept, the K-seq protocol was developed by Mata-Nicolás et al., 2021 and Ziarsolo et al., 2021, which relies on the amplification of genomic regions using two steps of Klenow amplification with short oligonucleotides, followed by standard PCR and Illumina

sequencing. The design of primers in this protocol can also be performed using the genome sequence of a closely related species by analyzing k-mer sets, making it applicable to genome-orphan species as well. Furthermore, the license-free nature of K-seq and low library preparation costs make it a valuable alternative to GBS and ddRADseq.

To date, the reference genomes of both Poppy anemone and Persian buttercup have not been released and the application of DNA-based techniques has been limited (Nissim et al., 2004; Fang and Paz, 2005; Laura et al., 2006; Martina et al., 2022a, 2022b). Here, we applied the K-seq protocol on these two species using primer sets obtained from the high-quality reference genome of the closely-related *Aquilegia oxysepala* var. *kansuensis* (Brühl.; Xie et al., 2020). The protocol was applied to breeding material commercialized by Biancheri Creazioni, specifically eighteen cultivars of *R. asiaticus* and six cultivars of *A. coronaria*, the latter being validated with previously published Single Sequence Repeats (SSR)-based fingerprinting (Martina et al., 2022a). Furthermore, the fine-tuning of the protocol showed to be cost-effecting for routine genotyping application.

2 Materials and Methods

2.1 Plant material and DNA extraction

Eighteen market varieties of *R. asiaticus*, known for their production of double and Pon-Pon® flowers, and six diploid cultivars of *A. coronaria*, which have been previously characterized using SSR markers (Martina et al., 2022a), were procured from "Biancheri Creazioni" as representative samples of their commercial breeding stock (as indicated in Supplemental Table 1). Genomic DNA was extracted from frozen leaves with the Plant

DNA Kit (E.Z.N.A.®), following manufacturer's instructions. DNA quality was assessed through the NanoDrop™ 2000 spectrophotometer (ratio 260/280 between 1.8 and 2.0; ratio 260/230 > 2.0), and the Qubit® 2.0 Fluorometer was used for DNA quantification.

2.2 Primers' design and library preparation

The recently published chromosome-scale reference genome of *Aquilegia oxysepala* var. *kansuensis*, a closely-related species to both species in study and belonging to the same family, was retrieved from the NCBI database. The Primer Explorer 2 pipeline (https://github.com/bioinfcomav/primer_explorer2) and used for oligonucleotide sets identification. Briefly, the 1000 most abundant k-mers with a GC content between 35 and 75% were selected and combined to create sequence sets (8-9bp each) suitable for PCR amplification. The pipeline provided a virtual PCR, producing a report (Supplemental Table 2) which included a number of predicted single and repetitive products for each primer pair within each oligonucleotide set. Three primer pairs were selected for *A. coronaria* and *R. asiaticus* genotyping, according to the predicted number of amplified fragments in A. oxysepala var. kansuensis (Supplemental Table 3). Library preparation was performed according to Mata-Nicolás et al., 2021 and Ziarsolo et al., 2021, halving enzymes concentrations and doubling reaction times in order to reduce genotyping costs. Briefly, the genomic DNA was subjected to denaturation and annealing at 37°C using forward primers. Then, the large Klenow fragment of DNA polymerase I (New England Biolabs, MA, USA) was utilized for the initial round of DNA synthesis. Following inactivation of the polymerase at 75°C, the remaining single-stranded DNA was degraded at

37°C using Exonuclease I (New England Biolabs, MA, USA) and inactivated at 80°C. The above steps were repeated with the reverse primers. The reaction product was employed as a template in a standard 15-cycle enrichment PCR utilizing in-house Illumina Nextera primers, which featured dual indexes. Finally, the indexed samples were combined, and the libraries were size-selected (300-700 bp) through gel excision and extraction (E.Z.N.A.® Gel Extraction Kit), as per the manufacturer's protocol.

2.3 Sequencing data analysis and SNPcalling

Libraries were Illumina sequenced (150PE) on NovaSeq 6000 (Illumina, San Diego, CA). After demultiplexing, reads were trimmed and cleaned using fastp (Danecek et al., 2021) and mock references (catalogs) were produced through the GBS-SNP-CROP pipeline (Melo et al., 2016). The six genotypes in study were used as input for *A. coronaria*, while for *R. asiaticus* just two genotypes, showing the highest sequencing output, and representative of the cultivars characterized by the so called Pon-Pon® (Pon-Pon® - 929) and double flower (Success® - 311), were used. Sequences for each genotype were mapped to the catalog file using the Burrows-Wheeler Aligner (BWA, v0.7.17) program and the ‘mem’ command with the default parameters (Li and Durbin, 2009). SNP calling was performed using Bcftools mpileup (v1.17; Danecek et al., 2021) and SNPs were filtered to remove those with a mean depth below 20 (MEAN(FMT/DP)>=20), a mean QUAL higher than 20 (MEAN(FMR/GQ)>20), no more than 20% of missing data (F_MISSING=<0.2), and a MAF >= 0.05. Markers were named according to the catalog sequence in which they were identified. Sequencing reads

were aligned also on the reference genome of A. oxysepala, following the same protocol mentioned above. After reads aligning, a window size of 100kb was set in the Mosdepth script (Pedersen and Quinlan, 2018). Putative k-mer fragments were retrieved through the K-seq bioinformatic pipeline for Aquiligea's genome, and a 100kb was applied using bedtools windows and bedtools coverage (v2.31.0; Quinlan and Hall, 2010), to calculate the bin coverage across the genome.

2.4 Molecular Fingerprinting and data analysis

After quality filtering, the identified markers were imported into MEGA11 software (Tamura et al., 2021) and used to construct a UPGMA-based dendrogram (Sneath and Sokal, 1973) with 1,000 bootstraps for *A. coronaria*, consistently with the previously reported SSR-based philogeny, and a NJ-based dendrogram for *R. asiaticus*. Principal component analysis (PCA) was also performed (SNPRelate v1.32.1; Zheng et al., 2012), displaying the multi-dimensional relationship between genotypes, and the two axes were graphically plotted, according to the extracted eigenvectors. To compare the previously published SSR-based fingerprint with the novel one, Mantel test (Mantel, 1967) was performed to evaluate the correlation index (r), between the two similarity matrices.

3 Results and Discussion

3.1 Primers' development and sequencing data analysis

The versatility of K-seq as a technique has been well established, as demonstrated by its ability to provide reliable genotyping results in closely related species using primer sets designed for a specific species (Mata-Nicolás et al., 2021; Ziarsolo et al., 2021). This makes K-seq particularly useful when working with genome-orphan species.

Sampling genomic loci - To date, the *Anemone* and *Ranunculus* genera lack high-quality reference genomes. As a result, primer mining was performed using the Primer Explorer 2 pipeline, which was based on the recently released genome sequence of *A. oxysepala* var. *kansuensis*, the most closely related species. The pipeline utilized the analysis of the 1000 most abundant k-mers, enabling the selection of three sets of primer pairs that provided the greatest number of amplicons in *A. oxysepala* (Supplemental Table 3).

Sequencing throughput - Overall, ~26M reads for the six cultivars of *A. coronaria* and ~164M reads for the 18 *R. asiaticus* cultivars were produced, generating a total of 29.3Gb of raw data. The number of reads obtained per sample showed variability among accessions. In *Anemone*, reads output ranged between 1.2 and 6.8 million, while in *Ranunculus* between 1.1 and 18.8 million (Supplemental Table 4). The presence of heterogeneity in the number of reads per sample has been previously reported by Mata-Nicolás et al., 2021 and Ziarsolo et al., 2021. This phenomenon is believed to be the result of the absence of DNA quantification of PCR-amplified products prior to library pooling. In order to mitigate this issue and maintain the cost-effectiveness of the protocol,

we conducted a direct quantification of the library using electrophoresis and densitometric analysis of band intensity, as well as comparison to a standard curve generated from DNA of known concentrations (ranging from 20 to 100 ng). Despite these efforts, our results indicated that this approach was not sufficient for standardizing the number of reads in the K-seq sample pooling step. Indeed, this step is crucial for producing standardized data and may require additional purification techniques such as SPRI bead purification after library enrichment to remove primer dimers and facilitate accurate quantification of products using fluorescence-based methods. According to our evaluations and in line with the current consumables pricing in our lab, by halving enzymatic concentrations and reducing the enrichment reaction volume to 20ul (the original protocol suggests 50ul), the per sample cost of the protocol can be decreased by a broad 50%, passing from 12€ to 6€ per library. In this scenario, we believe that SPRI beads might be taken into account for routine genotyping application of k-seq protocol, even with large sample sets. Such optimization allows for library quantification prior to samples pooling, requiring extra costs represented by fluorometer quantification. However, it must be noted that, after our adjustments, SPRI beads introduction in the protocol might impact on the single sample cost for a broad 30%, leaving space for their evaluation according to project budget and overall experimental design.

Not only library preparation, but also sequencing costs impact on project budget. In our experiment the technique was found to be reliable without SPRI optimizations and, with an average of 2M reads for *A. coronaria* and 4M reads for *R. asiaticus*, our results suggest that a sequencing output of

approximately 1Gb per sample represents a cost-effective solution for high-quality genotyping of complex species. With the current sequencing costs, such coverage appears to be reasonable for large-genome species, but might be considerably lowered for smaller ones.

3.2 SNP mining

Mock references (*catalogs*) were produced for the two species through the GBS-SNP-CROP pipeline. A whole of 2,374,786 contigs catalog was assembled in *A. coronaria*, while the *R. asiaticus* the catalog included 2,416,619 contigs. Cleaned reads were backaligned on the mock references and 629,004 and 8,021,225 SNPs were called, subsequently reduced to 1,752 and 11,542 after filtering for *A. coronaria* and *R. asiaticus* respectively and were then used for genetic diversity analyses. Heterozygosity was calculated (Supplemental Table 4), identifying a clear difference between the two species. The average heterozygosity was observed to be 0.5 in *Anemone* cultivars and was estimated at 0.34 in *Ranuculus*. These differences may be attributed to the distinct breeding practices applied for cultivar selection in the two species. Specifically, poppy anemone cultivars are usually produced by cross-pollinating selected parental lines, resulting in a substantial degree of morphological heterogeneity and without a focus on reducing heterozygosity. Conversely, Persian buttercup cultivars are commonly propagated by cloning and selected for their superior market appeal from a diverse range of segregating crosses. In this species, breeding strategies have also been directed towards reducing heterozygosity with the objective of achieving stable crosses (personal communication with Biancheri Creazioni). Although both species are known to possess a highly heterozygous genetic

background (Fang and Paz, 2005; Laura et al., 2006; Martina et al., 2022a, 2022b), the lower heterozygosity observed in Persian buttercup may be attributed to the breeding selection implemented at Biancheri Creazioni. The expected distribution of the k-mer fragments on the *A. oxysepala* genome was obtained through the K-seq native pipeline, predicting the three selected primer sets to produce a good coverage of all the seven chromosome of the species (Figure 1). Subsequently, the raw reads were aligned to the reference genome of *A. oxysepala*, and their distribution was compared to the predicted k-mer fragments to assess the amplified fragments in the two ornamental species (Figure 1). Approximately, 20% of the reads mapped to Aquilegia's genome in both species, while an average of 90% of the reads were successfully aligned to the two catalogs. As shown in Figure 1, the reads that mapped to the reference genome exhibited a broad coverage across the entire genome, consistent with the predicted fragments. Differences in overall coverage were observed, likely influenced by the number of individuals in the two panels and discrepancies between the two sequencing outputs. Nevertheless, it is plausible that the mapping reads may represent essential shared information among the three Ranunculaceae species, which diverged approximately 60 million years ago (Zhai et al., 2019).

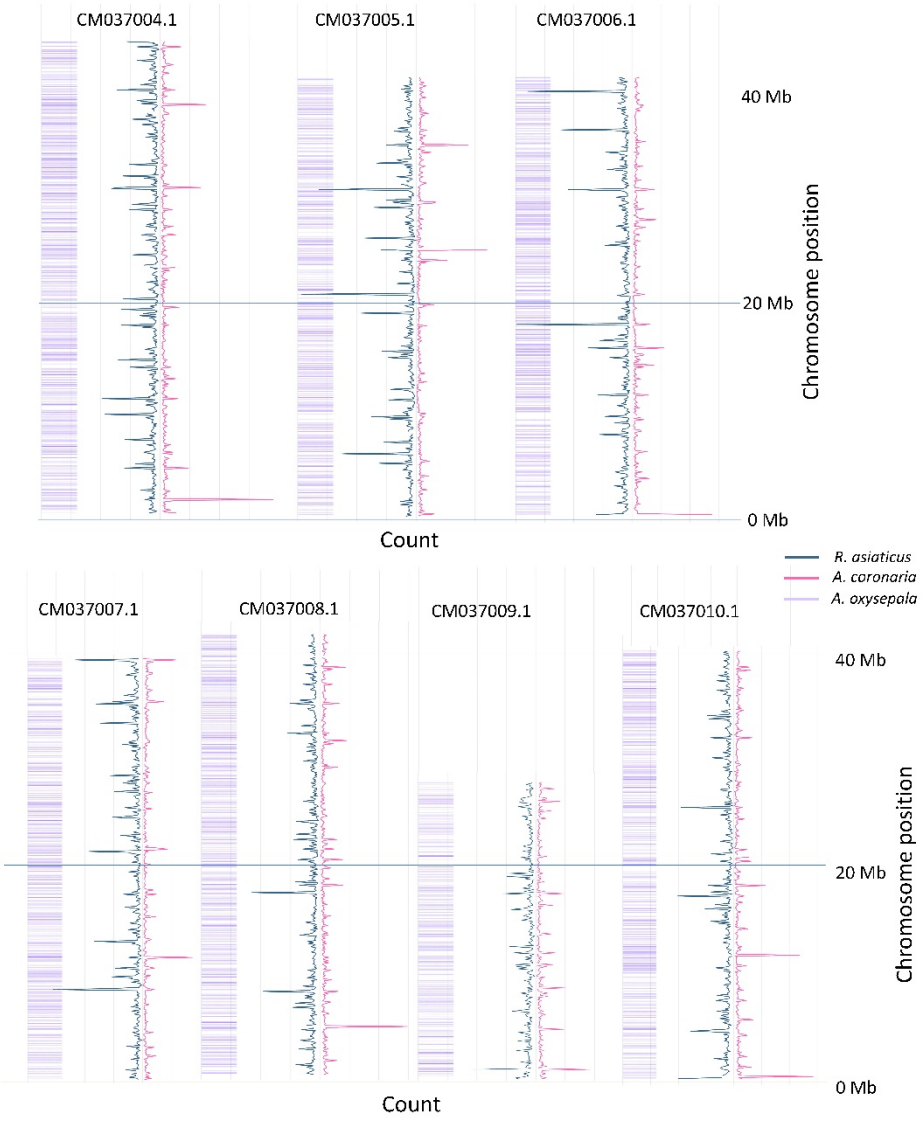


Figure 1. Genome-Wide coverage of the sequenced reads on the A. oxysepala genome. Only the seven chromosomes are presented for conciseness: *R. asiaticus* in blue, while *A. coronaria* in pink. The values for *A. coronaria*, for which only 6 samples were sequences, were re-scaled of a 3x factor, allowing a clear comparison in the figure. The predicted K-seq framents originated by the three sets of primers in A. oxysepala are reported in purple.

3.2 Genetic variation in cultivars of poppy anemone

An UPGMA dendrogram of the six *A. coronaria* cultivars was produced (Figure 1a), and the detected genetic relationships among the cultivars were in good agreement with the previously published SSR-based fingerprint (Figure 1b; Martina et al., 2022a), since Mantel correlation test highlighted a good fit between the similarity matrices (0.85). Two major clades were identified following the application of the K-seq protocol (Figure 2A). In the first branch, the cultivars “Rosa”, and “Edge” clustered with a bootstrap probability of 97%. In the second branch, the cultivars “Bordeaux”, “Magenta” and “Tigre” clustered with a bootstrap value of 92%. The cultivar “Tigre wine” was identified as the most distant variety, in contrast with the previously published SSR based fingerprint that grouped it with “Tigre” (Figure 2B). As previously reported (Xie et al., 2020), this demonstrates the effects associated with different genotyping platforms on the data interpretation. However, the two topologies were overall highly related, suggesting that K-seq represents a valuable tool to generate genotypic profiles. Furthermore, the K-seq procedure represents a valid and alternative approach to the more costly and time-consuming SSR-assays development and to the patented GBS-like approaches relying on restriction enzymes.

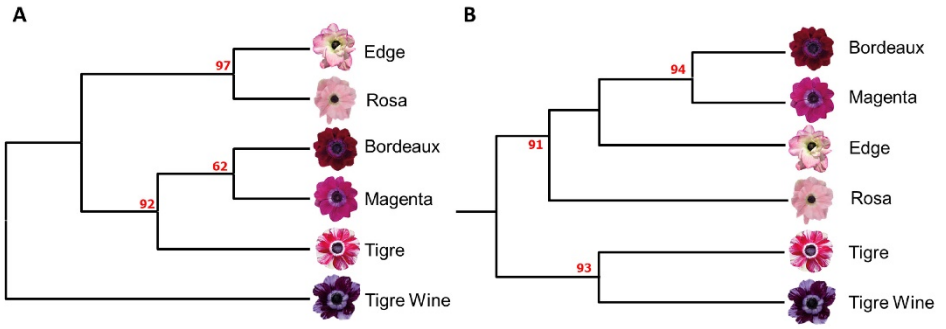


Figure 2. A) UPGMA dendrogram based on 1,752 SNPs identified by the K-seq protocol B) UPGMA dendrogram based on 62 SSR (adapted by Martina et al., 2022b). Bootstrap values (%) are reported in red.

3.3 Genetic variation in cultivars of persian buttercup

The same sets of primer used in poppy anemone was applied to obtain a NJ dendrogram of the eighteen cultivars of *R. asiaticus* (Figure 3A). Since no prior phylogenetic studies have been reported for Persian buttercup, our results are the first ones obtained in this species. A PCA analysis was performed on the genotyping data, highlighting that the first two axes (Figure 3B) explained 7.7 and 7.3% of the overall genetic variation. In Figure 3B the cultivars pairs “606” / “945”, and “929” / “8” clustered according to the first PCA component and were clearly differentiated from the others, while the cultivars “418” and “78” clustered according to the second one. Furthermore, except for the cultivar “945”, the NJ-based phylogeny clustered the cultivars defined Pon-Pon® in Clade I (Figure 3A), and the first component of the PCA grouped the varieties with brighter flower colors in the right side of the scatter plot and the darker ones in the left side.

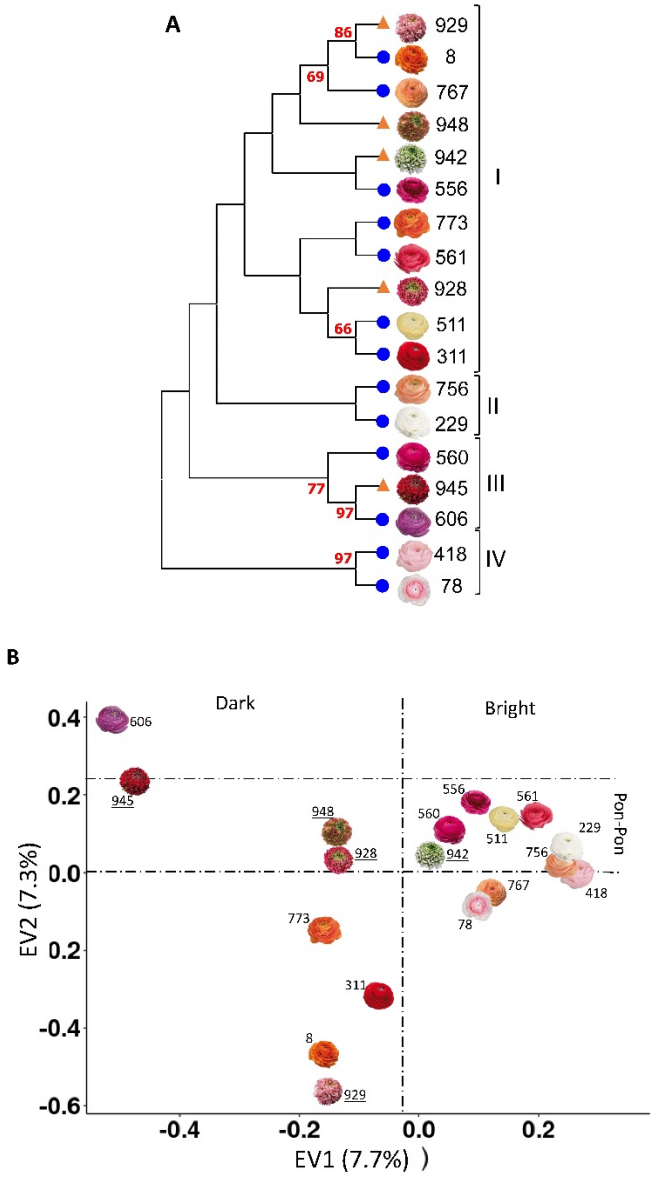


Figure 3. NJ dendrogram (A) and PCA analysis (B) of the eighteen varieties of *R. asiaticus*, based on 11,542 SNPs. Bootstrap values (%) higher than 60 are reported in red. Pon-Pon varieties are represented with an orange triangle in the NJ dendrogram (A), and their names are underlined in the PCA (B); on the other hand, double-flowered varieties are represented with a blue circle in the NJ dendrogram (B), to make it possible to easily discriminate between the two groups.

4. Conclusions

The application of K-seq in two Ranunculaceae species, characterized by large and highly repetitive genomes, has here been reported for the first time. The species, whose sequence has yet to be released, were found to be suitable for this genotyping protocol, which can be easily applied even in "orphan" species by utilizing genomic information from closely-related species. In 2019, Zhai et al. investigated the phylogeny and evolution of the family by sequencing the chloroplast genome of 35 species, representing 31 genera of the 14 tribes of the Ranunculaceae family. Apart from Glaucideae, Hydrastideae, and Coptideae, which diverged earlier (89.9 Mya), molecular clock analysis highlighted how the first divergence within the Ranunculaceae core occurred around 66.3 Mya, followed by subsequent independent branch-splitting event in the two lineages. As *Ranunculus* sp., *Anemone* sp., and *Aquilegia* sp. belong to the Ranunculaceae core, the applied methodology might be robust enough for comparing species across different genus. Furthermore, by reducing the enzymatic concentration in the reaction, the cost of the library preparation was significantly reduced. As previously reported (Ziarsolo et al., 2021), the number of K-seq fragments generated by Klenow amplification can deviate from the expected amount due to mismatched primers. By assessing the number of generated fragments through agarose gel visualization, or by using microfluidic-based electrophoresis, the best primer set can be identified, and the generated data also allow for the selection of the best number of primers sets applied for genotyping. However, our experiment highlighted that the protocol was reliable without such optimizations, demonstrating that an expected sequencing

output of approximately 1Gb per sample might be a cost-effective solution for achieving high-quality genotyping results in complex species. Our results confirm that K-seq is comparable or in some cases cheaper than GBS in terms of cost. In *A. coronaria* the estimated genetic diversity and the clustering of the eight cultivars in study were found consistent with those previously published and based on a SSR genotyping. Side by side this study provides the first assessment of genetic diversity in marketed varieties of *R. asiaticus* and novel information on their genetic relationships. The Pon-Pon® varieties did not cluster separately from the others, and thus they do not appear genetically differentiated. As confirmed by the breeders of the Company 'Biancheri Creations', the peculiar flower trait not always showed to be stable in different environmental conditions and it is thus presumably attributable to epigenetic mechanisms, at least partly influenced by growth conditions.

The user-friendly nature of this technique, together with its cost-effectiveness, may find applications in cultivar fingerprinting and legal disputes, having the potential to facilitate the implementation of molecular-assisted breeding programs in this two complex species, promoting the efficient development of new cultivars and enabling the identification of the genetic factors underlying complex traits.

5. Data Availability Statement

Sequencing data used in this study are openly available in the NCBI database (PRJNA954204).

6. Author Contributions

Conceptualization, M.M., L.B., E.P and S.L.; methodology, M.M., L.B. and A.A.; software, M.M., L.B. and A.A..; validation, M.M., L.B. and E.P..; formal analysis, M.M. and E.P.; investigation, M.M and E.P..; data curation, M.M.; writing—original draft preparation, M.M. and A.A.; writing-review and editing, L.B., E.P and S.L.; visualization, A.A. and E.P.; supervision, E.P and S.L. All authors have read and agreed to the published version of the manuscript.

7. Acknowledgments

The authors thank the company Biancheri Creazioni for supplying the plant material, and in particular Dr. Mario Rabaglio for his technical support.

8. Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9. Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

10. Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1187205/full#supplementary-material

11. References

Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., et al. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. PLoS ONE 12, e0180774. doi: 10.1371/journal.pone.0180774.

Acquadro, A., Portis, E., Valentino, D., Barchi, L., and Lanteri, S. (2020). "Mind the Gap": Hi-C Technology Boosts Contiguity of the Globe Artichoke Genome in Low-Recombination Regions. G3 (Bethesda) 10, 3557–3564. doi: 10.1534/g3.120.401446.

Barabaschi, D., Tondelli, A., Desiderio, F., Volante, A., Vaccino, P., Valè, G., et al. (2016). Next generation breeding. Plant Science 242, 3–13. doi: 10.1016/j.plantsci.2015.07.010.

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., et al. (2019). Single Primer Enrichment Technology (SPET) for High-Throughput Genotyping in Tomato and Eggplant Germplasm. Front. Plant Sci. 10, 1005. doi: 10.3389/fpls.2019.01005.

Barchi, L., Lanteri, S., Portis, E., Valè, G., Volante, A., Pulcini, L., et al. (2012). A RAD Tag Derived Marker Based Eggplant Linkage Map and the Location of QTLs Determining Anthocyanin Pigmentation. PLoS ONE 7, e43740. doi: 10.1371/journal.pone.0043740.

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., et al. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. The Plant Journal 107, 579–596. doi: 10.1111/tpj.15313.

Buckler, E. S., Ilut, D. C., Wang, X., Kretzschmar, T., Gore, M., and Mitchell, S. E. (2016). rAmpSeq: Using repetitive sequences for robust genotyping. 096628. doi: 10.1101/096628.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience 10, giab008. doi: 10.1093/gigascience/giab008.

Davey, J. W., and Blaxter, M. L. (2010). RADSeq: next-generation population genetics. Brief Funct Genomics 9, 416–423. doi: 10.1093/bfgp/elq031.

Fang, J., and Paz, T. (2005). Modified AFLP technique for fingerprinting of *Anemone coronaria*. Plant Genetic Resources 3, 73–75. doi: 10.1079/PGR200456.

Hibrand Saint-Oyant, L., Ruttink, T., Hamama, L., Kirov, I., Lakhwani, D., Zhou, N. N., et al. (2018). A high-quality genome sequence of Rosa chinensis to elucidate ornamental traits. Nature Plants 4, 473–484. doi: 10.1038/s41477-018-0166-1.

Jiang, N., Zhou, Z., Yang, J. B., Zhang, S. D., Guan, K. Y., Tan, Y. H., & Yu, W. B. (2017). Phylogenetic reassessment of tribe Anemoneae (Ranunculaceae): Non-monophyly of *Anemone* sl revealed by plastid datasets. PLoS One, 12(3), e0174792.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467. doi: 10.1038/nature06148.

Jordon-Thaden, I. E., Beck, J. B., Rushworth, C. A., Windham, M. D., Diaz, N., Cantley, J. T., et al. (2020). A basic ddRADseq two-enzyme protocol performs well with herbarium and silica-dried tissues across four genera. Appl Plant Sci 8. doi: 10.1002/aps3.11344.

Laura, M., Allavena, A., Magurno, F., Lanteri, S., and Portis, E. (2006). Genetic variation of commercial *Anemone coronaria* cultivars assessed by AFLP. The Journal of Horticultural Science and Biotechnology 81, 621–626. doi: 10.1080/14620316.2006.11512114.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi: 10.1093/bioinformatics/btp324.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Res 27, 209–220.

Martina, M., Acquadro, A., Barchi, L., Gulino, D., Brusco, F., Rabaglio, M., et al. (2022a). Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in *Anemone coronaria* L. International Journal of Molecular Sciences 23, 3126. doi: 10.3390/ijms23063126.

Martina, M., Acquadro, A., Gulino, D., Brusco, F., Rabaglio, M., Portis, E., et al. (2022b). First genetic maps development and QTL mining in *Ranuculus asiaticus* L. through ddRADseq. Frontiers in Plant Science 13. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1009206 [Accessed November 28, 2022].

Mata-Nicolás, E., Montero-Pau, J., Gimeno-Paez, E., García-Pérez, A., Ziarsolo, P., Blanca, J., et al. (2021). Discovery of a Major QTL Controlling Trichome IV Density in Tomato Using K-Seq Genotyping. Genes 12, 243. doi: 10.3390/genes12020243.

Melo, A. T. O., Bartaula, R., and Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics 17, 29. doi: 10.1186/s12859-016-0879-y.

Nakano, M., Hirakawa, H., Fukai, E., Toyoda, A., Kajitani, R., Minakuchi, Y., et al. (2021). A chromosome-level genome sequence of Chrysanthemum seticuspe, a model species for hexaploid cultivated chrysanthemum. Commun Biol 4, 1–11. doi: 10.1038/s42003-021-02704-y.

Nashima, K., Shirasawa, K., Ghelfi, A., Hirakawa, H., Isobe, S., Suyama, T., et al. (2021). Genome sequence of Hydrangea macrophylla and its application in analysis of the double flower phenotype. DNA Research 28, dsaa026. doi: 10.1093/dnares/dsaa026.

Nissim, Y., Jinggui, F., Arik, S., Neta, P., Uri, L., and Avner, C. (2004). Phenotypic and genotypic analysis of a commercial cultivar and wild populations of *Anemone cornaria*. Euphytica 136, 51–62. doi: 10.1023/B:EUPH.0000019520.19707.59.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841-842.

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics, 34(5), 867-868.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLOS ONE 7, e37135. doi: 10.1371/journal.pone.0037135.

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., et al. (2018). The Rosa genome provides new insights into the domestication of modern roses. Nat Genet 50, 772–777. doi: 10.1038/s41588-018-0110-3.

Sahu, P. K., Sao, R., Mondal, S., Vishwakarma, G., Gupta, S. K., Kumar, V., et al. (2020). Next Generation Sequencing Based Forward Genetic Approaches for Identification and Mapping of Causal Mutations in Crop Plants: A Comprehensive Review. Plants 9, 1355. doi: 10.3390/plants9101355.

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641. doi: 10.1038/nature11119.

Sim, S.-C., Van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganal, M. W., et al. (2012). High-Density SNP Genotyping of Tomato (Solanum lycopersicum L.) Reveals Patterns of Genetic Variation Due to Breeding. PLoS ONE 7, e45520. doi: 10.1371/journal.pone.0045520.

Sneath, P. H. A., and Sokal, R. R. (1973). Numerical Taxonomy: The Principles and Practice of Numerical Classification. W. H. Freeman and Co.

Sunde, J., Yıldırım, Y., Tibblin, P., and Forsman, A. (2020). Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies. Front. Genet. 11. doi: 10.3389/fgene.2020.00218.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38, 3022–3027. doi: 10.1093/molbev/msab120.

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., et al. (2020). A New Intra-Specific and High-Resolution Genetic Map of Eggplant Based on a RIL Population, and Location of QTLs Related to Plant Anthocyanin Pigmentation and Seed Vigour. Genes 11, 745. doi: 10.3390/genes11070745.

Unamba, C. I. N., Nag, A., and Sharma, R. K. (2015). Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. Frontiers in Plant Science 6. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2015.01074 [Accessed November 28, 2022].

van Geest, G., Voorrips, R. E., Esselink, D., Post, A., Visser, R. G., and Arens, P. (2017). Conclusive evidence for hexasomic inheritance in chrysanthemum based on analysis of a 183 k SNP array. BMC Genomics 18, 585. doi: 10.1186/s12864-017-4003-0.

Víquez-Zamora, M., Vosman, B., van de Geest, H., Bovy, A., Visser, R. G., Finkers, R., et al. (2013). Tomato breeding in the genomics era: insights from a SNP array. BMC Genomics 14, 354. doi: 10.1186/1471-2164-14-354.

Vukosavljev, M., Arens, P., Voorrips, R. E., van ‘t Westende, W. P., Esselink, G. D., Bourke, P. M., et al. (2016). High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68k rose SNP array. Hortic Res 3, 1–8. doi: 10.1038/hortres.2016.52.

Wright, B., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., and Grueber, C. E. (2019). From reference genomes to population genomics: comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. BMC Genomics 20, 453. doi: 10.1186/s12864-019-5806-y.

Xie, J., Zhao, H., Li, K., Zhang, R., Jiang, Y., Wang, M., et al. (2020). A chromosome-scale reference genome of Aquilegia oxysepala var. kansuensis. Hortic Res 7, 1–13. doi: 10.1038/s41438-020-0328-y.

Zhai, W., Duan, X., Zhang, R., Guo, C., Li, L., Xu, G., ... & Ren, Y. (2019). Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. Molecular Phylogenetics and Evolution, 135, 12-21.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28(24), 3326-3328.

Ziarsolo, P., Hasing, T., Hilario, R., Garcia-Carpintero, V., Blanca, J., Bombarely, A., et al. (2021). K-seq, an affordable, reliable, and open Klenow NGS-based genotyping technology. Plant Methods 17, 30. doi: 10.1186/s13007-021-00733-6.

Chapter V

The first reference genome of *Ranunculus asiaticus* L. reveals a key region related to anthocyanin pigmentation

Matteo Martina[1], Ezio Portis[1], Alberto Acquadro[1], Luciana Gaccione[1], Edoardo Vergnano[1], Marie Bolgher[2], Björn Usadel[2], Lorenzo Barchi[1,*] and Sergio Lanteri[1]

1 Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA), Plant Genetics, University of Turin, Grugliasco, Italy
2 IBG-4 Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany

Keywords: *Ranuculus asiaticus*; genome assembly; long-reads sequencing; nanopore.

Abstract

Persian buttercup (*Ranuculus asiaticus* L.; 2n=2x=16; estimated genome size: 7.6Gb) is an ornamental and perennial crop native to Asia Minor and Mediterranean basin. It is marketed both as cut flower or potted plant. In 2019, its production counted for the 0,4% of the total turnover of cut flowers and foliage, with the highest values in Italy (132 million of stem) and 300-350 ha of cultivated area. The large size of the genome as well as its high frequency of repetitive sequences and the high level of heterozygosity has made challenging the assembly of a reference genome sequence based on a short-reads sequencing approach. Thanks to the recent advances in long-reads sequencing (e.g. Oxford Nanopore Technology, ONT), we report on the first version of the *R. asiaticus* genome assembled and annotated. Furthermore, the previously developed genetic map, based on two

F1 mapping populations sharing a common male parent, was aligned with the assembled genome, allowing the identification of quantitative trait loci (QTLs) affecting anthocyanin pigmentation.

Introduction

Persian buttercup (*Ranunculus asiaticus* L.; 2n=2x=16; estimated genome size: 7.6Gb -Goepfert, 1974) a member of the Ranunculaceae family, is an outcrossing, ornamental, and perennial crop. With approximately 600 species distributed worldwide, the genus *Ranuculus* is valued for its applications in traditional medicine and ornamental purposes. The species originated from Asia Minor and the Mediterranean basin and was introduced to Western Europe in the sixteenth century. Extensive cultivation as an ornamental plant in Europe during the eighteenth century led to the development of over 500 distinctive ornamental varieties obtained from intensive breeding programs in the nineteenth century. Currently, Persian buttercup production is concentrated in Italy, France, the Netherlands, Israel, South Africa, California, and Japan, with smaller production niches in Ecuador, Ethiopia, Kenya, Tunisia, and Turkey. This species represents approximately 0.4% of the cut bulb market. Italy is the primary global producer, over area ranging from 300 to 350 hectares and with a production of 132 million plants annually (Beruto et al., 2018). *R. asiaticus* life cycle involves active growth during the cold season and a dormant phase in the hotter and drier periods of summer. Underground rhizomes ensure its survival and facilitate germination in the following season. The species can be propagated by seeds, however this practice is not commonly applied, particularly in Mediterranean climates, due to the inability to yield plants in the economically advantageous winter season, unless costly cultivation

systems, such as soil cooling, are implemented. Thus, the primary propagation method involves dividing the rhizomes produced by the plants prior to entering dormancy during the summer season. Rhizome production is influenced by the size of the rosette leaf habit, with larger rosettes producing more rhizomes. The harvested rhizomes undergo drying and subsequent commercialization.

Companies in Italy, Netherlands, France, and Japan are leading the market, and their breeding activities are focused on improving traits relevant to produce cut flowers obtained from bulbs. Flower traits, including diameter, morphology, and color, are of particular interest to respond to a market that is always looking for novelty. As an example, the cultivar "Pon Pon" is renowned for its fringed petals and striking bicolored appearance and currently accounts for approximately 30% of commercially available *Ranuculus* cultivars. Additionally, achieving high bulb yield, early production, and prolonged shelf life are important aspects of cut bulb production. *R. asiaticus* is also marketed as a potted plant, which requires the targeting to additional traits such as compact plant structure and an extended flowering period. Conventional breeding techniques, including the crossing of parental plants and subsequent selection of offspring, are primarily used to produce new varieties. In vitro propagation techniques also play a significant role, enabling the clonal propagation of plants with traits of interest, identified in the segregating populations and ensuring the commercialization of stable genotypes in the short to medium term (Beruto et al., 2018).

To meet the increasing market demand and enhance breeding efficiency, a deeper understanding of buttercup genetics and the development of molecular breeding techniques are crucial. Recent progress has been made by

Martina et al. (2022a) in constructing molecular-genetic maps of *R. asiaticus*. The double digest restriction-site associated DNA (ddRAD) approach was applied for genotyping two F1 mapping populations, whose female parents were a 'Pon Pon' and a 'Double flower' plants, while the common male parental ('Cipro') was a genotype producing a simple flower. This approach made it possible to identify eight linkage groups, corresponding to the haploid chromosome number of the species (Figure 1). In addition, a major locus associated with purple pigmentation in buttercup flowers was identified through a QTL analysis, providing promising application for marker-assisted breeding strategies. These findings highlight the potential of molecular tools and genetic mapping in enhancing breeding efforts for superior varieties.

Due to the complexity and size of the *Ranuculus* genome, a reference sequence has not been released until now. In this study, we aimed at developing the first draft genome of *R. asiaticus* using the latest Oxford Nanopore long-read sequencing protocols. Long-read sequencing allows for higher contiguity of the assembled genome, facilitating functional gene annotation and the development of tools for genetic map construction. The advancement in understanding buttercup genetics and the harnessing of molecular tools, may thus contribute to streamlining the breeding process and contribute to meet the growing demand for superior varieties in the market.

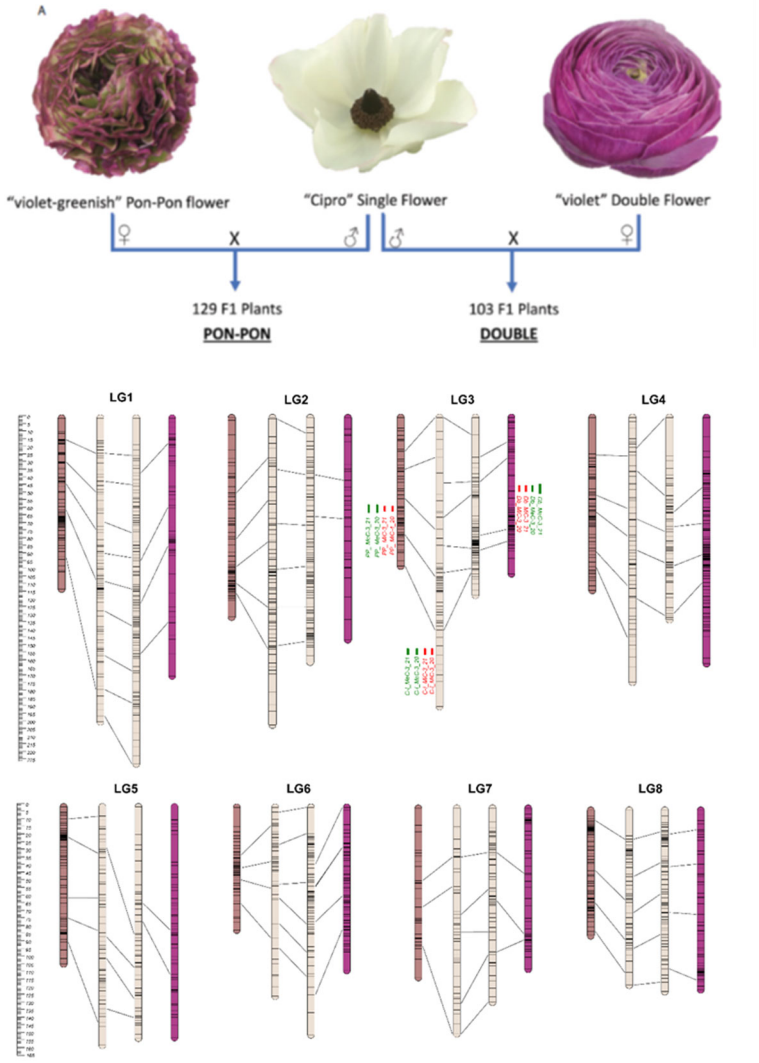


Figure 1 The first genetic map in *R. asiaticus* L., published by Martina et al. (2022a). (A) Crossing scheme for the development of the two mapping populations (PON-PON and DOUBLE); (B) Eight linkage groups (LG; corresponding to the haploid chromosome number of the species) were identified within each parental map. Four maps per LG are presented, representing, from the left to the right: the three parental lines (Pon Pon, Cipro-I, Cipro-II, and Double flowered). QTL regions for flower pigmentation are reported as green (Main External Color) and red (Main Internal Color) bars on the side of the associated LGs.

Methods

Plant Material

A clone of the cultivar 'Niveo' was selected to obtainthe first *R. asiaticus* genome sequence. Plants were obtained from one year old rhizomes and were grown in a cold greenhouse between October and February (Camporosso, IM, Italy. 43.79, 7.63), according to standard cultivation practices. Young plants were etiolated for three days in the dark for optimizing the HMW DNA extraction procedure. Leaves were harvested and used as fresh starting material for the DNA extraction.

HMW DNA Extraction

HMW DNA Extraction was carried out using a modified SDS DNA extraction procedure, as reported by Russo et al. (2022). Briefly, 100 mg of leaf material was ground in liquid nitrogen and the powder was immediately transferred to a sterile 2 ml plastic tube using a chilled metal spatula. Subsequently, 600 µl of the pre-warmed (65°C) SDS lysis buffer was added to the tube containing the ground plant material. The sample was vortexed for 3–5 seconds and incubated in a thermomixer with gentle agitation (400 rpm) at 55°C for 20 minutes. This step aimed to inactivate DNases and remove polyphenols that could bind DNA. Following the incubation, 4 µl of 100 mg/ml DNase-free RNase A (Qiagen, Germantown, MD, United States) were added to the sample, which was then incubated for an additional 10 minutes at 55°C. To precipitate proteins and polysaccharides that form complexes with SDS, 200 µl of 5 M potassium acetate (KAc) were added to the sample, and the solution was mixed by gently inverting the tube 25 times. The sample was subjected to a phenol/chloroform extraction for further purification. First, under a fume hood, 800 µl of a phenol:chloroform:isoamyl alcohol mixture (25:24:1 v/v, pH 8) was

added to the sample, which was then incubated for 10 minutes at room temperature (RT) with gentle agitation on a tube rotator at 20 rpm. Subsequently, the sample was centrifuged for 10 minutes at 10,000 × g at RT to separate the aqueous phase. The resulting supernatant was carefully transferred to a new 2 ml tube using a wide-bore pipette tip with a volume of 1,000 µl to prevent DNA shearing. A second purification step was initiated by adding 800 µl of chloroform:isoamyl alcohol (24:1 v/v), followed by another 10-minute incubation at room temperature and 20 rpm, as well as centrifugation at 10,000 × g for 10 minutes. Finally, the supernatant was transferred to a new 2 ml tube, and 10% of 5M sodium acetate were added to help precipitation, followed by 1 volume of chilled absolute ethanol. The nucleic acid was then precipitate by a centrifugation step at 8,000 x g for 10 minutes and the obtained pellet was washed for two consecutive times with 70% ethanol. Finally, after drying the pellet, the sample was eluted in 100ul of TE.

Genome Sequencing and Assembly

Genomic DNA was visualized on an agarose gel for assessing shearing. DNA was size-selected using BluePippin PAC20KB cassette (Sage Science) and prepared based on 'Oxford Nanopore Technologies' standard ligation sequencing kit SQK-LSK114. FLO-MIN114 (R10.4.1) flowcells were used for sequencing on the MinIon Mk1C platform.

After sequencing, the raw signal intensity data was used for both simplex and duplex base-calling using dorado (https://github.com/nanoporetech/dorado - version 0.3.0) from Oxford Nanopore Technologies, using super accuracy models for 400bps sequencing, and reads with a mean qscore (quality) higher than 9 were retained for genome assembly.

Raw nanopore sequencing reads were assembled with the genome assembler Flye (https://github.com/fenderglass/Flye), and the initial draft assembly was polished using the Medaka software (https://github.com/nanoporetech/medaka) from Oxford Nanopore Technologies.

SSR-Mining

The obtained assembly was used for SSR mining. Scaffolds were chopped into manageable pieces using SciRoKo tool (Kofler et al., 2007- v3.4; https://kofler.or.at/bioinformatics/SciRoKo, accessed on 22 June 2023), and perfect, compound, and imperfect SSRs were identified in silico using the SciRoKo pipeline. A minimum of four repetitions together with a minimum length of 15 nt were requested. Any sequence was considered as a perfect SSR when a motif was repeated at least fifteen times (1 nt motif), eight times (2 nt), five times (3 nt), or four times (4–6 nt), allowing for only one mismatch. For compound repeats, the maximum default interruption (spacer) length was set at 100 bp.

Genome annotation and ddRADseq data mapping

The obtained assembly was annotated using the deep-learning software Helixer (Stiehler et al., 2021). Briefly, this tool employs pre-trained Deep Neural Networks to predict genic class and phase on a base-wise level using DNA sequences as input. These predictions are further processed by a Hidden Markov Model to generate the primary gene model. Assembly quality was assessed using the BUSCO pipeline, using eudicots dataset as reference (Simão et al., 2015). Furthermore, the previously published catalog (Martina et al., 2022a), based on ddRADseq data, was aligned with the assembled genome using minimap2 (Li, 2018), allowing the identification of genomic regions for anthocyanin pigmentation.

Results and Discussion

Draft Genome Assembly and Annotation

Since *R. asiaticus* is a highly heterozygous species, the sequence divergence between alleles in a diploid genotype may hinder a reliable contig assembly of its genome sequence. In order to overcome this hurdle, we applied long-reads DNA sequencing. Overall, 65 Gb (~8X coverage) of cleaned reads were generated and used as input for genome assembly. The obtained draft assembly consisted of 211,118 scaffolds (N50 = 57,914 bp) for a total genome size of 5,1 Gb, representing ~70% of the expected genome size for this species.

Successively, we annotated the 'Niveo' assembly with the Helixer pipeline (Stiehler et al., 2021), identifying and overall number of 112,507 genes. Such number of genes is not expected in plants genome and might be attributed to Helixer difficulties in resolving gene annotation in fragmented contigs. For sure, classical annotation approaches, such as the RepeatMaker pipeline, might lead to a more coherent result. However, according to the completeness of the analysis performed using BUSCO (Simão et al., 2015), 75% of the expected genes are already presents in the assembled genome. For sure, an implementation is required, and higher coverage and scaffolding technology will be necessary to improve overall completeness and contiguity, but in our opinion the present assembly already represents a satisfactory representation of the Persian buttercup genome.

The SSR Content of the Persian Buttercup Genome

n the assembled genome, a total of 101,420 perfect SSR motifs (density of 19,88 SSR/Mb) and 93,280 imperfect SSR motifs were identified, including 25,899 compound SSRs. The identified microsatellites were grouped according to the number of repeats of each motif, as can be observed in Table 1.

Table 1. Frequency of the main identified SSR motifs (considering sequence complementary).

Repeats	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	>20	Total
A/T	-	-	-	-	-	-	-	-	-	-	-	5253	4023	3032	2388	1875	1566	5209	23346
C/G	-	-	-	-	-	-	-	-	-	-	-	1758	1116	745	417	278	167	327	4808
AC/GT	-	-	-	-	1549	899	591	431	323	272	205	194	165	150	120	112	113	2657	7781
AG/CT	-	-	-	-	5404	3372	2158	1428	1105	969	840	836	746	699	703	663	652	15942	35517
AT/AT	-	-	-	-	5836	4208	3056	2149	1489	1090	915	777	663	632	581	455	509	5788	28148
CG/CG	-	-	-	-	33	11	7	3	7	4	3	5	1	2	2	2	2	28	110
AAG/CTT	-	11285	2707	920	453	282	198	162	125	116	107	106	105	69	121	66	97	2105	19024
AAC/GTT	-	3288	1331	700	379	291	193	133	111	84	78	77	55	58	54	48	39	1513	8432
ATC/ATG	-	3842	1050	379	212	139	87	71	48	39	20	32	21	20	32	22	16	587	6617
AAT/ATT	-	3199	1151	543	322	182	131	78	78	50	47	32	34	22	22	27	17	158	6093
AGC/CTG	-	3129	1104	231	75	36	18	8	2	3	1	1	-	-	1	-	-	1	4610
AAAT/ATTT	2548	505	124	32	16	6	5	1	1	-	1	-	-	-	-	-	-	0	3239
AAAG/CTTT	1088	212	117	66	45	37	35	38	37	22	23	16	13	14	9	7	5	36	1820
AATG/ATTC	567	167	87	45	33	30	24	25	22	10	14	10	13	14	10	10	13	163	1257
AGAT/ATCT	642	125	66	22	19	25	16	21	13	21	8	11	10	13	13	14	15	197	1251
AAAC/GTTT	826	126	74	35	23	11	12	11	7	2	8	3	3	4	2	2	3	4	1156
AAAAT/ATTTT	2712	259	72	20	8	1	1	1	1	2	-	-	-	-	-	-	-	1	3078
AAAAG/CTTTT	356	72	21	16	5	2	9	1	2	-	1	-	-	-	-	1	-	0	486
AACCC/GGGTT	281	59	16	16	5	1	-	-	-	1	-	-	-	-	-	-	-	0	379
AACAC/GTGTT	264	45	19	2	1	3	-	2	-	-	1	-	-	1	-	1	1	1	341
AAGGG/CCCTT	133	93	6	3	1	1	-	-	-	-	-	-	-	-	-	-	-	0	237
ATATCC/ATATGG	897	335	182	103	106	118	92	76	91	82	71	66	74	60	60	59	50	258	2780
AAACAC/GTGTTT	149	106	75	70	38	51	37	34	28	32	20	24	27	22	27	18	22	466	1246
AAATAC/ATTTGT	138	85	61	40	22	22	26	22	17	24	21	19	12	11	9	15	10	265	819
AAAGTC/ACTTTG	448	130	66	29	23	27	12	6	3	5	5	3	3	1	1	-	-	1	763
AAGCCT/AGGCTT	401	150	54	49	17	15	12	11	7	4	4	10	8	1	3	3	1	2	752
Other Repeats	11138	9032	3090	1626	989	674	517	444	362	293	265	246	199	170	146	154	128	1137	30610

Six classes of perfect SSRs were evaluated (from mono- to hexanucleotide) for their abundance in the assembled genome. Dinucleotides were the most abundant, in accordance with what has been previously reported in literature (Tóth et al., 2000; Mun et al., 2006; Scaglione et al., 2009; Cheng et al., 2016; Portis et al., 2016; Sahu and Chattopadhyay, 2017; Portis et al., 2018; Patil et al., 2021; Martina et al., 2022b), representing 36.75% of the identified SSRs. Trinucleotides were the second most abundant class (27.92%), followed by mononucleotides (14.96%). Tetra-, Penta- and Hexa- covered the remaining percentage and showed analogous frequency ranging from 4.10 to 9.06% (Figure 2). The most represented dinucleotide motifs, AG/CT, AT/AT, and AC/GT, accounted respectively for 49.63%, 39.33%, and 10.87% (Figure 2), while CG/GC motifs were approximately absent. Within the trinucleotide repeat motifs, the most abundant were AAG/CTT, accounting for 34.98%, followed by AAC/GTT for 15.50% (Figure 2).

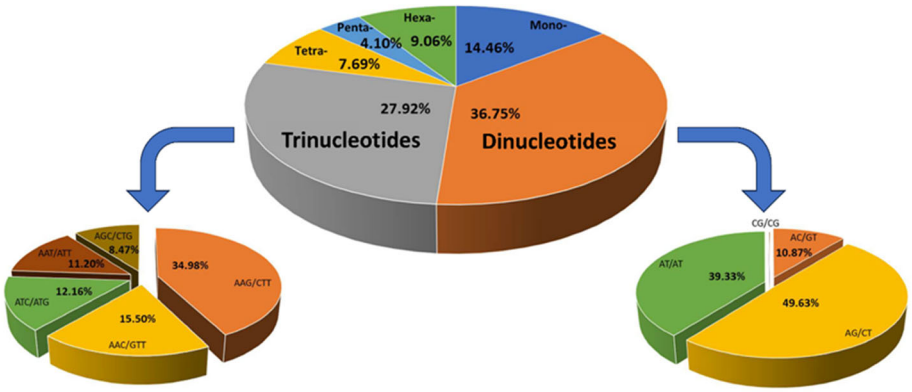


Figure 2. Microsatellites distribution in the poppy anemone genome. (a) Percentage distribution of the most frequent classes of SSRs; (b) dinucleotides motifs e and (c) main trinucleotides motifs identified by SciRoKo.

GBS data mapping and identification of potential candidate genes for *Ranuculus* pigmentation

The mock genome (catalog) that was used in Martina et al. (2022a) for ddRAD-seq alignment was mapped on the obtained assembly using minimap2, allowing genome anchoring of the linkage map previously published. Catolog's scaffolds position were leveraged to the annotated scaffolds in the new assembly, and markers position was converted consequently, allowing the identification of the genomic regions associated with the main QTLs for anthocyanins highlighted in the two F1 population (Table 2).

Table 2 QTLs detected in the mapping population by Martina et al. 2022a for Main internal Color (MiC) and Main external Color (MeC). The identified contigs are reported, as well as the percentage of explained variability.

Map	Trait	Year	Name	Group	Catalog Contig	Assembly Contig	PVE
PP	MiC	20	*PP_MiC-3_20*	LG3	S0004445	contig_170159	48
		21	*PP_MiC-3_21*	LG3	S0004445	contig_170159	67
	MeC	20	*PP_MeC-3_20*	LG3	S0004445	contig_170159	49
		21	*PP_MeC-3_21*	LG3	S0004445	contig_170159	72
C-I	MiC	20	*C-I_MiC-3_20*	LG3	S0038600	contig_252602	58
		21	*C-I_MiC-3_21*	LG3	S0038600	contig_252602	79
	MeC	20	*C-I_MeC-3_20*	LG3	S0038600	contig_252602	66
		21	*C-I_MeC-3_21*	LG3	S0038600	contig_252602	84
Db	MiC	20	*Db_MiC-3_20*	LG3	S0004445	contig_170159	32
		21	*Db_MiC-3_21*	LG3	S0004445	contig_170159	45
	MeC	20	*Db_MeC-3_20*	LG3	S0004445	contig_170159	59
		21	*Db_MeC-3_21*	LG3	S0004445	contig_170159	59

By annotating the identified contigs using the Augustus software (Stanke and Morgenstern, 2005), three transposable elements were identified in the two regions, namely a PREDICTED: Transposon, a Homeobox protein knotted-1-like 3 isoform X3, and a Transposon Ty3-G Gag-Pol polyprotein. Such genomic elements, even when not in LTR regions, have been reported to play a role in anthocyanin expression in many species (Huang et al., 2019; Jung et al., 2019). By increasing the contiguity of our genome assembly (by scaffolding the genome using linkage maps data and by applying HiC technology), the identified region might reveal functional genes involved in anthocyanin biosynthesis, depicting the role of the reported QTLs.

Conclusion

Persian buttercup has resulted in a wide range of cultivars, distinguishable by flower traits such as size, shape, color. Conventional breeding techniques and in vitro micro-propagation have been instrumental in developing premium varieties and cloning genotypes with desired traits. However, to meet the increasing market demand and improve breeding efficiency, a deeper understanding of buttercup genetics and the application of molecular tools is pivotal. The construction of genetic maps and the identification of genes, QTLs and molecular markers underlying key traits will help accelerate the breeding process and will lead to develop varieties with new flower traits, compact plant structure, extended flowering period, and other desirable characteristics such as tolerance to unfavorable environmental conditions and resistance to diseases.

Additionally, the availability of the first genome of *R. asiaticus* based on long-read sequencing technology represents a further key advance for supporting the accurate sequencing and assembly at chromosome-level of the species, facilitates functional gene annotation and contributes to the validation and fine mapping of the already mapped QTL narrowing down the genomic regions governing target traits.

References

Beruto, M., Rabaglio, M., Viglione, S., Van Labeke, M.-C., Dhooghe, E., 2018. *Ranuculus*, in: Van Huylenbroeck, J. (Ed.), Ornamental Crops, Handbook of Plant Breeding. Springer International Publishing, Cham, pp. 649–671. https://doi.org/10.1007/978-3-319-90698-0_25

Cheng, J., Zhao, Z., Li, B., Qin, C., Wu, Z., Trejo-Saavedra, D.L., Luo, X., Cui, J., Rivera-Bustamante, R.F., Li, S., Hu, K., 2016. A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum. Sci Rep 6, 18919. https://doi.org/10.1038/srep18919

Goepfert, D., 1974. Karyotypes and DNA content in species of *Ranuculus* L. and related genera. Botaniska notiser.

Huang, D., Yuan, Y., Tang, Z., Huang, Y., Kang, C., Deng, X., Xu, Q., 2019. Retrotransposon promoter of Ruby1 controls both light- and cold-induced accumulation of anthocyanins in blood orange. Plant, Cell & Environment 42, 3092–3104. https://doi.org/10.1111/pce.13609

Jung, S., Venkatesh, J., Kang, M.-Y., Kwon, J.-K., Kang, B.-C., 2019. A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in Capsicum annuum. Plant Sci 287, 110181. https://doi.org/10.1016/j.plantsci.2019.110181

Kofler, R., Schlötterer, C., Lelley, T., 2007. SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics 23, 1683–1685. https://doi.org/10.1093/bioinformatics/btm157

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Martina, M., Acquadro, A., Gulino, D., Brusco, F., Rabaglio, M., Portis, E., Lanteri, S., 2022a. First genetic maps development and QTL mining in *Ranuculus asiaticus* L. through ddRADseq. Frontiers in Plant Science 13.

Martina, M., Acquadro, A., Barchi, L., Gulino, D., Brusco, F., Rabaglio, M., Portis, F., Portis, E., Lanteri, S., 2022b. Genome-Wide Survey and Development of the First Microsatellite Markers Database (*AnCorDb*) in *Anemone cornaria* L. International Journal of Molecular Sciences 23, 3126. https://doi.org/10.3390/ijms23063126

Mun, J.-H., Dong-Jin, K., Hong-Kyu, C., Gish, J., al, et, 2006. Distribution of Microsatellites in the Genome of Medicago truncatula: A Resource of Genetic Markers That Integrate Genetic and Physical Maps. Genetics 172, 2541–55.

Patil, P.G., Singh, N.V., Bohra, A., Raghavendra, K.P., Mane, R., Mundewadikar, D.M., Babu, K.D., Sharma, J., 2021. Comprehensive Characterization and Validation of Chromosome-Specific Highly Polymorphic SSR Markers from Pomegranate (Punica granatum L.) cv. Tunisia Genome. Front. Plant Sci. 12. https://doi.org/10.3389/fpls.2021.645055

Portis, E., Lanteri, S., Barchi, L., Portis, F., Valente, L., Toppino, L., Rotino, G.L., Acquadro, A., 2018. Comprehensive Characterization of Simple Sequence Repeats in Eggplant (Solanum melongena L.) Genome and Construction of a Web Resource. Front. Plant Sci. 9. https://doi.org/10.3389/fpls.2018.00401

Portis, E., Portis, F., Valente, L., Moglia, A., Barchi, L., Lanteri, S., Acquadro, A., 2016. A Genome-Wide Survey of the Microsatellite Content of the Globe Artichoke Genome and the Development of a Web-Based Database. PLoS One 11, e0162841. https://doi.org/10.1371/journal.pone.0162841

Sahu, K.K., Chattopadhyay, D., 2017. Genome-wide sequence variations between wild and cultivated tomato species revisited by whole genome sequence mapping. BMC Genomics 18, 430. https://doi.org/10.1186/s12864-017-3822-3

Scaglione, D., Acquadro, A., Portis, E., Taylor, C.A., Lanteri, S., Knapp, S.J., 2009. Ontology and diversity of transcript-associated microsatellites mined from a globe artichoke EST database. BMC Genomics 10, 454. https://doi.org/10.1186/1471-2164-10-454

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research 33, W465–W467. https://doi.org/10.1093/nar/gki458

Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A.P.M., Denton, A.K., 2021. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics 36, 5291–5298. https://doi.org/10.1093/bioinformatics/btaa1044

Tóth, G., Gáspári, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10, 967–981. https://doi.org/10.1101/gr.10.7.96

Closing remarks and future perspectives

In this thesis, the genetic knowledge of two important ornamental crops in Italy, namely *Anemone coronaria* and *Ranunculus asiaticus*, has been significantly advanced. The aim of this research was to enhance the understanding of the genetic background of these species and provide valuable insights for the development of targeted breeding programs. This thesis has been divided into five chapters, each contributing to the overall goal. In this final chapter, the main findings are summarized, and their implications discussed, in the light of the significance of the research, and present and future perspectives for further exploration in the field of ornamental plant breeding.

Chapter I focused on *Anemone coronaria* (poppy anemone) and it presented the development of the first draft genome of this perennial plant with a large genome size. The establishment of the "*Anemone coronaria* Microsatellite DataBase" (*AnCorDb*) provided a valuable resource of SSR markers, which were successfully validated and can applied for various purposes such as cultivar fingerprinting, assessing intra-cultivar variability, and supporting applications like Breeding Rights disputes, genetic mapping, marker-assisted breeding (MAS), and phylogenetic studies. This research significantly expanded our understanding of the genetic diversity in this species.

Chapter II introduced a technique for discriminating between double-haploid (DH) plants and somatic embryogenesis-derived plants in *Anemone coronaria*. This method facilitated the identification of DH plants at early developmental stages, contributing to the production of F_1 hybrids and the selection of homozygous plants in breeding programs. The application of microsatellites in this context provided a practical and efficient tool for

enhancing the efficiency of breeding programs and the development of improved cultivars.

Chapter III focused on *Ranunculus asiatic*us, commonly known as persian buttercup. The development of molecular-genetic maps using a double digest restriction-site associated DNA (ddRAD) approach allowed for the identification of major quantitative trait loci (QTLs) associated with purple pigmentation of the flower. These maps, constructed based on two F_1 mapping populations, provided valuable insights into the genetic basis of pigmentation traits and paved the way for future marker-assisted selection and breeding efforts.

Chapter IV introduced the application of a low-cost Klenow NGS-based protocol, termed K-seq, for diversity analysis in persian buttercup and poppy anemone. This innovative approach enabled the identification of thousands of single nucleotide polymorphisms (SNPs) and provided a comprehensive assessment of the genetic diversity of commercial varieties. By comparing the results with a previously published SSR-based fingerprinting in poppy anemone (from Chapter I), the efficiency of K-seq for genotyping complex genetic backgrounds was demonstrated. This research broadened our understanding of genetic diversity and laid the foundation for improved breeding strategies in ornamental plants.

Finally, ***Chapter V*** presented the successful assembly of the first version of the *R. asiaticus* genome. This milestone achievement provided researchers with a valuable tool for further investigations into the molecular basis of different trait of interest in the species. By aligning the genetic maps developed in Chapter III with the assembled genome, specific genomic regions influencing pigmentation were identified, opening new avenues

for understanding the underlying mechanisms and facilitating future breeding efforts.

These findings have practical implications for the development of improved cultivars, efficient breeding programs, and the preservation of genetic diversity. Moreover, the availability of the draft genomes and molecular markers will facilitate further research, such as functional genomics, transcriptomics, and comparative genomics, providing a deeper understanding of these species and their potential applications.

Looking ahead, several promising avenues for future research in this field are identified. Firstly, the utilization of advanced genomic technologies, such as long-read sequencing and chromosome conformation capture (e.g., Hi-C), can contribute to improving the quality and completeness of the genome assemblies. This would enable more accurate and comprehensive identification of key genomic regions related to important traits beyond anthocyanin pigmentation. Additionally, the integration of genomic data with phenotypic and environmental information through approaches like genome-wide association studies (GWAS) and genomic selection could enhance the efficiency of breeding programs, enabling the development of tailored varieties with improved traits. Furthermore, exploring the functional significance of identified QTLs and candidate genes associated with specific traits, including flower quality, disease resistance, and flowering time, would provide a deeper understanding of the underlying mechanisms and potential targets for genetic manipulation. Functional validation experiments, such as gene expression studies and gene editing techniques like CRISPR/Cas9, can shed light on the precise roles of these genes and their potential applications in breeding programs. Moreover, fostering col-

laboration and knowledge exchange among researchers, breeders, and industry stakeholders is crucial for the successful translation of genetic research into practical applications. Partnerships with commercial producers and breeders can facilitate the implementation of marker-assisted selection, the exchange of germplasm resources, and the development of improved cultivars tailored to specific market demands.

By combining cutting-edge genetic tools with traditional breeding approaches, we can enhance the efficiency, precision, and sustainability of ornamental plant breeding, meeting the demands of consumers and ensuring the future success of the poppy anemone and persian buttercup industries in Italy, particularly in the Ligurian region.