



## Article

# Genomic Resource Development for Hydrangea (*Hydrangea macrophylla* (Thunb.) Ser.)—A Transcriptome Assembly and a High-Density Genetic Linkage Map

Xingbo Wu<sup>1,2</sup>, Amanda M. Hulse-Kemp<sup>1,3</sup>, Phillip A. Wadl<sup>4</sup>, Zach Smith<sup>5</sup>, Keithanne Mockaitis<sup>6</sup>, Margaret E. Staton<sup>7</sup>, Timothy A. Rinehart<sup>8</sup> and Lisa W. Alexander<sup>2,\*</sup>

<sup>1</sup> Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695, USA; xwu33@ncsu.edu (X.W.); amanda.hulse-kemp@usda.gov (A.M.H.-K.)

<sup>2</sup> USDA-ARS, U.S. National Arboretum, Floral and Nursery Plants Research Unit, Otis L. Floyd Nursery Research Center, McMinnville, TN 37110, USA

<sup>3</sup> USDA-ARS, Genomics and Bioinformatics Research Unit, Raleigh, NC 27695, USA

<sup>4</sup> USDA-ARS, U. S. Vegetable Laboratory, Charleston, SC 29414, USA; phillip.wadl@usda.gov

<sup>5</sup> Beckman Coulter Life Sciences, Indianapolis, IN 46268, USA; ZSMITH@beckman.com

<sup>6</sup> Root Cause Scientific Inc., Indianapolis, IN 46268, USA; km13@cornell.edu

<sup>7</sup> Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA; mstaton1@utk.edu

<sup>8</sup> USDA-ARS, Crop Production and Protection, Beltsville, MD 20705, USA; tim.rinehart@usda.gov

\* Correspondence: lisa.alexander@usda.gov



**Citation:** Wu, X.; Hulse-Kemp, A.M.; Wadl, P.A.; Smith, Z.; Mockaitis, K.; Staton, M.E.; Rinehart, T.A.; Alexander, L.W. Genomic Resource Development for Hydrangea (*Hydrangea macrophylla* (Thunb.) Ser.)—A Transcriptome Assembly and a High-Density Genetic Linkage Map. *Horticulturae* **2021**, *7*, 25. <https://doi.org/10.3390/horticulturae7020025>

Academic Editor: Johan Van Huylenbroeck

Received: 11 January 2021

Accepted: 2 February 2021

Published: 5 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** Hydrangea (*Hydrangea macrophylla*) is an important ornamental crop that has been cultivated for more than 300 years. Despite the economic importance, genetic studies for hydrangea have been limited by the lack of genetic resources. Genetic linkage maps and subsequent trait mapping are essential tools to identify and make markers available for marker-assisted breeding. A transcriptomic study was performed on two important cultivars, Veitchii and Endless Summer, to discover simple sequence repeat (SSR) markers and an F<sub>1</sub> population based on the cross 'Veitchii' × 'Endless Summer' was established for genetic linkage map construction. Genotyping by sequencing (GBS) was performed on the mapping population along with SSR genotyping. From an analysis of 42,682 putative transcripts, 8780 SSRs were identified and 1535 were validated in the mapping parents. A total of 267 polymorphic SSRs were selected for linkage map construction. The GBS yielded 3923 high quality single nucleotide polymorphisms (SNPs) in the mapping population, resulting in a total of 4190 markers that were used to generate maps for each parent and a consensus map. The consensus linkage map contained 1767 positioned markers (146 SSRs and 1621 SNPs), spanned 1383.4 centiMorgans (cM), and was comprised of 18 linkage groups, with an average mapping interval of 0.8 cM. The transcriptome information and large-scale marker development in this study greatly expanded the genetic resources that are available for hydrangea. The high-density genetic linkage maps presented here will serve as an important foundation for quantitative trait loci mapping, map-based gene cloning, and marker-assisted selection of *H. macrophylla*.

**Keywords:** *Hydrangea macrophylla*; SSR; SNP; linkage map



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hydrangea is one of the most widely cultivated plants in the horticulture industry and is grown for floriculture, nursery, landscape, and cut flower markets. Introduced as a cultivated ornamental into Europe from Asia around 1788 by Sir Joseph Banks [1], *Hydrangea macrophylla* is now represented by more than 500 extant cultivars in the United States and Europe. Among the twenty-three *Hydrangea* species that are recognized, less than half are common in cultivation and the most economically important among these by far is *H. macrophylla* [2]. In recent years, hydrangeas have become a dominant product

in the flowering shrub market in the United States (US), worth 91.3 million US dollars in 2014 [3]. The economic value has stimulated an increasing demand for specialized hydrangeas, such as the remontant type (blooming more than once per year), and thus has energized hydrangea breeding in the US.

As a typical woody plant, hydrangea breeding requires a long generation time in a conventional breeding process. Traits that are most important to *H. macrophylla* breeding are unique inflorescence architecture and flower color variation for all markets, floral induction (reblooming for landscape use and year-round floral production for florist's hydrangeas), drought and sun tolerance in landscape breeding, and powdery mildew resistance for greenhouse and landscape grown plants. Selections in hydrangea breeding historically have focused mainly on novel flower types that can be chosen in the first generation for incremental improvements [4]. An additional, current goal of hydrangea breeding is to introduce resistance to existing and emerging diseases and pests for nursery producers [5]. As a highly heterozygous horticultural crop, hydrangea breeding is mainly based on phenotypic selection, which is a long and difficult process, particularly when coupled with long generation times [4].

Conventional breeding can be improved and generation interval times can be reduced by incorporating marker-assisted selection (MAS) or marker-assisted breeding. One prerequisite of MAS is the development of genomic resources, such as genetic markers, that can be used to identify markers linked to the traits of interest to segregate populations through genetic mapping analyses [6]. Codominant markers such as microsatellites (simple sequence repeats, SSRs) and single nucleotide polymorphisms (SNPs) are markers of choice for such an application as they occur randomly throughout the genome in relatively abundant levels that can allow for estimation of additive and dominance allelic effects. However, little genetic information is currently known for *H. macrophylla*. Development of genetic information has been difficult given its highly heterozygous genomic composition, differing ploidy levels with base chromosome number ( $n = 18$ ), and large diploid genome size (~2.2 Gbp) [7].

A previous transcriptomic study focusing on aluminum stress of hydrangea discovered thousands of SSRs; however, none were validated [8,9]. Only 39 SSR markers were available for hydrangea through 2010 [10–12]. Even with 226 newly developed SSR markers in 2018 [13], the total number of SSRs is still under 300, which is too low for most genetic studies in hydrangea. The insufficient number of molecular markers and unavailability of any genotyping platform has largely hampered the ability to perform linkage and association mapping for molecular-trait dissection studies in hydrangea. To date, only one linkage map was reported in hydrangea, but with low SSR marker density [13], which is much lower than the number of linkage maps reported in roses (eight) [14], peony (three) [15], and chrysanthemum (three) [16]. In addition to providing tools for MAS, high-density linkage maps could also be used to assist genome assembly in scaffolding and quality assessments [17]. Further marker development and high-density genetic linkage maps are needed for potential genetic improvement as well as genomic studies in hydrangea.

Transcriptome sequencing is an efficient and economical way to capture sequences of the genic regions in plants with complex, large genomes. Specialty crops such as *H. macrophylla* benefit from this approach, which creates massive expression datasets that include candidate coding sequences that can be used to produce an abundance of markers. Recent advances in the next-generation sequencing (NGS) methods have also made detection of SNPs, the most abundant marker type in the largest numbers, possible through high-throughput sequence-based genotyping. Sequence-based genotyping is a powerful tool for plant breeding and is especially helpful for those plant species that lack genome information [18]. Genotyping by sequencing (GBS) allows for the detection and genotyping of tens of thousands of SNPs in many individuals simultaneously, resulting in an unparalleled cost per data point when screening for codominant polymorphisms in large panels and for constructing highly saturated genetic maps [19]. GBS has been employed for linkage mapping in woody ornamentals such as roses [20] and apples [21]. In hydrangea,

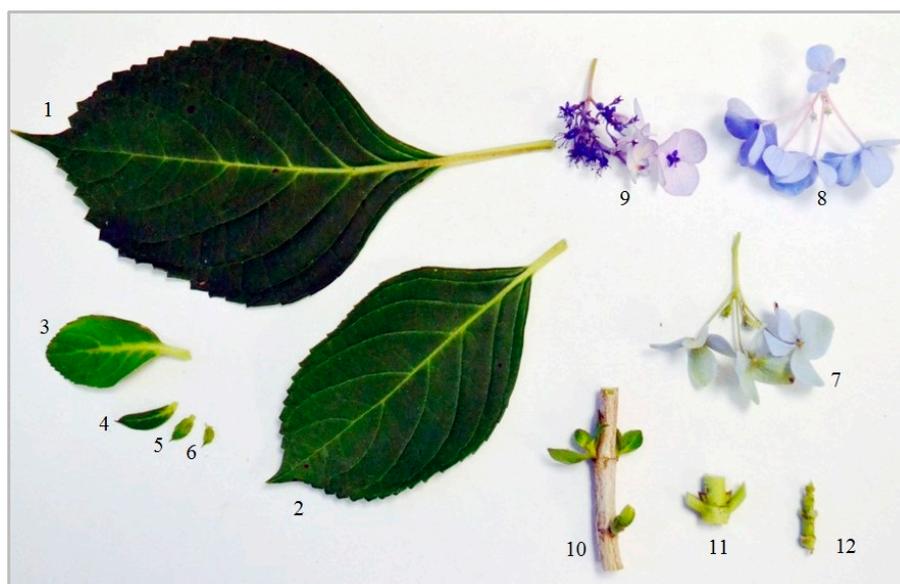
it has been used for genetic diversity [22] and association mapping studies [23,24], but no linkage map was constructed with this method.

Genomic resource development is urgently needed for future genetic mapping and quantitative genetics studies in *H. macrophylla*. Genomics approaches to hydrangea breeding offer opportunities to make immediate advances in trait improvement, especially for complex traits such as environmental tolerance and disease resistance. The objectives of this study were to generate genomic resources through transcriptome sequencing and genotyping by sequencing for *H. macrophylla*, and to demonstrate the utility of these resources with a high-density linkage map.

## 2. Materials and Methods

### 2.1. Plant Materials and Sample Preparation

Two important hydrangea cultivars, Veitchii and Endless Summer, were chosen as representatives for genomic resource development. 'Veitchii' is a lacecap cultivar used for the cut flower market and is tolerant to powdery mildew and *Cercospora* leaf spot but only flowers once a year (once-flowering), while 'Endless Summer' (formerly also identified as 'Bailmer') is a mophead cultivar which can flower twice or more per year (remontant) and is popular in the floriculture market but susceptible to powdery mildew [25,26]. In order to get broader representation in gene expression and better coverage of the genome, four tissue types (leaves, stems, buds, and open flowers) were collected for transcriptome sequencing from three year old 'Veitchii' and 'Endless Summer' plants grown in three gallon plastic containers (24.1 × 27.9 × 27.9 cm) containing composted pine bark medium under drip irrigation and 50% shade at the United States Department of Agriculture Agricultural Research Service (USDA-ARS) Thad Cochran Southern Horticultural Laboratory, Poplarville, MS, USA (30° 50' 02.68" N 89° 32' 50.54" W) (Figure 1). The tissues were collected in spring 2010 and immediately frozen in liquid nitrogen and stored at −80 °C. Total RNA was isolated from the tissues using the RNeasy Plant Mini Kit (Qiagen; Valencia, CA, USA) according to the manufacturer's instructions. The isolated RNA was treated with DNase I (New England BioLabs; Ipswich, MA, USA). RNA quantity and quality were determined by fluorimetry and by microfluidic electrophoresis (BioAnalyzer, Agilent; Santa Clara, CA, USA).



**Figure 1.** Example of tissue types sampled from a single plant of 'Veitchii' and 'Endless Summer' for transcriptome sequencing. 1–6: leaf tissues collected at different developmental stages. 7–9: flower tissues collected at different flowering times; 10–12: stem tissues collected from different positions. Bud tissues are not listed in this figure.

For genotyping by sequencing, an F<sub>1</sub> mapping population comprising 90 individuals, as well as the two parents, 'Veitchii' and 'Endless Summer', were used to construct a genetic linkage map. Trifoliolate leaves of both parents and the F<sub>1</sub> population were collected into 2 mL Eppendorf tubes with lysate buffer and ground in a homogenizer (FastPrep-24 5G; MP Biomedicals, Santa Ana, CA, USA). Genomic DNA was isolated from the lysed tissue using the DNeasy Plant Mini Kit, followed by RNase treatment (Qiagen, Valencia, CA, USA). Nucleic acid was evaluated in 1% agarose gel and quantified using a spectrometer (NanoDrop 2000; Thermo Scientific, Wilmington, DE, USA). Enzyme activity was evaluated using 6 U of EcoRI on 300 ng DNA of eight randomly selected samples for 2 h at 37 °C and separation using gel electrophoresis on a 1% agarose gel. A lyophilized aliquot of 1.5 mg DNA was prepared for each sample for subsequent SSR genotyping and GBS.

## 2.2. Transcriptome Sequencing, Assembly, and Analyses

Total RNA from each cultivar was pooled in approximately equimolar amounts prior to construction of a customized cDNA library enriched for polyA and 5' cap containing transcripts, with the double-stranded cDNA library intermediate partially normalized by DSN treatment (Evrogen, Moscow, Russia) to reduce representation of the high abundance transcripts. Prior to adaptor ligation, the library was sheared by nebulization to optimize template length distribution for the sequencing system and assessed on a Bioanalyzer DNA7500 chip (Agilent; Santa Clara, CA, USA). According to the manufacturer (Roche/454 Sequencing, Branford, CT, USA) emulsion PCR was performed and the library sequenced to below saturation on a GS FLX Titanium PicoTitre™ plate to 800 cycles.

## 2.3. Transcriptome Assembly

Sequence reads were cleaned of adaptor sequences [27] and assembled into isotigs using NEWBLER v2.3 GS De Novo Assembler [28] (Roche 454 Sequencing, Branford, CT, USA) with default parameters for cDNA (40 bp overlap; 90% identity). TransDecoder [29] was used to predict open reading frames (ORFs) from the isotig assembly. Predictions were improved by including the Pfam matching information. The matches were generated by comparing the isotigs to the Pfam database version 3.1b1 [30] with HMMer hmmscan version 3.1b2 ("HMMER" 2016) [31]. The nucleotide sequences of the isotigs were compared to the Uniprot Swissprot database and the plant proteins from Uniprot Trembl [32] with BLAST version 2.2.26. Hits were filtered at an e-value of 0.00001. InterProScan version 5.4–47.0 [33] was run for all protein sequences. Assemblies were assessed for completeness with BUSCO version 1.2 and the early access plant dataset [34]. Raw sequence data were uploaded to NCBI under BioProject PRJNA661039.

## 2.4. EST-SSR Identification and Genotyping

The assembled *H. macrophylla* transcriptome was searched for microsatellite motifs or SSRs using the program SSR Finder [35]. For the searches and comparison of microsatellites, SSRs were defined as being mononucleotide repeats (MNRs) with  $\geq 10$  repeats and di- (DNRs), tri- (TNRs), tetra- (TTNRs), penta-, and hexanucleotide repeats  $\geq 6$  repeats; criteria for composite SSRs was an interval of bases  $\leq 100$  bp.

Primers were designed and used to amplify SSR loci in genomic DNA from 'Veitchii' and 'Endless Summer' using a modified three-primer protocol originally described in Waldbieser et al. [36]. In summary, DNA was extracted from 1 cm<sup>2</sup> pieces of fresh leaf tissue using Qiagen Plant Mini Kit (Qiagen). Genomic DNA was quantified using a NanoDrop Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). Forward primers were 5' tailed with the sequence 5'-CAGTTTTCCCAGTCACGAC-3' to permit product labeling, and reverse primers were tailed at the 5' end with the sequence 5'-GTTT-3' to promote non-template adenylation [37]. A primer, 5'-CAGTTTTCCCAGTCACGAC-3', was labeled with 6-carboxy-fluorescein (FAM) (Integrated DNA Technologies, Coralville, IA, USA) and added to the amplification reaction, which included Advantage2 Taq DNA Polymerase (Clontech, Mountain View, CA, USA) according to previously published proto-

cols [6]. Fluorescence-labeled PCR fragments were visualized by automated capillary gel electrophoresis on an ABI3100-Avant (Applied Biosystems, Foster City, CA, USA) using a ROX-500 size standard. GeneMapper version 4.0 (Applied Biosystems) was used to recognize and size peaks. Loci that produced high quality data, that were polymorphic between 'Veitchii' and 'Endless Summer', and that produced allele combinations compatible with the pseudo test cross strategy were selected for linkage mapping with the mapping population.

### 2.5. Genotyping by Sequencing and SNP Identification

GBS was performed at the University of Wisconsin-Madison Biotechnology Center (UWBC; Madison, WI, USA) following the protocols described in Elshire et al. [19]. The 4-base cutter enzyme, ApeKI (G \* CWGC), was found to be efficient in producing fragments that were < 500 bp in sample DNA digestion and was selected for library preparation. The digested DNA fragments were then ligated with optimal barcodes and common adapters followed by standard PCR. Single-end sequencing (100 bp) of the library was performed on the Illumina HiSeq 2000 (Illumina Inc. San Diego, CA, USA). Raw sequence data for the total library were uploaded to NCBI under BioProject PRJNA656177.

The reference-free Java program TASSEL-UNEAK [38] was utilized for GBS data analysis given there was no reference genome available for *H. macrophylla*. Briefly, the barcodes were first removed from the raw sequence reads. The remaining sequences were analyzed by the TASSEL-UNEAK pipeline to further produce 64 bp length sequence tags and aligned with each other based on unique sequence tags. SNP discovery was performed directly within pairs of matched sequence tags and filtered through network analysis. The minimum tag count required for output was set to 5. To reduce the sequencing error, the error tolerance rate was set to 0.03. Raw SNPs were filtered to ensure high-quality genotype calling using Trait Analysis by Association, Evolution and Linkage (TASSEL) version 5.0 [39]. SNPs that had more than 10% missing data and minimum minor allele frequency (MAF) of 0.05 were excluded from the dataset, individuals that possessed more than 20% missing data were filtered out. SNPs that were polymorphic between the parents of the mapping population were retained for linkage mapping.

### 2.6. Linkage Map Construction

The allele data for all selected markers were tested for segregation distortion and genotypic data similarities. The linkage map was created using the cross-pollinated (CP) model in JoinMap v5.0 [40]. In the case of SSRs, five segregation types ( $lm \times ll$ ,  $nn \times np$ ,  $hk \times hk$ ,  $ef \times eg$ , and  $ab \times cd$ ) were used based on the software instructions, while three segregation types ( $lm \times ll$ ,  $nn \times np$ , and  $hk \times hk$ ) were used to code SNPs. Distortion of each marker from their expected frequencies was observed by a goodness-of-fit (Chi-square) test. Identical progenies, identical markers, and severely distorted loci ( $p < 0.001$ ) were observed and excluded from mapping construction. Linkage groups (LGs) were constructed using the regression mapping method with default parameters. The independence LOD (logarithm of the odds) score was set from 2 to 15 for grouping purposes and groups were selected at the LOD score that produced group numbers equivalent to the number of chromosomes. Map distances within individual linkage groups were converted to centiMorgans (cM) using the Kosambi mapping function.

The two-step strategy was used to construct the linkage map. First, alleles segregating only in the female ( $lm \times ll$ ) or male ( $nn \times np$ ) parents were separated as individual datasets using the embed grouping function. Individual maps for each of the parents were constructed using the parameters mentioned above. For consensus map construction, the maternal and paternal maps were integrated by combining linkage groups sharing the same SNP alleles classified as ( $hk \times hk$ ). The map construction, marker order, and distance for each LG in the consensus map were calculated using the same parameters as those for the parental maps. The consensus map was graphically represented using MapChart V2.2 [41].

### 3. Results

#### 3.1. Transcriptome Analysis and EST-SSR Identification

The hydrangea libraries were constructed from normalized cDNA and generated a total of 1.17 million reads (444 Mb) that were de novo assembled into 42,682 isotigs, putative transcripts, comprising 24,201 isogroups, and putative unigene sets of transcripts. Open reading frames (ORFs) were found in 42,620 isotigs (>99%). Functional annotation of these data was performed using BLAST to publicly available databases. This resulted in matches for 67% of the isotigs against the SwissProt database [42] and for 85% of the isotigs against plant proteins from the Trembl database [32]. InterProScan [33] was used to gain further functional information in protein sequences derived from the ORFs, estimating 199,709 additional putative functions for 34,389 of the isotigs. InterProScan also yielded 171,059 gene ontology (GO) term [43] assignments with assignments of 1679 unique terms to 21,818 isotigs. To assess the completeness of the transcriptome, the software BUSCO was utilized. Based on the analysis, the *H. macrophylla* assembly includes coverage of 92% of the 956 BUSCO conserved genes (429 as complete single copy transcripts, 358 as complete but multiple transcripts, and 91 as fragmented genes).

EST-SSR development identified 30,169 new SSR loci for *H. macrophylla*. Primers could be designed for 5986 loci and 2794 loci in the contigs and singletons (non-assembled reads), respectively. The total number of new SSR loci identified was 8780 (Supplementary Table S1). The majority of these loci contained tri-nucleotide repeats (52.82%) as would be expected for coding regions. The next largest class of repeat motifs was di-nucleotide (37.15%) (Table 1). The remaining (<10%) motifs were larger than tri-nucleotides.

**Table 1.** The length and distribution frequency of simple sequence repeat (SSR) types in *H. macrophylla*.

Type \ Length	4	5	6	7	8	10	>10	Total	Distribution Frequency (%)
Dinucleotide	0	0	1040	692	861	395	274	3262	37.15
Trinucleotide	2675	997	434	246	202	60	24	4638	52.82
Tetranucleotide	226	57	23	15	9	2	0	332	3.78
Pentanucleotide	125	25	6	5	0	0	0	161	1.83
Hexanucleotide	241	93	29	12	6	0	0	381	4.34
Other	6	0	0	0	0	0	0	6	0.07

#### 3.2. EST-SSR Genotyping

A selection of the identified SSR loci (1535) were evaluated with genomic DNA from 'Veitchii' and 'Endless Summer'. Of these, 1226 produced high-quality data, but 35% of those loci were monomorphic and therefore not suitable for genetic mapping purposes. A total of 267 SSRs were selected for genetic mapping of the F<sub>1</sub> population. Among them, 115 (43.1%) showed maternal segregation (Im × Il) and 98 (36.7%) showed paternal segregation (nn × np), while 53 (19.9%) were heterozygous in both parents (44 with four distinct alleles, ab × cd type, and 9 with 3 alleles, ef × eg type), and one (0.3%) was partially informative as both parents had the same heterozygous genotype (hk × hk). Information of the 267 markers are highlighted in Supplementary Table S1.

#### 3.3. GBS-SNP Discovery

The GBS library generated an average of 2.7 million reads per F<sub>1</sub> progeny and the TASSEL-UNEAK pipeline analysis resulted in 376,153 raw SNPs. By removing SNPs that showed MAF < 0.05 and more than 10% missing data, the total SNP number was reduced to 12,214. Extra criteria were also used to assure the SNP quality as follows: SNPs that were missing in either parent or homozygous in both parents were removed from further analyses. Furthermore, SNPs showing segregation patterns in the progeny that were not in agreement with the parental genotypes were discarded from the mapping process. Finally, a group of 3923 high quality informative SNPs were obtained, consisting of 1160 SNPs (29.6%) segregating in the female parent (Im × Il), 881 SNPs (22.5%) segregating in the

male parent (nn × np), and 1882 SNPs (47.9%) that were heterozygous and segregating in both parents (hk × hk). The SNP sequence information is presented in Supplementary Dataset S1.

### 3.4. Construction of Genetic Linkage Maps

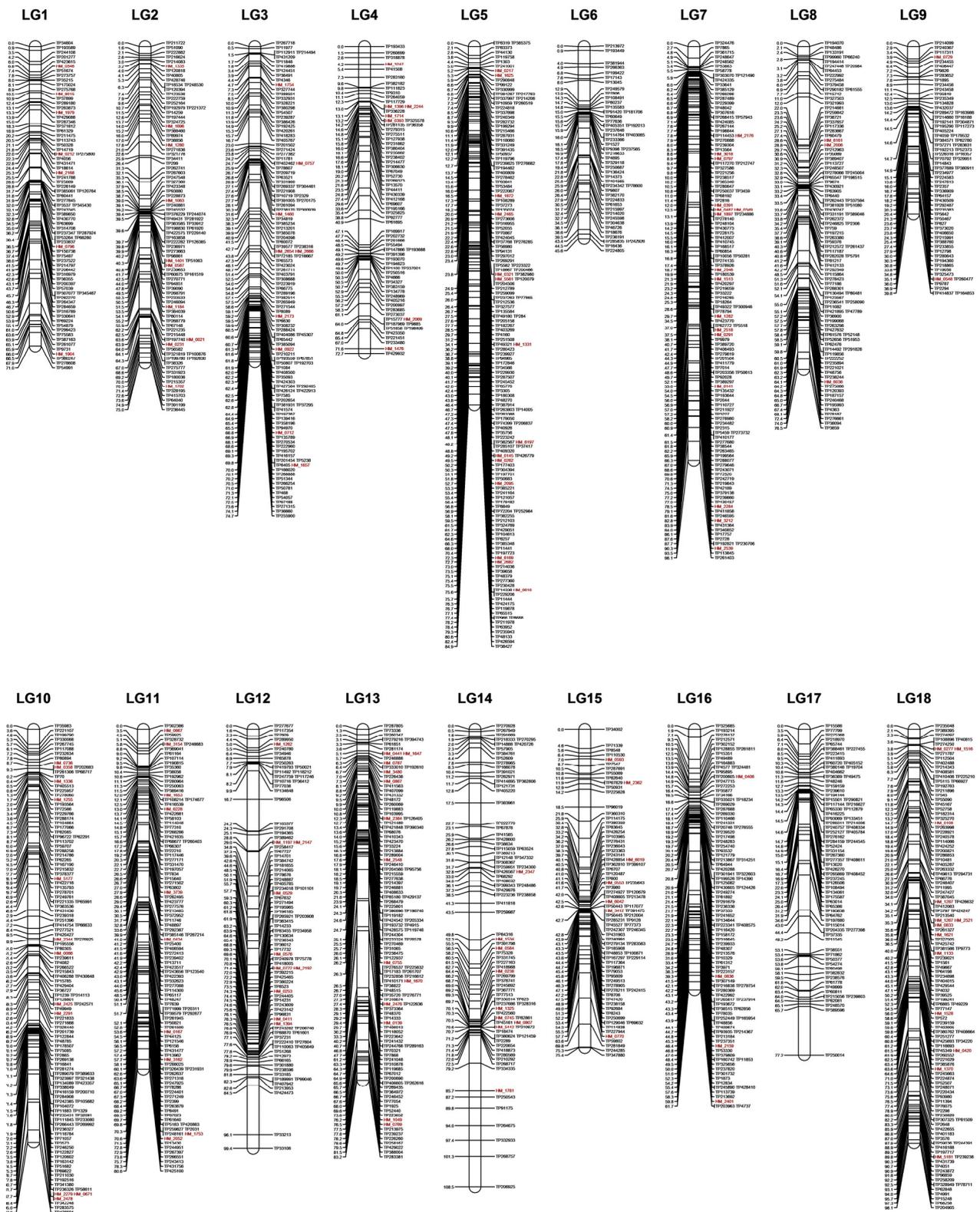
A total of 4190 high quality markers were used for linkage mapping, including 267 SSR markers developed from transcriptome sequencing and 3923 SNP markers developed from GBS (Supplementary Table S2). Of them, 472 markers were identical to other markers and were excluded from further analyses. No identical samples were found in the mapping population.

The maternal and paternal maps were determined at LOD of 8, with each map containing 18 LGs. The maternal map consisted of 1686 markers with a total genetic length of 1396.9 cM, while the paternal map consisted of 1396 markers with a total genetic length of 1400.8 cM (Supplementary Table S3). The maternal and paternal map visualizations can be found in Supplementary Figure S1. Generally, the paternal map tended to have smaller sizes for each of the linkage groups, which resulted in an overall smaller size than the maternal and consensus maps. When comparing the markers of individual LGs between consensus map and parent maps, the maternal map contributed most of the markers on LGs 5, 8, 9, 10, 14, 15, and 17 of the consensus map, while the paternal map contributed largely to LGs 2 and 7 of the consensus map. The LGs 1, 3, 4, 6, 11, 12, 13, 16, and 18 of the consensus map had a similar number of markers derived from the corresponding LGs of each parent map.

A consensus map was generated that contained a total of 1767 markers, consisting of 1621 SNP and 146 SSR markers. The total length of the consensus map was 1383.4 cM with LG 14 (108.5 cM) being the largest group and LG 9 (39.8 cM) being the smallest group (Table 2). The number of markers per linkage group varied from 152 (LG 5) to 49 (LG 6), with an average of 98 markers per linkage group. Most LGs were made of both SNP and SSR markers, with LG 17 being an exception and only containing SNPs. Four large gaps, that were more than 10 cM, were observed on LG 8 (10.0 cM), LG 12 (11.5 cM), LG 13 (13.5 cM), and LG 17 (11.5 cM). Overall, the average distance between markers was 0.8 cM. The distribution of SNPs and SSRs on consensus map LGs is shown in Figure 2.

**Table 2.** Summary of the *Hydrangea* consensus map.

Linkage Group (LG)	Marker Count (#)	SNP Count (#)	SSR Count (#)	Length (cM)	Average Marker Interval (cM)	Largest Gap (cM)
1	77	70	7	71.0	0.9	5.9
2	93	83	10	75.0	0.8	6.3
3	120	111	9	74.7	0.6	6.4
4	72	65	7	72.7	1.0	6.2
5	152	138	14	84.9	0.6	3.0
6	49	49	0	45.6	0.9	4.9
7	125	109	16	98.1	0.8	4.8
8	101	98	3	76.5	0.8	10.0
9	68	66	2	39.8	0.6	3.0
10	126	114	12	97.9	0.8	6.6
11	108	98	10	80.6	0.7	6.3
12	92	82	10	99.4	1.1	11.5
13	114	101	13	83.2	0.7	13.5
14	86	77	9	108.5	1.3	7.2
15	83	76	7	75.4	0.9	6.1
16	101	97	4	61.7	0.6	5.2
17	78	78	0	77.2	1.0	11.5
18	122	109	13	61.3	0.5	6.4
Total	1767	1621	146	1383.4	0.8	13.5



**Figure 2.** SSR- and single nucleotide polymorphism (SNP)-based high-density genetic consensus linkage map of *H. macrophylla*. A total of 1621 SNP markers (black) and 146 SSR markers (red) are shown in 18 linkage groups.

#### 4. Discussion

Recently, hydrangea has been gaining popularity as an attractive plant for public and private gardens and has become an important ornamental crop for nursery growers. MAS could directly accelerate the breeding of new hydrangea cultivars to meet the consumer needs for new and novel plants [44]. Previous studies on hydrangeas mainly focused on phenotypic evaluations on a physiologic level, which have not led to an acceleration in the hydrangea breeding progress. The lack of genomic resources is a common restraint to molecular breeding in nursery crops, and *H. macrophylla* is no exception to this trend. The rapid development of sequencing technology provides opportunities of marker development not only in a fast and convenient way, but also at low cost. In order to expand the genomic resources of *H. macrophylla*, the present study implemented a transcriptomics study for SSR development, as well as a sequencing-based marker discovery method, GBS, to identify SNP markers in the population developed by crossing these two cultivars for genetic linkage mapping.

A total of 8780 SSRs were identified through the transcriptomics of important hydrangea cultivars, providing the most extensive set of genic sequences so far for this species. Among the identified SSRs, 1535 of them were tested and 267 were found to be polymorphic. SSR markers are robust tools for genetic mapping and molecular breeding applications. The polymorphic SSRs validated in this study significantly increased the availability of useful markers for genetic studies in *H. macrophylla*. In addition, high transferability of SSRs were reported between hydrangea species [10,11,45], so the discovered SSRs here should immediately enable multiple molecular applications in the numerous related species.

GBS offers an efficient way to decipher hydrangea genetics as it can discover SNPs and facilitate genotyping at the same time [19]. GBS has been widely used in popular ornamental crops, such as roses [20], for linkage mapping and subsequent quantitative trait loci (QTL) analyses. In hydrangea, valuable traits such as inflorescence shape, double flower, and recurrent blooming are thought to be quantitatively inherited and suitable for QTL studies to identify markers for MAS [13,23]. The genotyping of two hydrangea cultivars and their F<sub>1</sub> progeny identified 3923 SNP markers through a GBS approach, providing the possibility to facilitate construction of a high-density linkage map for future QTL studies in hydrangea.

Genetic linkage maps are powerful tools to locate genes and QTL regions. As a perennial woody plant, linkage map construction in hydrangea is more complicated than that of homozygous inbred annual plants due to its high heterozygosity and self-incompatibility. F<sub>1</sub> full-sibling populations have become a choice for linkage maps in woody plant species. In this study, an F<sub>1</sub> population was derived from the cross of two important cultivars, Veitchii and Endless Summer, and was used to build the linkage map of hydrangea with 1767 molecular markers. The genetic map contained 18 linkage groups, corresponding to the haploid number of chromosomes identified in the karyotype of diploid *H. macrophylla* (2n = 36) [7]. To date, only one linkage map was reported in hydrangea and it included only 147 SSRs [13]. The linkage map developed in this study consists of 146 SSRs and 1621 SNPs (Table 2), with an average of 98 markers per linkage group, which was far more than the markers used in the previous study [13]. The total length (1383.4 cM) and small marker interval (0.8 cM in average) indicated the high quality of the developed map when compared to the total genetic distance of 980 cM and average marker interval of 6.8 cM in the previously reported map [13]. With more total mapped markers, a larger genetic distance, and the small marker interval, the reported linkage map represents the most saturated genetic map, and will serve as a valuable tool for future genetic and breeding study in hydrangea.

All of the linkage groups identified in the current map consist of both SSR and SNP markers. However, the selected 267 SSRs alone were not sufficient to fully construct a linkage map of hydrangea with the number of linkage groups equal to the number of chromosomes. Studies in woody ornamentals, such as roses [20] and blueberry [46], have

used SSRs as the bridge for crosstalk between genetic maps, thus, more SSRs need to be validated in order to build a high-density consensus map using multiple hydrangea populations. Several regions in one linkage group showed higher density than other regions, which may indicate the uneven marker polymorphism and recombination frequency on a certain chromosome between the mapping parents. This phenomenon has also been observed in grapes [47] and Mei [48]. Though the average genetic distance was 0.8 cM, four gaps located on LG 8, LG 12, LG 13, and LG 17 were found to be larger than 10 cM (Table 2, Figure 2). Large gaps in genetic maps are usually caused by low marker detection and polymorphism in similar chromosome segments of mapping parents or centromeric regions of the chromosome [49], which could be filled by a high-density consensus map construction using different marker types [50].

Despite the substantial value of the industry, technologies available to augment the development of new and improved ornamental plants lag far behind most other agricultural crops. Paucity of genomic resources such as reference genome sequences and comprehensive transcriptome datasets are especially problematic for breeding programs focused on woody ornamentals. Challenges posed by long juvenile periods, obligate or nearly obligate out-crossing breeding systems, inbreeding depression, high heterozygosity, and poor understanding of the genetic control of traits variation are major impediments for effective application of traditional breeding approaches to improve woody ornamental cultivars. Genome-enabled research and breeding programs would greatly accelerate the development of new durable and high-value cultivars. MAS provides the opportunity to assess young plants at the genotype level for multiple traits of interest, which can greatly reduce costs associated with growing the plants to maturity. The mapping parents used in this study, 'Veitchii' and 'Endless Summer', are two important hydrangea cultivars that have contrasting phenotypes in inflorescence type, flowering habits, powdery mildew tolerance, and drought and cold stress, and the markers developed in this study will be useful for interrogation of these traits in the future. Integration of these traits in cultivars through conventional breeding methods has proved difficult but it is likely more feasible through MAS.

## 5. Conclusions

This study performed large-scale SSR and SNP marker discovery in a  $F_1$  population through transcriptome and genotyping by sequencing for genetic map construct in hydrangea. The consensus linkage map contained 1767 positioned markers (146 SSRs and 1621 SNPs), spanned 1383.4 centiMorgans (cM), and was comprised of 18 linkage groups, with an average mapping interval of 0.8 cM. This study provided not only the most comprehensive marker resource currently available in hydrangea, but also an important foundation for QTL mapping, map-based gene cloning, and MAS of *H. macrophylla*. For traits where variation is readily available in cultivated germplasm, this work defined the strategies for improving hydrangea through conventional breeding. It is these traits where advanced genetic methods such as molecular markers could increase efficiency in long term, high risk projects such as germplasm improvement and new cultivar development.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2311-7524/7/2/25/s1>, Supplemental Figure S1—Genetic linkage maps of maternal ('Veitchii') and paternal ('Endless Summer') *H. macrophylla* cultivars. Supplementary Table S1—Details of 8780 EST-SSR markers developed from the transcriptome of hydrangea cultivars Veitchii and Endless Summer. Supplementary Table S2—Genotype information for identified and linkage-mapped SSR and SNP markers. Supplementary Table S3—Number of markers and genetic lengths of 18 LGs in maternal, paternal, and consensus maps. Supplementary Dataset S1—Information of SNP sequence discovered by genotyping by sequencing.

**Author Contributions:** L.W.A. and X.W. conceived the experiment and developed and sequenced the mapping population. K.M. and Z.S. designed and performed transcriptome sequencing. K.M., Z.S., M.E.S., T.A.R., and P.A.W. performed transcriptome analysis and developed SSR markers. X.W. and A.M.H.-K. performed linkage mapping and produced the genetic map. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the U.S. Department of Agriculture (USDA) Floriculture and Nursery Research Initiative and the Agricultural Research Service including ARS project number 6066-21310-005-00D. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendors that also may be suitable. USDA is an equal opportunity employer and provider.

**Data Availability Statement:** The transcriptome sequence reads were deposited into NCBI under PRJNA661039. The Illumina sequencing data were deposited to NCBI under BioProject PRJNA656177.

**Acknowledgments:** The authors would like to acknowledge Carrie Witcher, Benjamin Moore, and Ram Podicheti for invaluable technical support. Two anonymous reviewers provided contributions that greatly improved the manuscript. In addition, we acknowledge Indiana University and the Biotechnology Center at University of Wisconsin-Madison for their sequencing and data process services.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Wilson, E.H. The hortensias *Hydrangea macrophylla* DC. and *Hydrangea serrata* DC. *J. Arnold Arbor.* **1923**, *4*, 233–246.
- Griffiths, M. *The New Royal Horticultural Society Dictionary: Index of Garden Plants*; Macmillan Press Ltd.: London, UK, 1994.
- The 2014 Census of Horticultural Specialties*; U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS): Washington, DC, USA, 2015; Volume 3, Special Studies, Part 3, AC-12-SS-3.
- Rinehart, T.; Wadl, P.; Staton, M. An update on *Hydrangea macrophylla* breeding targets and genomics. In Proceedings of the III International Symposium on Woody Ornamentals of the Temperate Zone 1191, Minneapolis, MI, USA, 2–5 August 2016; pp. 217–224.
- Leus, L. *Breeding for Disease Resistance in Ornamentals in Ornamental Crops 97-125*; Springer: Cham, Switzerland, 2018.
- Mohan, M.; Nair, S.; Bhagwat, A.; Krishna, T.G.; Yano, M.; Bhatia, C.R.; Sasaki, T. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* **1997**, *3*, 87–103. [[CrossRef](#)]
- Cerbah, M.; Cerbah, M.; Mortreau, E.; Brown, S.; Siljak-Yakovlev, S.; Bertrand, H.; Lambert, C. Genome size variation and species relationships in the genus *Hydrangea*. *Appl. Genet.* **2001**, *103*, 45–51. [[CrossRef](#)]
- Chen, H.; Lu, C.; Jiang, H.; Peng, J. Global transcriptome analysis reveals distinct aluminum-tolerance pathways in the Al-accumulating species *Hydrangea macrophylla* and marker identification. *PLoS ONE* **2015**, *10*, e0144927. [[CrossRef](#)]
- Negishi, T.; Oshima, K.; Hattori, M.; Kanai, M.; Mano, S.; Nishimura, M.; Yoshida, K. Tonoplast-and plasma membrane-localized aquaporin-family transporters in blue hydrangea sepals of aluminum hyperaccumulating plant. *PLoS ONE* **2012**, *7*, e43189. [[CrossRef](#)]
- Rinehart, T.A.; Scheffler, B.E.; Reed, S.M. Genetic diversity estimates for the genus *Hydrangea* and development of a molecular key based on SSR. *J. Am. Soc. Hortic. Sci.* **2006**, *131*, 787–797. [[CrossRef](#)]
- Rinehart, T.A.; Scheffler, B.E.; Reed, S.M. Ploidy variation and genetic diversity in *Dichroa*. *HortScience* **2010**, *45*, 208–213. [[CrossRef](#)]
- Reed, S.M.; Rinehart, T.A. Simple sequence repeat marker analysis of genetic relationships within *Hydrangea macrophylla*. *J. Am. Soc. Hortic. Sci.* **2007**, *132*, 341–351. [[CrossRef](#)]
- Waki, T.; Kodama, M.; Akutsu, M.; Namai, K.; Iigo, M.; Kurokura, T.; Yamamoto, T.; Nashima, K.; Nakayama, M.; Yagi, M. Development of DNA Markers Linked to Double-Flower and Hortensia Traits in *Hydrangea macrophylla* (Thunb.) Ser. *Hortic. J.* **2018**, *87*, 264–273. [[CrossRef](#)]
- Li, S.; Yang, G.; Yang, S.; Just, J.; Yan, H.; Zhou, N.; Jian, H.; Wang, Q.; Chen, M.; Qiu, X.; et al. The development of a high-density genetic map significantly improves the quality of reference genome assemblies for rose. *Sci. Rep.* **2019**, *9*, 5985. [[CrossRef](#)] [[PubMed](#)]
- Yang, Y.; Sun, M.; Li, S.; Chen, Q.; Da Silva, J.A.T.; Wang, A.; Yu, X.; Wang, L. Germplasm resources and genetic breeding of *Paeonia*: A systematic review. *Hortic. Res.* **2020**, *7*, 1–19. [[CrossRef](#)]
- Song, X.; Xu, Y.; Gao, K.; Fan, G.; Zhang, F.; Deng, C.; Dai, S.; Huang, H.; Xin, H.; Li, Y. High-density genetic map construction and identification of loci controlling flower-type traits in *Chrysanthemum* (*Chrysanthemum* × *morifolium* Ramat.). *Hortic. Res.* **2020**, *7*, 1–13. [[CrossRef](#)] [[PubMed](#)]
- Walve, R.; Rastas, P.; Salmela, L. Kermit: Linkage map guided long read assembly. *Algorithms Mol. Biol.* **2019**, *14*, 8. [[CrossRef](#)]
- He, J.; Zhao, X.; Laroche, A.; Lu, Z.-X.; Liu, H.; Li, Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* **2014**, *5*, 484. [[CrossRef](#)] [[PubMed](#)]

19. 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. [[CrossRef](#)]
20. Yan, M.; Byrne, D.H.; Klein, P.E.; Yang, J.; Dong, Q.; Anderson, N. Genotyping-by-sequencing application on diploid rose and a resulting high-density SNP-based consensus map. *Hortic. Res.* **2018**, *5*, 17. [[CrossRef](#)]
21. Di Pierro, E.A.; Gianfranceschi, L.; Di Guardo, M.; Koehorst-van Putten, H.J.; Kruisselbrink, J.W.; Longhi, S.; Troggio, M.; Bianco, L.; Muranty, H.; Pagliarani, G.; et al. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for out-crossing species. *Hortic. Res.* **2016**, *3*, 1–13. [[CrossRef](#)] [[PubMed](#)]
22. Wu, X.; Alexander, L.W. Genetic Diversity and Population Structure Analysis of Bigleaf Hydrangea Using Genotyping-by-sequencing. *J. Am. Soc. Hortic. Sci.* **2019**, *144*, 257–263. [[CrossRef](#)]
23. Wu, X.; Alexander, L.W. Genome-wide association studies for inflorescence type and remotancy in *Hydrangea macrophylla*. *Hortic. Res.* **2020**, *7*, 1–9. [[CrossRef](#)]
24. Tränkner, C.; Krüger, J.; Wanke, S.; Naumann, J.; Wenke, T.; Engel, F. Rapid identification of inflorescence type markers by genotyping-by-sequencing of diploid and triploid F1 plants of *Hydrangea macrophylla*. *BMC Genet.* **2019**, *20*, 60. [[CrossRef](#)]
25. Li, Y.; Mmbaga, M.; Zhou, B.; Joshua, J.; Rotich, E.; Parikh, L. *Diseases of Hydrangea. Handbook of Florists' Crops Diseases. Handbook of Plant Disease Management*; Springer: Cham, Germany, 2016; pp. 1–19.
26. Adkins, J.A.; Dirr, M.A. Remotant flowering potential of ten *Hydrangea macrophylla* (Thunb.) Ser. cultivars. *HortScience* **2003**, *38*, 1337–1340. [[CrossRef](#)]
27. Tae, H.; Ryu, D.; Sureshchandra, S.; Choi, J.-H. ESTclean: A cleaning tool for next-gen transcriptome shotgun sequencing. *BMC Bioinf.* **2012**, *13*, 247. [[CrossRef](#)]
28. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.-J.; Chen, Z.; et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nat. Cell Biol.* **2005**, *437*, 376–380. [[CrossRef](#)] [[PubMed](#)]
29. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)]
30. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [[CrossRef](#)]
31. Finn, R.D.; Clements, J.; Arndt, W.; Miller, B.L.; Wheeler, T.J.; Schreiber, F.; Bateman, A.; Eddy, S.R. HMMER web server: 2015 update. *Nucleic Acids Res.* **2015**, *43*, W30–W38. [[CrossRef](#)]
32. Apweiler, R. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, 115D–119D. [[CrossRef](#)] [[PubMed](#)]
33. 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. [[CrossRef](#)]
34. 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. [[CrossRef](#)] [[PubMed](#)]
35. Stieneke, D.; Eujayl, I. Imperfect SSR Finder, Version 1.0. Available online: <http://ssr.nwisrl.ars.usda.gov/> (accessed on 12 April 2017).
36. Waldbieser, G.C.; Quiniou, S.M.; Karsi, A. Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. *Biotechniques* **2003**, *35*, 976–979. [[CrossRef](#)]
37. Brownstein, M.J.; Carpten, J.D.; Smith, J.R. Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *Biotechniques* **1996**, *20*, 1004–1010. [[CrossRef](#)]
38. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **2014**, *9*, e90346. [[CrossRef](#)]
39. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]
40. Van Ooijen, J. *JoinMap 4. Software for the Calculation of Genetic Linkage Maps in Experimental Populations*; Kyazma BV: Wageningen, The Netherlands, 2006; p. 33.
41. Voorrips, R. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [[CrossRef](#)]
42. Consortium, U. UniProt: A hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204–D212. [[CrossRef](#)]
43. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
44. Bailey, D.A. *Hydrangea Production*; Timber Press, Inc.: Portland, OR, USA, 1989.
45. Reed, S.M.; Jones, K.D.; Rinehart, T.A. Production and characterization of intergeneric hybrids between *Dichroa febrifuga* and *Hydrangea macrophylla*. *J. Am. Soc. Hortic. Sci.* **2008**, *133*, 84–91. [[CrossRef](#)]
46. McCallum, S.; Graham, J.; Jorgensen, L.; Rowland, L.J.; Bassil, N.V.; Hancock, J.F.; Wheeler, E.J.; Vining, K.J.; Poland, J.; Olmstead, J.W.; et al. Construction of a SNP and SSR linkage map in autotetraploid blueberry using genotyping by sequencing. *Mol. Breed.* **2016**, *36*, 41. [[CrossRef](#)]
47. Wang, N.; Fang, L.; Xin, H.; Wang, L.; Li, S. Construction of a high-density genetic map for grape using next generation restriction-site associated DNA sequencing. *BMC Plant Biol.* **2012**, *12*, 148. [[CrossRef](#)]

48. Sun, L.; Yang, W.; Zhang, Q.; Cheng, T.; Pan, H.; Xu, Z.; Zhang, J.; Chen, C. Genome-Wide Characterization and Linkage Mapping of Simple Sequence Repeats in Mei (*Prunus mume* Sieb. et Zucc.). *PLoS ONE* **2013**, *8*, e59562. [[CrossRef](#)]
49. Guajardo, V.; Solís, S.; Sagredo, B.; Gainza-Cortés, F.; Muñoz, C.; Gasic, K.; Hinrichsen, P. Construction of High Density Sweet Cherry (*Prunus avium* L.) Linkage Maps Using Microsatellite Markers and SNPs Detected by Genotyping-by-Sequencing (GBS). *PLoS ONE* **2015**, *10*, e0127750. [[CrossRef](#)] [[PubMed](#)]
50. Kirov, I.; Van Laere, K.; De Riek, J.; De Keyser, E.; Van Roy, N.; Khrustaleva, L. Anchoring Linkage Groups of the Rosa Genetic Map to Physical Chromosomes with Tyramide-FISH and EST-SNP Markers. *PLoS ONE* **2014**, *9*, e95793. [[CrossRef](#)]