

Review

Mining and Development of Novel SSR Markers Using Next Generation Sequencing (NGS) Data in Plants

Sima Taheri ^{1,*}, Thohirah Lee Abdullah ^{1,*}, Mohd Raffi Yusop ^{1,2} , Mohamed Musa Hanafi ^{2,3,4}, Mahbod Sahebi ², Parisa Azizi ² and Redmond Ramin Shamshiri ⁵

¹ Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia; mraffii@upm.edu.my

² Laboratory of Climate-Smart Food Crop Production, Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia; mmhanafi@upm.edu.my (M.M.H.); mahbod_sahebi@yahoo.com (M.S.); bahar3236@yahoo.com (P.A.)

³ Laboratory of Plantation Science and Technology, Institute of Plantation Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

⁴ Department of Land Management, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

⁵ Smart Farming Technology Research Center, Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia; raminshamshiri@upm.edu.my

* Correspondence: sima_taheri65@yahoo.com (S.T.); thohirah@upm.edu.my (T.L.A.); Tel.: +60-11-2067-5170 (S.T.)

Received: 28 December 2017; Accepted: 13 January 2018; Published: 13 February 2018

Abstract: Microsatellites, or simple sequence repeats (SSRs), are one of the most informative and multi-purpose genetic markers exploited in plant functional genomics. However, the discovery of SSRs and development using traditional methods are laborious, time-consuming, and costly. Recently, the availability of high-throughput sequencing technologies has enabled researchers to identify a substantial number of microsatellites at less cost and effort than traditional approaches. Illumina is a noteworthy transcriptome sequencing technology that is currently used in SSR marker development. Although 454 pyrosequencing datasets can be used for SSR development, this type of sequencing is no longer supported. This review aims to present an overview of the next generation sequencing, with a focus on the efficient use of de novo transcriptome sequencing (RNA-Seq) and related tools for mining and development of microsatellites in plants.

Keywords: SSR markers; de novo transcriptome; RNA-Seq; microsatellite; Illumina; short tandem repeat (STR)

1. Introduction

Advances in sequencing technologies, commonly referred to as next-generation sequencing (NGS), generate millions of sequences that can be read in a very cost-effective manner. NGS has paved the way for the large-scale discovery of genetic markers [1].

Within breeding programs, various types of molecular markers, such as random amplified polymorphic DNA (RAPD), ribosomal DNA (rDNA), inter-simple sequence repeat (ISSR), sequence characterised amplified region (SCAR), and simple sequence repeat (SSR), have been utilized [2–7]. Notably, SSRs and single nucleotide polymorphism (SNP) markers are propounded in genetic and plant breeding applications [8]. Furthermore, the advent of NGS has facilitated the development

of SSRs or microsatellites across the genome, while being quick, efficient, and cost-effective even in non-model plant populations with limited or having any background genetic information [9–11].

In recent years, generating transcriptome data through RNA sequencing have been successfully reported for SSR marker development in non-model plants with no reference genome as de novo sequencing [12]. Accordingly, microsatellite markers have several uses in marker-assisted selection (MAS), linkage mapping or quantitative trait loci (QTL) mapping, phylogenetic, positional cloning, genetic divergence appraisal, genotypic profiling, and so forth [13,14].

The following discussion aims to review the application of next generation sequencing technologies specifically de novo transcriptome sequencing (RNA-Seq) in mining and development of SSR markers for genetic research.

1.1. Importance of Microsatellites and Their Use as Genetic Markers

Microsatellites are a subcategory of tandem repeats consisting of 1–6 nucleotides in length (motifs) found in genomes of all prokaryotes and eukaryotes [15]. Among individual genotypes, the number of repeat units may vary since the tandem arrays of SSR motifs change. Accordingly, with additional repeated units, the genotypic variety also increases. Likewise, motif length also affects the number of repeats as shorter motifs contain a higher number of repeats than larger (e.g., tetranucleotide) motifs. Notwithstanding, in smaller motifs, there is a greater feasibility of genotyping errors due to slipped-strand mispairing (stuttering) during the polymerase chain reaction (PCR), while longer and perfect SSR loci display more prominent allelic fluctuation [16,17].

There are a vast number of SSR loci spread out all over the genome, specifically in the euchromatin of eukaryotes, and in coding and non-coding nuclear and organellar DNA [18]. In a comparative study of rice and *Arabidopsis thaliana*, SSR distribution has been shown to be highly organised, varying in different regions of the genes [19]. Microsatellites have been utilized liberally over previous years since they are profoundly informative with a high mutation rate per locus per generation (10^{-7} to 10^{-3}) [16], locus specificity, high intraspecific polymorphism, high reproducibility, ease of scoring, multiallelic, and frequent transpacific presence across related taxa. Additionally, the co-dominance nature of SSRs allows for the direct measurement of heterozygosity and only requires small amounts of DNA for data collection, another characteristic of SSRs (1 ng of DNA per reaction) [20–23]. Notably, they have been widely applied for different purposes, such as (1) genetic diversity; (2) discovering quantitative trait loci (QTL); (3) linkage map construction between gene and marker; (4) marker assisted selection for desired traits (MAS); (5) forensics and parentage analysis (SSRs with core repeats three to five nucleotides long are preferred); (6) cultivar DNA fingerprinting [24]; (7) genome-wide association study (GWAS); (8) gene flow estimation and crossing over rates; (9) marker assisted breeding (MAS) [25]; (10) haplotype determination; (11) harnessing heterosis; (12) germplasm characterization; and (13) genetic diagnostics, characterization of transformants, and the study of genome organization [14,26–29]. However, the high cost for SSR development, the presence of more null alleles, and the occurrence of homoplasmy are some of the weak points of microsatellites [30].

SSRs are assorted based on their source, i.e., genomic SSRs (g-SSRs) and expressed sequence tags SSRs (EST-SSRs), which are located in the coding region and are identified from transcribed RNA sequences [31]. The EST-SSRs generate higher quality patterns with almost 70% having a distinct polymorphic fragment of the supposed size [32] as opposed to 36% in g-SSRs [33]. Furthermore, generating SSR markers using express sequence tags (EST) has been accelerated through sequencing technology advancements in various plant species [34–38]. Some characteristics of EST-SSRs such as their inexpensive development, a higher level of genetic diversity, and higher transferability to related taxa, are because of the additional conservation of sequences that contain EST-SSRs, thereby making them advantageous for biodiversity studies [39]. In contrast to the EST-SSRs, genomic SSRs have less interspecific transferability because of the repeat region or degeneracy of the primer binding sites [40,41]. Although a major weak point of the EST-SSRs is the sequence redundancy that yields multiple sets of markers at the same locus, this problem can be handled by assembling the ESTs into

a unigene [41]. Accordingly, EST-SSRs markers have been developed and used in many plant species, such as rice, wheat, barley, sorghum, tomato, coffee, rubber, castor bean, and sesame [42–51].

1.2. Next-Generation Sequencing (NGS)

Since its commercial availability in 2005, next-generation sequencing (NGS) technology has assisted researchers in recent years, providing excellent opportunities for life sciences [52]. Before NGS, the development process of SSRs was labor-intensive, economically costly, and time-consuming due to the necessity of building up genomic libraries for targeted SSR motifs in creating recombinant DNA molecules using restriction enzymes for DNA fragmentation. Additionally, the cloning of DNA fragments into a vector was performed, as well as sequencing of clones carrying SSRs [11,53,54]. Secondly, one of the most significant impediments to primer design for PCR in the validation of SSR markers procedures was the necessity of background information of genome sequences containing SSR repeats [55–57]. Thirdly, successful SSR development relied strongly on the amplification of the target locus by a primer designed from a single SSR locus to generate obvious polymorphism [55]. High-throughput NGS technologies as a powerful, quick, cost-effective, and reliable tool, transformed the field of discovery and development of molecular markers by generating an enormous amount of sequence data [58–61].

There are different NGS technologies such as 454 Roche (<http://www.my454.com>) as the first commercially NGS platform that was utilized, mostly for bacterial and viral genomes. Next, there is the Illumina genome analyzer (<http://www.illumina.com>) used for complex genomes (human, plant, and mouse), ABI SOLID (<http://www.thermofisher.com/my/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html/>), Pacific Bioscience (<http://www.pacb.com/>), Ion Torrent (<http://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing.html/>), Oxford Nanopore (<http://www.nanoporetech.com>), and Qiagen GeneReader (<http://www.genereaderngs.com/>) [62,63]. In all, these NGS technologies are applied for different uses, such as for multiplex-PCR products, whole genome sequencing, de novo assembly sequencing, RNA-Seq, somatic mutation detection, methylation detection, validation of point mutations, and metagenomics [63,64]. Currently, sequencing by synthesis (e.g., Illumina) is the most widely utilized NGS platform for SSR marker development [11,29,65]. Although the 454-pyrosequencing dataset is still being used in some laboratories, it is mostly being phased out and will soon be redundant.

Illumina technology has been upgraded in recent years, revolutionizing NGS by establishing the HiSeq series (2500/3000/4000) sequencing system. The latest Illumina HiSeq 4000 sequencing system with patterned one or two flow-cells, can produce up to 100 million reads per sample. Moreover, it has a reading length of 50/75/150 bp for data yields of 210–250 Gb, 650–750 Gb, and 1300–1500 Gb per flow cell in less than 3.5 days' runtime, and with an accuracy greater than 99%, as compared to the original HiSeq and MiSeq systems (www.illumina.com). Furthermore, only Illumina can generate paired-end sequencing reads leading to high-quality sequence data due to enhancing the possibility of the alignment of the reference genome. Moreover, Illumina facilitates the detection of genomic Indels, inversions, novel transcripts, and genes. Moreover, in de novo sequencing, it can produce longer contigs by filling the gaps in the consensus sequence [66,67]. Every laboratory using the HiSeq 3000/HiSeq 4000 Systems can access the latest sequencing technology and increase their genomics power.

1.3. SSR Discovery by Transcriptome Sequencing (RNA-Seq)

SSR development can be reliant on either genomic DNA sequences or double-stranded DNA synthesised from single-strand RNA (cDNA) depending on the project objectives, the future research scheme, and the researcher's ability to manage output data [68]. Although direct sequencing using DNA instead of RNA is more straightforward, as it does not require library construction and normalization, sequence assembly, annotation, and integration of unigenes [69–73], transcriptome

sequencing (RNA-Seq) as a successful and effective approach can be used for transcriptome profiling, gene expression analysis, and the detection of functional genes [74,75]. Furthermore, it is usable for SSR mining, especially for plants without a reference genome (de novo assembly) [76–78]. Moreover, high reproducibility and few systematic differences among technical replicates make RNA-Seq data more profitable [79]. Even in non-model organisms with no reference genome, large amounts of expressed sequence data can be obtained using RNA-Seq technology [80,81], where the generated readouts of billions of bases each day from a solitary instrument can be utilized in the development of high throughput EST-SSRs [82]. Accordingly, this speeds up transcriptomes assembly, allowing for the identification of expressed genes including gene isoforms and gene products to be completed accurately and extensively [83–89]. In RNA-Seq, in the presence of a reference genome, the output reads align to a reference genome or to reference transcripts, while in the absence of reference genome or transcriptome information, it is required to map a genome-scale transcription comprised of both the transcript structure and the level of expression for each gene at any specific developmental stage [90–93]. As de novo transcriptome assembly functions independently from existing genomic sequences, it can be particularly useful for the analysis of non-model species containing large nuclear genomes, such as polyploids [85].

Transcriptome sequencing is an efficient way to generate superior resources for the vast discovery and development of SSR loci in plants and has provided an improved understanding of them (see Table 1). In a recent study, researchers developed SSR in Guar (*Cyamopsis tetragonoloba*, L. Taub.) using Illumina HiSeq 2000 technology and found 5773 SSR loci from 62,146 non-redundant unigenes. In this study, 20 primer pairs were designed and synthesised, with a total of 13 primer pairs successfully amplified in two target guar varieties, M-83 and RGC-1066. Amplification failure in the other seven SSR markers was attributed to the possibility of flanking primers extending across a splice site with a large intron or chimeric cDNA contigs [8,94]. In a study by Wei et al. (2016) [80], they identified 9933 EST-SSR markers among 39,298 unigenes in colored calla lily (*Zantedeschia rehmannii* Engl.) using an Illumina HiSeq 2000 instrument. Accordingly, out of 200 designed primer pairs, 58 were polymorphic among 21 accessions of colored calla lily [80]. In 2012, Li and colleagues performed another example using de novo transcriptome sequencing for providing EST datasets used for the development of SSR molecular markers. In that study, a total of 39,257 EST-SSRs from the rubber tree were identified using data generated by Illumina HiSeq 2000 [49]. RNA-Seq as a simple, straightforward, and reliable approach has been applied for EST-SSR development in many other species such as sesame [51], sweet potato [95], carrot [96], bamboo [97], peanut [98], pea [99], common bean [100], mungbean (*Vigna radiata*) [101], and *Hemarthria* species [89] (see Table 1).

Table 1. Developed simple sequence repeat (SSR) markers using Illumina, and 454 sequencing technologies in plants.

Species	SSR Type	No. Unigenes	NGS Technology	Total No. of Discovered SSRs	Total No. of SSR Primer Designed	Total Polymorphic SSR Primers	Reference
<i>Jatropha curcas</i>	SSR	115,611	Roche 454 Genome Sequencer	9798	262	33	[102]
Guar (<i>Cyamopsis tetragonoloba</i> , L. Taub.)	SSR	62,146	Illumina HiSeq 2000 sequencing platform	5773	20	13	[8]
Red clover (<i>Trifolium pratense</i> L.)	SSR	80,328/83,489/ 84,545/84,442	Illumina HiSeq 2000 sequencing platform	15	n/a	15	[103]
Winged bean (<i>Psophocarpus tetragonolobus</i>)	EST-SSR	97,241	Roche 454 Genome Sequencer FLX (Titanium chemistry)	12,956	2994	n/a	[104]
Colored calla lily (<i>Zantedeschia rehmannii</i> Engl.)	EST-SSR	39,298	Illumina HiSeq 2000 sequencing platform	9933	200	58	[80]
<i>Salix psammophila</i>	EST-SSR	71,458	Illumina HiSeq2500 platform	6346	168	27	[105]
Sainfoin (<i>Onobrychis vicifolia</i>)	SSR	92,772	Illumina HiSeq 2000 sequencing platform	3823	100	n/a	[106]
Two <i>Hemarthria</i> Species	SSR	137,142/77,150	Illumina HiSeqTM 2500 sequencing platform	10,888	4846	34	[89]
Oak (<i>Quercus austrocochinchinensis</i>) & (<i>Q. kerrii</i>)	SSR	49,845/50,767	Illumina MiSeq sequencing platform	13,762/13,430	5196/5021	18	[107]
<i>Dipteronia oliver</i> (Aceraceae)	SSR	99,358	Illumina HiSeq 2000 sequencing platform	12,377	4179	97	[108]
<i>Elymus sibiricus</i> L.	EST-SSR	94,458	Illumina HiSeq2000 sequencing platform	8769	500	112	[109]
<i>Argyranthemum broussonetii</i> ,		80,620		2282	30	8	
<i>Echium wildpretii</i> ,	SSR	58,526	Illumina MiSeq sequencing platform	1284	n/a	n/a	[110]
<i>Descurainia bourgaeana</i>		44,287		1972	n/a	n/a	
<i>Boea clarkeana</i> Hemsl. (<i>Boea</i> , Gesneriaceae)	EST-SSR	91,449	Illumina HiSeqTM 2000 sequencing platform	8563	436	17	[111]
<i>Diabella</i> (Caprifoliaceae)	EST-SSR	58669	Illumina HiSeqTM 2000 sequencing platform	n/a	2746	13	[112]
<i>Paris polyphylla</i> Smith	EST-SSR	56,095	Illumina HiSeq2000 sequencing platform	3853	80	9	[113]
Pummelo (<i>Citrus grandis</i> (L.) Osbeck)	SSR	57,212	Illumina HiSeq2000 sequencing platform	10,276	1174	29	[114]
Chinese walnut (<i>Juglans cathayensis</i> L.)	EST-SSR	116814	Illumina HiSeq2000 sequencing platform	22,484	62	12	[115]
Chinese cabbage (<i>Brassica rapa</i> L. ssp. <i>pekinensis</i>)	EST-SSR	51,694	Solexa/Illumina	10,420	24	17	[116]
<i>Lotus (Nelumbo nucifera)</i>	SSR	105,834	Illumina HiSeqTM 2000 sequencing platform	11,178	6568	80	[117]
<i>Carthamus tinctorius</i> L. (Safflower)	SSR	2,043,956	Illumina HiSeqTM 2000 sequencing platform	23,067	325	93	[118]
<i>Phalaenopsis aphrodite</i> subsp. <i>formosana</i>	EST-SSR	22,598	Illumina HiSeqTM 2000 sequencing platform	1439	1051	10	[119]
<i>Neolitsea sericea</i> (Lauraceae)	EST-SSR	68,624	Illumina HiSeqTM 2000 sequencing platform	13,213	1191	13	[120]
Mango (<i>Mangifera indica</i>).	SSR	66,288	Illumina HiSeq 2000 sequencing platform	106,049	84,118	90	[121]
Adzuki bean (<i>Vigna angularis</i>)	EST-SSR	112 million	Illumina HiSeq2000 sequencing platform	7947	296	38	[41]
<i>Quercus pubescens</i>	SSR	96,006	Illumina HiSeq 2000 sequencing platform	14,202	10,864	20	[122]
<i>Brassica oleracea</i> L. var. <i>capitata</i> L.	EST-SSR	34,688 and 40,947	454 GS FLX Titanium Sequencer	2405	937	116	[123]
<i>Hevea brasiliensis</i>	SSR	19,708	Roche 454 sequencing platform	1397	n/a	n/a	[124]
<i>Medicago sativa</i>	EST-SSR	54,278	Illumina HiSeqTM 2000 sequencing platform	4493	837	372	[125]
<i>Paspalum dilatatum</i> Poir.	EST-SSR	20169	GS FLX Titanium technology	2339	96	32	[126]
Red clover (<i>Trifolium pratense</i> L.)	SSR	45181	Illumina HiSeq2000 sequencing platform	3127	2193	n/a	[76]
<i>Eulaliopsis binata</i>	SSR	59,134	Illumina HiSeq 2000 sequencing platform	6681	5,723	24	[127]
Common vetch (<i>Vicia sativa</i> subsp. <i>sativa</i>)	cDNA-SSR (cSSR)	n/a	454 Pyrosequencing platform	3811	300	65	[128]
Faba bean (<i>Vicia faba</i> L.)	cDNA-SSR (cSSR)	n/a	454 Pyrosequencing platform	1729	240	55	[129]
lentil (<i>Lens culinaris</i> Medik.)	SSR	55,463	Illumina Genome Analyzer II platform	8722	5,673	23	[130]
<i>Amorphophallus</i> (Araceae)	SSR	135,822	Illumina HiSeq™ 2000 sequencing platform	19,596	10,754	205	[31]
Tea (<i>Camellia sinensis</i>)	SSR	75,531	Illumina HiSeq™ 2000 platform	12,582	2439	431	[131]

Table 1. Cont.

Species	SSR Type	No. Unigenes	NGS Technology	Total No. of Discovered SSRs	Total No. of SSR Primer Designed	Total Polymorphic SSR Primers	Reference
Faba bean (<i>Vicia faba</i> L.)	cDNA-SSR (cSSR)	n/a	454 Pyrosequencing platform	1729	240	55	[129]
Tea (<i>Camellia sinensis</i>)	EST-SSR	25,637	Roche/454 Genome Sequencer FLX Instrument	3767	100	36	[132]
Rubber tree (<i>Hevea brasiliensis</i> Muell. Arg.)	EST-SSR	22,756	Illumina HiSeq™ 2000 sequencing platform	39,257	110	61	[49]
Peanut (<i>Arachis hypogaea</i> L.)	SSR	59,077	Solexa HiSeq™ 2000 sequencing platform	3919	160	65	[79]
<i>Bituminaria bituminosa</i>	SSR	3838	Roche 454 sequencing platform	3419	240	21	[133]
(<i>Sesamum indicum</i> L.)	EST-SSR	86,222	Illumina HiSeq2000 sequencing platform	7702	50	40	[51]
Pigeonpea [<i>Cajanus cajan</i> (L.) Millspaugh]	SSR	43,324	454 GS-FLX sequencing platform	3771	2877	20	[72]
Chickpea (<i>Cicer arietinum</i> L.)	SSR and SNP	103,215	Roche/454 and Illumina/Solexa	26,252	3172	42	[71]
Lentil (<i>Lens culinaris</i> Medik.)	EST-SSR	25,592	Roche 454 GS-FLX Titanium platform	1.38 × 10 ⁶	2393	51	[134]
<i>Hevea brasiliensis</i>	EST-SSR	113,313	454 pyrosequencing platform	17,819	430	47	[135]
Sweet potato (<i>Ipomoea batatas</i>)	cDNA SSR (cSSR)	56,516	Illumina paired-end sequencing platform	4114	100	92	[86]

2. Overview of the Process of SSR Development through Transcriptome de Novo Assembly Using the Illumina Platform

The transcriptome de novo assembly process includes RNA extraction, cDNA library construction, sequencing, data filtering and quality control, de novo assembly, unigene annotation, SSR search and primer design, and marker validation (see Figure 1). After extraction of total RNA and its treatment with DNase I, Oligo(dT) is used to isolate mRNA. mRNAs are fragmented by fragmentation buffer and are used as a template for cDNA synthesis. Then, short fragments are purified and resolved with elution buffer (EB) for end reparation and single nucleotide A (adenine) addition. Next, adaptors are conjoined to short fragments, and suitable fragments are selected for PCR amplification. After quantification and qualification of the sample library during the QC steps, the library is then sequenced using an Illumina HiSeq 2000/2500/3000/4000, or another sequencer if necessary. After sequencing, the low-quality, adaptor-polluted, and high content of unknown base (N) reads will be filtered to obtain clean reads and are then saved in the FASTQ format [136]. Next, de novo assembly is performed with the clean reads to obtain the unigenes.

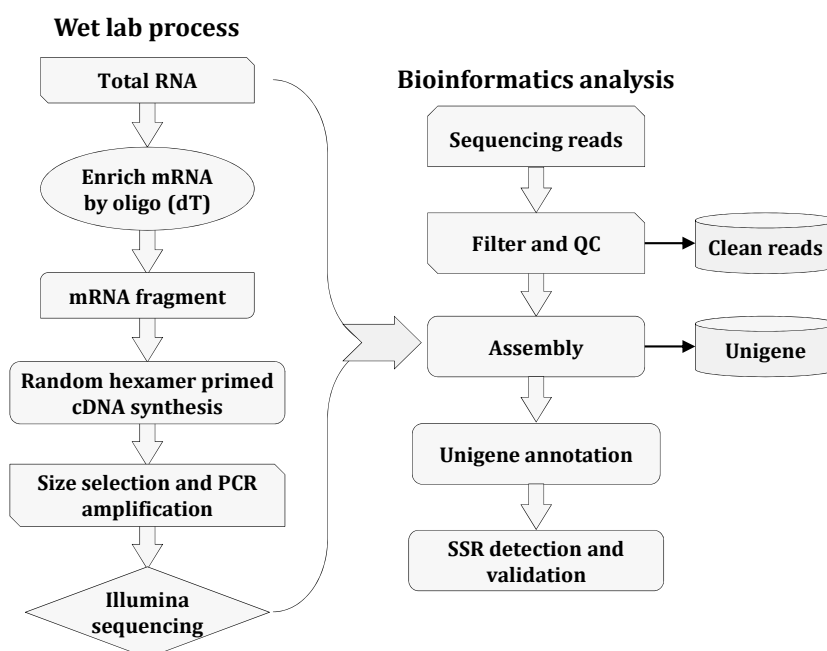


Figure 1. Schematic overview of a de novo transcriptome sequencing and assembly process.

2.1. de Novo Assembly

There are several tools used for de novo assembly of RNA-Seq reads, such as Multiple-k [137], Rnnotator [138], Trans-ABYSS [139], Velvet-Oases [140], and SOAPdenovo-Trans (<http://soap.genomics.org.cn/SOAPdenovo-Trans.html>). A tool that has recently been gaining popularity for de novo assembly of transcriptomes is Trinity [141,142], which generates individual de Bruijn graphs for sequence reads. Accordingly, each de Bruijn graph indicates the transcriptional complexity of a certain gene or locus, which is processed separately to obtain full-length splicing isoforms and to tease apart transcripts extracted from paralogous genes. Moreover, this process distinguishes Trinity from other available transcriptome de novo assembly tools. Additionally, Trinity sequentially applies three software applications, namely, Inchworm, Chrysalis, and Butterfly, to manage the enormous quantity of reads [138,143]. The process is briefly described below:

1. **Inchworm:** assembles the reads set into the unique sequences of transcripts by extending the sequences with the most abundant k-mers and then only reports the unique portions of differently spliced transcripts.

2. **Chrysalis:** groups the overlapping Inchworm contigs by overlaps of $k - 1$ into clusters to construct de Bruijn graph components for each cluster, representing the full transcriptional complexity of a given gene or genes with the common sequence. Next, chrysalis partitions the full read set between clusters.
3. **Butterfly:** resolves spliced and paralogous transcripts independently in parallel, ultimately reporting full-length transcripts.

The transcripts generated by Trinity are applied to gene family clustering with the TGICL (TIGR Gene Indices clustering tools) pipeline [144]. Moreover, to obtain the final unigenes (if there is more than one sample), TGICL will execute again with each sample's unigene to attain the final unigene (for downstream analyses). The unigenes will be divided into (a) clusters containing several clusters with more than 70% similarity and (b) singletons. Figure 2 illustrates the schematic overview of the process.

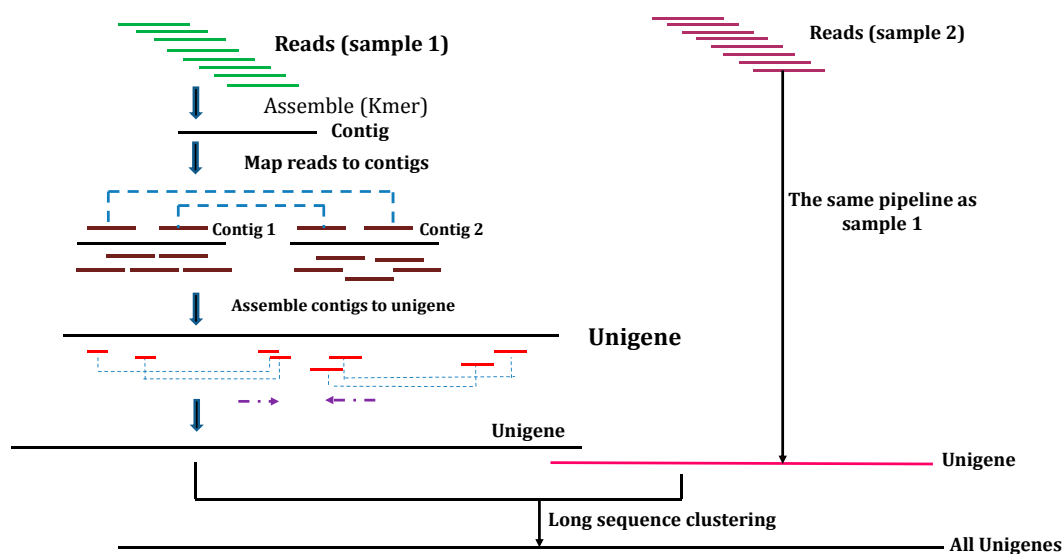


Figure 2. Schematic overview of the de novo transcriptome assembly process.

2.2. Unigene Functional Annotation

The functional databases used include the non-redundant nucleotide sequence database (NT), and the non-redundant protein sequence database (NR) of the National Centre for Biotechnology Information (NCBI), (<http://www.ncbi.nlm.nih.gov>). Additionally, the Swiss-Prot protein, Protein family (Pfam), Eukaryotic Orthologous Groups of proteins (KOG), Gene Ontology (GO), and the Kyoto Encyclopaedia of Genes and Genomes (KEGG). All databases are used to align assembled unigenes using Blast [145–147] (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to obtain the annotated functions of each unigene. With the NR annotation, gene ontology annotations of the unigenes can be acquired using Blast2GO [148] or AmiGO [149]. The Gene Ontology (GO) project is a major bioinformatics collaboration to address the need of knowledge for descriptions of encoding biological functions by genes at the molecular, cellular, and tissue system levels across databases (<http://www.geneontology.org>).

2.3. Microsatellites Mining and Identification Tools

For SSR mining and identification in unigenes, tools such as MISA (MICroSATellite; <http://pgrc.ipk-gatersleben.de/misa>) [45,150] and SSR Locator [151] have been developed. However, these tools are not able to process large genomes efficiently and produce poor statistics. Additionally, as a platform-dependent tool, MISA does not provide a graphical interface or SSR Locator. The development of the Genome-wide Microsatellite Analysing Tool (GMATo) overcomes

the abovementioned weak points, given it is faster and more accurate than MISA and SSR Locator. Furthermore, GMATo is an appropriate, powerful tool for complete SSR characterization in any genome size [152]. Recently, a novel software package, GMATA, was developed that provides new strategies and comprehensive solutions for fast SSR analyses, marker development, and polymorphism screening by mapping and graphically, displaying the results in a genome browser with other genic features. Furthermore, this software also provides high-quality statistical graphics to incorporate in publications [153]. Notably, GMATA is the first tool that generates results that enable viewing SSR loci and SSR marker information along with other genome features in a genome browser. Current software/tools, such as SSR Locator cannot easily design primers that flank each SSR locus in a large genome sequence because the genome sequence at the chromosome level is too large to be directly used as a template for primer design, as for large genomes, primer design can be quite difficult. The GMATA software only uses the flanking sequence as a template for designing PCR primers, thereby reducing computing memory and accelerates the design process for large data sequences. Furthermore, not all primer pairs are unique at the genome scale because duplicated DNA sequences have arisen during evolution. The mining of SSRs from the whole genome provides valuable information on the abundance of SSRs in various genomic regions and will also facilitate the development of markers for genetic analysis and related applications, such as marker-assisted breeding and linkage mapping [154]. Additionally, the Whole Genome Sequencing (WGS)-SSR Annotation Tool (WGSSAT) provides a graphical user interface (GUI) pipeline, mining and characterizing SSR from whole genome data.

The sequences will be searched for perfect mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs. Based on previous studies, dinucleotide and trinucleotide repeat motifs are the most frequent SSR repeats in *Hemarthria* species [89], *Dipteronia Oliver* [108], *Amorphophallus* [31], and pigeon pea [72]. Mono-nucleotide repeats will be excluded since they can result from sequencing errors or mismatches. Furthermore, distinguishing mononucleotides from polyadenylation might be difficult. From the unigenes, primers can then be designed using Primer 3 (<http://bioinfo.ut.ee/primer3>) [155], or Premier 5.0 (PREMIER Biosoft International, Palo Alto, CA, USA), or similar software. Designing primers should meet some criteria, such as the size of the PCR product range between 100 and 280/300 bp; a primer length of 18–21/28 nucleotides; a GC content of 40–70% with 50% as the optimum, and with an annealing temperature between 50 and 70 °C, with 55 °C as the optimum melting temperature [31,108].

2.4. DNA Isolation, PCR Amplification, and SSR Validation

In order to validate the SSRs, the DNA will need to be isolated from plant leaves. DNA integrity will be checked by gel electrophoresis (1% agarose gel). Accordingly, all designed SSR primers should be tested for amplification in different plant varieties or accessions through polymerase chain reaction (PCR). The successful primers will then be selected for genetic diversity studies.

2.5. Genotyping STRs in Next-Generation Data: Challenges and Solutions

Short tandem repeats (STRs) or microsatellites are highly variable elements that play a crucial role in population genetics applications as molecular markers [156]. However, there is a limitation on genotyping STRs from high-throughput sequencing data (for a review, see Treangen and Salzberg, 2012) [157]. From a bioinformatics perspective, if whole reads carrying STRs are mapped due to high mismatch/indel resulting from different STR lengths, some reads will not be mapped with those at the corresponding positions in the reference genome. This leads to a much less accurate estimation of the allele frequency and the real level of STR variation in the genome [158]. More recently, a number of software tools have been developed to profile STRs in NGS data, such as LobSTR [159], RepeatSeq [160], STRViper [161], STR-FM [158], PSR [162], rAmpSeq [163], and STRScan [164]. LobSTR has a fast running time and considers PCR stutter noise during the genotyping stage. However, LobSTR sensitivity is low for mononucleotide STRs and STRs shorter than 25 bp. Additionally, LobSTR uses a mapping algorithm that is fixed in the program [157].

Therefore, an STR-profiling tool was needed to customize a mapping algorithm that can evaluate and correct the STR errors generated by NGS technology [154].

The RepeatSeq tool was released using informed error profiles from inbred *Drosophila* lines [160]. The tool utilizes the reads mapped by other programs, such as Burrows-Wheeler Aligner (BWA) [165] and Bowtie [166], and predicts the most probable genotype at a locus based on the STR motif, length, and base quality. However, RepeatSeq's limitation is in using the whole-read mapping approach, which introduces a bias toward the STR length in the reference genome and thus might obscure the true STR variation spectrum. To profile the full spectrum of STR lengths in human and other genomes, and to correct for NGS-associated STR errors, STR-FM (short tandem repeat profiling using a flank-based mapping approach) was developed as a flexible pipeline for detecting and genotyping STRs from short-read sequencing data. Moreover, this pipeline can detect STRs of any length, including short ones (as short as only two repeats), and includes an error-correcting module, which can combine any NGS mapping algorithm with paired-end mapping capability, thereby making it adaptable to new mapping methods as they become available [158].

Another method that exploits paired-end information for the detection of STR variation from in-depth sequencing data is STRViper [161]. STRViper predicts the polymorphic repeats across a population of genomes and uncovers several polymorphic repeats including the locus of the only known repeat expansion in *A. thaliana*. All tools require prealigned data, except lobSTR, which uses its own aligner. STRViper's performance largely depends on the fragment size variance. Therefore, regarding running time, once reads were aligned, both lobSTR and RepeatSeq performances were poor on moderate variation sizes. Notably, STRViper needed <4 min to process 10-fold coverage reads [161].

All tools mentioned above are used mainly for profiling microsatellites from SAM/BAM data that they identify gSSR alleles at each locus in short reads NGS data. However, they have difficulties in the correct identification of polymorphic SSRs. Unlike the tools above, polymorphic SSR retrieval (PSR) was developed to identify polymorphic SSRs from NGS data where, in the non-model plant species, they use de novo transcriptome assembly as a first sequence resource for SSR mining more effectively [162]. In 2016, Buckler et al. [163] developed the rAmpSeq tool for repeat amplification sequencing that is applicable for genotyping in most species, using low-quality DNA and generating several markers, thereby facilitating whole genome sequencing at less cost per sample. In the last decade, genomics has been used in scientific discovery of thousands of species, but breeding or conservation applications were strongly felt for only a few dozen species. Another software tool, STRScan, was developed for in silico mining STRs from genome sequences with higher sensitivity compared to lobSTR and STR-FM. It uses a specific algorithm for targeted STR profiling in NGS data on the whole genome sequencing (WGS) data from both the Sanger sequencer [167] and the Illumina sequencer (generated by the 1000 Genomes Project [168]). The results showed that STRScan could profile 20% more STRs in the target set, which were missed by lobSTR, in less computation time.

3. Conclusions

Molecular markers are tools used to detect genetic polymorphism at specific loci and an entire genome level in plant species. Among the various molecular markers, SSRs are remarked as being among the most important in genetic and plant breeding programs. However, limited numbers of SSRs are known for some species, thus limiting the capacity of plant breeding approaches. The ability of next-generation sequencing accelerated microsatellite identification and facilitated their variation discovery. Presently, the utilization of RNA-Seq or transcriptome profiling as a reliable and robust tool brings interesting opportunities in the identification and development of a substantial number of SSR markers, being faster, easier, and more cost-effective compared to traditional SSR development processes. The RNA-Seq provides an extensive collection of transcriptomes (expressed sequences), which are believed to be more transmissible among tightly related species as compared to genomic markers because of their presence in more-conserved transcribed regions of the genome. Several studies

on SSR development have demonstrated that Illumina is the most frequently used platform to generate millions of transcriptome sequences, which vary in length. Illumina HiSeq4000 has higher accuracy and is less expensive compared to Illumina HiSeq 2500/3000 sequencing and is the best platform to isolate EST-SSRs markers. Over the years, to support the management of vast amounts of NGS sequence data, and for the profiling and genotyping of short tandem repeats, new specific tools have been developed. Therefore, the utilization of NGS technologies in the development of SSRs is an effective method for the plant community, especially in non-model plants which no genetic information is known.

Acknowledgments: We are grateful to the bioinformatics staff of Malaysia Genome Institute & National Institutes of Biotechnology Malaysia.

Author Contributions: All authors had substantial contributions to the conception, design, and drafting of this work as individual experts in their fields. In particular, S.T. and T.L.A. contributed to writing and organizing the contents of the article and revising it critically. M.R.Y. and M.M.H. revised the manuscript critically. M.S. and P.A. contributed to the writing, figure preparation, and formatting of the article. R.R.S. revised and proofread the article.

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

References

1. Singh, V.K.; Singh, A.K.; Singh, S.; Singh, B.D. Next-Generation Sequencing (NGS) Tools and Impact in Plant Breeding. In *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*; Springer: Cham, Switzerland, 2015; pp. 563–612.
2. Punia, A.; Yadav, R.; Arora, P.; Chaudhury, A. Molecular and morphophysiological characterization of superior cluster bean (*Cymopsis tetragonoloba*) varieties. *J. Crop Sci. Biotechnol.* **2009**, *12*, 143–148. [[CrossRef](#)]
3. Pathak, R.; Singh, S.; Singh, M.; Henry, A. Molecular assessment of genetic diversity in cluster bean (*Cyamopsis tetragonoloba*) genotypes. *J. Genet.* **2010**, *89*, 243–246. [[CrossRef](#)] [[PubMed](#)]
4. Kuravadi, N.A.; Tiwari, P.B.; Tanwar, U.K.; Tripathi, S.K.; Dhugga, K.S.; Gill, K.S.; Randhawa, G.S. Identification and Characterization of EST-SSR Markers in Cluster Bean (spp.). *Crop Sci.* **2014**, *54*, 1097–1102. [[CrossRef](#)]
5. Kuravadi, N.A.; Yenagi, V.; Rangiah, K.; Mahesh, H.; Rajamani, A.; Shirke, M.D.; Russiachand, H.; Loganathan, R.M.; Lingu, C.S.; Siddappa, S. Comprehensive analyses of genomes, transcriptomes and metabolites of neem tree. *PeerJ* **2015**, *3*, e1066. [[CrossRef](#)] [[PubMed](#)]
6. Pathak, R. Genetic Markers and Biotechnology. In *Clusterbean: Physiology, Genetics and Cultivation*; Springer: Singapore, 2015; pp. 125–143.
7. Kumar, S.; Parekh, M.J.; Patel, C.B.; Zala, H.N.; Sharma, R.; Kulkarni, K.S.; Fougat, R.S.; Bhatt, R.K.; Sakure, A.A. Development and validation of EST-derived SSR markers and diversity analysis in cluster bean (*Cyamopsis tetragonoloba*). *J. Plant Biochem. Biotechnol.* **2016**, *25*, 263–269. [[CrossRef](#)]
8. Tanwar, U.K.; Pruthi, V.; Randhawa, G.S. RNA-Seq of Guar (*Cyamopsis tetragonoloba*, L. Taub.) Leaves: De novo Transcriptome Assembly, Functional Annotation and Development of Genomic Resources. *Front. Plant Sci.* **2017**, *8*, 91. [[CrossRef](#)] [[PubMed](#)]
9. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [[CrossRef](#)] [[PubMed](#)]
10. Sakiyama, N.S.; Ramos, H.C.C.; Caixeta, E.T.; Pereira, M.G. Plant breeding with marker-assisted selection in Brazil. *Crop Breed. Appl. Biotechnol.* **2014**, *14*, 54–60. [[CrossRef](#)]
11. Zalapa, J.E.; Cuevas, H.; Zhu, H.; Steffan, S.; Senalik, D.; Zeldin, E.; McCown, B.; Harbut, R.; Simon, P. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* **2012**, *99*, 193–208. [[CrossRef](#)] [[PubMed](#)]
12. Singh, V.; Goel, R.; Pande, V.; Asif, M.H.; Mohanty, C.S. De novo sequencing and comparative analysis of leaf transcriptomes of diverse condensed tannin-containing lines of underutilized *Psophocarpus tetragonolobus* (L.) DC. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]

13. Rosazlina, R.; Jacobsen, N.; Ørgaard, M.; Othman, A.S. Utilizing next generation sequencing to characterize microsatellite loci in a tropical aquatic plant species *Cryptocoryne cordata* var. *cordata* (Araceae). *Biochem. Syst. Ecol.* **2015**, *61*, 385–389. [[CrossRef](#)]
14. Zhao, D.-W.; Yang, J.-B.; Yang, S.-X.; Kato, K.; Luo, J.-P. Genetic diversity and domestication origin of tea plant *Camellia taliensis* (Theaceae) as revealed by microsatellite markers. *BMC Plant Biol.* **2014**, *14*, 1. [[CrossRef](#)] [[PubMed](#)]
15. Taheri, S.; Abdullah, T.L.; Ahmad, Z.; Abdullah, N.A.P. Effect of acute gamma irradiation on *Curcuma alismatifolia* varieties and detection of DNA polymorphism through SSR Marker. *BioMed Res. Int.* **2014**, *2014*. [[CrossRef](#)] [[PubMed](#)]
16. Buschiazzo, E.; Gemmell, N.J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* **2006**, *28*, 1040–1050. [[CrossRef](#)] [[PubMed](#)]
17. Kelkar, Y.D.; Tyekucheva, S.; Chiaromonte, F.; Makova, K.D. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res.* **2008**, *18*, 30–38. [[CrossRef](#)] [[PubMed](#)]
18. Phumichai, C.; Phumichai, T.; Wongkaew, A. Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild *Hevea* rubber. *Plant Mol. Biol. Rep.* **2015**, *33*, 1486–1498. [[CrossRef](#)]
19. Lawson, M.J.; Zhang, L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol.* **2006**, *7*, R14. [[CrossRef](#)] [[PubMed](#)]
20. Oliveira, E.J.; Pádua, J.G.; Zucchi, M.I.; Vencovsky, R.; Vieira, M.L.C. Origin, evolution and genome distribution of microsatellites. *Genet. Mol. Biol.* **2006**, *29*, 294–307. [[CrossRef](#)]
21. Selkoe, K.A.; Toonen, R.J. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* **2006**, *9*, 615–629. [[CrossRef](#)] [[PubMed](#)]
22. Fan, L.; Zhang, M.-Y.; Liu, Q.-Z.; Li, L.-T.; Song, Y.; Wang, L.-F.; Zhang, S.-L.; Wu, J. Transferability of newly developed pear SSR markers to other Rosaceae species. *Plant Mol. Biol. Rep.* **2013**, *31*, 1271–1282. [[CrossRef](#)] [[PubMed](#)]
23. Mason, A.S. SSR genotyping. In *Plant Genotyping. Methods in Molecular Biology (Methods and Protocols)*; Batley, J., Ed.; Humana Press: New York, NY, USA, 2015; pp. 77–89.
24. Kalia, R.K.; Rai, M.K.; Kalia, S.; Singh, R.; Dhawan, A. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* **2011**, *177*, 309–334. [[CrossRef](#)]
25. Zargar, S.M.; Raatz, B.; Sonah, H.; Bhat, J.A.; Dar, Z.A.; Agrawal, G.K.; Rakwal, R. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *J. Crop Sci. Biotechnol.* **2015**, *18*, 293–308. [[CrossRef](#)]
26. Gao, H.; Jiang, K.; Geng, Y.; Chen, X.-Y. Development of microsatellite primers of the largest seagrass, *Enhalus acoroides* (Hydrocharitaceae). *Am. J. Bot.* **2012**, *99*, e99–e101. [[CrossRef](#)] [[PubMed](#)]
27. Jain, S.M.; Brar, D.S.; Ahloowalia, B. *Molecular Techniques in Crop Improvement*; Springer: Dordrecht, The Netherlands, 2010.
28. Antigueira, L.M.O.R. Application of Microsatellite Molecular Markers in Studies of Genetic Diversity and Conservation of Plant Species of Cerrado. *J. Plant Sci.* **2013**, *1*, 1–5.
29. Vieira, M.L.C.; Santini, L.; Diniz, A.L.; Munhoz, C.D.F. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **2016**, *39*, 312–328. [[CrossRef](#)] [[PubMed](#)]
30. Nadeem, M.A.; Nawaz, M.A.; Shahid, M.Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equipment* **2017**, 1–25. [[CrossRef](#)]
31. Zheng, X.; Pan, C.; Diao, Y.; You, Y.; Yang, C.; Hu, Z. Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). *BMC Genom.* **2013**, *14*, 490. [[CrossRef](#)] [[PubMed](#)]
32. Nicot, N.; Chiquet, V.; Gandon, B.; Amilhat, L.; Legeai, F.; Leroy, P.; Bernard, M.; Sourdille, P. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theor. Appl. Genet.* **2004**, *109*, 800–805. [[CrossRef](#)] [[PubMed](#)]
33. Röder, M.S.; Plaschke, J.; König, S.U.; Börner, A.; Sorrells, M.E.; Tanksley, S.D.; Ganal, M.W. Abundance, variability and chromosomal location of microsatellites in wheat. *Mol. Gen. Genet.* **1995**, *246*, 327–333. [[CrossRef](#)] [[PubMed](#)]

34. Ronning, C.M.; Stegalkina, S.S.; Ascenzi, R.A.; Bougri, O.; Hart, A.L.; Utterbach, T.R.; Vanaken, S.E.; Riedmuller, S.B.; White, J.A.; Cho, J. Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.* **2003**, *131*, 419–429. [[CrossRef](#)] [[PubMed](#)]
35. Kurata, N.A.; Nagamura, Y.; Yamamoto, K.; Harushima, Y.; Sue, N.; Wu, J.; Antonio, B.; Shomura, A.; Shimizu, T.; Lin, S.Y. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nat. Genet.* **1994**, *8*, 365–372. [[CrossRef](#)] [[PubMed](#)]
36. Qi, L.; Echaliier, B.; Chao, S.; Lazo, G.; Butler, G.; Anderson, O.; Akhunov, E.; Dvořák, J.; Linkiewicz, A.; Ratnasiri, A. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* **2004**, *168*, 701–712. [[CrossRef](#)] [[PubMed](#)]
37. Ellis, J.; Burke, J. EST-SSRs as a resource for population genetic analyses. *Heredity* **2007**, *99*, 125–132. [[CrossRef](#)] [[PubMed](#)]
38. Varshney, R.K.; Graner, A.; Sorrells, M.E. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol.* **2005**, *23*, 48–55. [[CrossRef](#)] [[PubMed](#)]
39. Jo, K.M.; Jo, Y.; Chu, H.; Lian, S.; Cho, W.K. Development of EST-derived SSR markers using next-generation sequencing to reveal the genetic diversity of 50 chrysanthemum cultivars. *Biochem. Syst. Ecol.* **2015**, *60*, 37–45. [[CrossRef](#)]
40. Rungis, D.; Bérubé, Y.; Zhang, J.; Ralph, S.; Ritland, C.E.; Ellis, B.E.; Douglas, C.; Bohlmann, J.; Ritland, K. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor. Appl. Genet.* **2004**, *109*, 1283–1294. [[CrossRef](#)] [[PubMed](#)]
41. Chen, H.; Liu, L.; Wang, L.; Wang, S.; Somta, P.; Cheng, X. Development and validation of EST-SSR markers from the transcriptome of adzuki bean (*Vigna angularis*). *PLoS ONE* **2015**, *10*, e0131939. [[CrossRef](#)] [[PubMed](#)]
42. Temnykh, S.; DeClerck, G.; Lukashova, A.; Lipovich, L.; Cartinhour, S.; McCouch, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* **2001**, *11*, 1441–1452. [[CrossRef](#)] [[PubMed](#)]
43. Eujayl, I.; Sorrells, M.; Baum, M.; Wolters, P.; Powell, W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica* **2001**, *119*, 39–43. [[CrossRef](#)]
44. Yu, J.-K.; Dake, T.M.; Singh, S.; Bensch, D.; Li, W.; Gill, B.; Sorrells, M.E. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome* **2004**, *47*, 805–818. [[CrossRef](#)] [[PubMed](#)]
45. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)] [[PubMed](#)]
46. Ramu, P.; Kassahun, B.; Senthilvel, S.; Kumar, C.A.; Jayashree, B.; Folkertsma, R.; Reddy, L.A.; Kuruvinashetti, M.; Haussmann, B.; Hash, C. Exploiting rice–sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. *Theor. Appl. Genet.* **2009**, *119*, 1193–1204. [[CrossRef](#)] [[PubMed](#)]
47. Areshchenkova, T.; Ganai, M. Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. *Theor. Appl. Genet.* **2002**, *104*, 229–235. [[CrossRef](#)] [[PubMed](#)]
48. Poncet, V.; Rondeau, M.; Tranchant, C.; Cayrel, A.; Hamon, S.; De Kochko, A.; Hamon, P. SSR mining in coffee tree EST databases: Potential use of EST-SSRs as markers for the *Coffea* genus. *Mol. Genet. Genom.* **2006**, *276*, 436–449. [[CrossRef](#)] [[PubMed](#)]
49. Li, D.; Deng, Z.; Qin, B.; Liu, X.; Men, Z. De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genom.* **2012**, *13*, 192. [[CrossRef](#)] [[PubMed](#)]
50. Qiu, L.; Yang, C.; Tian, B.; Yang, J.-B.; Liu, A. Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol.* **2010**, *10*, 278. [[CrossRef](#)] [[PubMed](#)]
51. Wei, W.; Qi, X.; Wang, L.; Zhang, Y.; Hua, W.; Li, D.; Lv, H.; Zhang, X. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genom.* **2011**, *12*, 451. [[CrossRef](#)] [[PubMed](#)]

52. Taheri, S.; Abdullah, T.L.; Jain, S.M.; Sahebi, M.; Azizi, P. TILLING, high-resolution melting (HRM), and next-generation sequencing (NGS) techniques in plant mutation breeding. *Mol. Breed.* **2017**, *37*, 40. [[CrossRef](#)]
53. Squirrell, J.; Hollingsworth, P.; Woodhead, M.; Russell, J.; Lowe, A.; Gibby, M.; Powell, W. How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* **2003**, *12*, 1339–1348. [[CrossRef](#)] [[PubMed](#)]
54. Zane, L.; Bargelloni, L.; Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **2002**, *11*, 1–16. [[CrossRef](#)] [[PubMed](#)]
55. Zhu, H.; Senalik, D.; McCown, B.; Zeldin, E.; Speers, J.; Hyman, J.; Bassil, N.; Hummer, K.; Simon, P.; Zalapa, J. Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theor. Appl. Genet.* **2012**, *124*, 87–96. [[CrossRef](#)] [[PubMed](#)]
56. 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. [[CrossRef](#)] [[PubMed](#)]
57. Csencsics, D.; Brodbeck, S.; Holderegger, R. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *J. Hered.* **2010**, *101*, 789–793. [[CrossRef](#)] [[PubMed](#)]
58. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145. [[CrossRef](#)] [[PubMed](#)]
59. Ekblom, R.; Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **2011**, *107*, 1–15. [[CrossRef](#)] [[PubMed](#)]
60. Stapley, J.; Reger, J.; Feulner, P.G.; Smadja, C.; Galindo, J.; Ekblom, R.; Bennison, C.; Ball, A.D.; Beckerman, A.P.; Slate, J. Adaptation genomics: The next generation. *Trends Ecol. Evol.* **2010**, *25*, 705–712. [[CrossRef](#)] [[PubMed](#)]
61. Duan, X.; Wang, K.; Su, S.; Tian, R.; Li, Y.; Chen, M. De novo transcriptome analysis and microsatellite marker development for population genetic study of a serious insect pest, *Rhopalosiphum padi* (L.) (Hemiptera: Aphididae). *PLoS ONE* **2017**, *12*, e0172513. [[CrossRef](#)] [[PubMed](#)]
62. Egan, A.N.; Schlueter, J.; Spooner, D.M. Applications of next-generation sequencing in plant biology. *Am. J. Bot.* **2012**, *99*, 175–185. [[CrossRef](#)] [[PubMed](#)]
63. Mardis, E.R. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* **2017**, *12*, 213–218. [[CrossRef](#)] [[PubMed](#)]
64. Lee, C.-Y.; Chiu, Y.-C.; Wang, L.-B.; Kuo, Y.-L.; Chuang, E.Y.; Lai, L.-C.; Tsai, M.-H. Common applications of next-generation sequencing technologies in genomic research. *Transl. Cancer Res.* **2013**, *2*, 33–45.
65. Grohme, M.A.; Soler, R.F.; Wink, M.; Frohme, M. Microsatellite marker discovery using single molecule real-time circular consensus sequencing on the Pacific Biosciences RS. *BioTechniques* **2013**, *55*, 253–256. [[CrossRef](#)] [[PubMed](#)]
66. Ambardar, S.; Gupta, R.; Trakroo, D.; Lal, R.; Vakhlu, J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J. Microbiol.* **2016**, *56*, 394–404. [[CrossRef](#)] [[PubMed](#)]
67. Ray, S.; Satya, P. Next generation sequencing technologies for next generation plant breeding. *Front. Plant Sci.* **2014**, *5*, 367. [[CrossRef](#)] [[PubMed](#)]
68. Addisalem, A.; Esselink, G.D.; Bongers, F.; Smulders, M. Genomic sequencing and microsatellite marker development for *Boswellia papyrifera*, an economically important but threatened tree native to dry tropical forests. *AoB Plants* **2015**, *7*. [[CrossRef](#)] [[PubMed](#)]
69. Parchman, T.L.; Geist, K.S.; Grahnen, J.A.; Benkman, C.W.; Buerkle, C.A. Transcriptome sequencing in an ecologically important tree species: Assembly, annotation, and marker discovery. *BMC Genom.* **2010**, *11*, 180. [[CrossRef](#)] [[PubMed](#)]
70. Blanca, J.; Cañizares, J.; Roig, C.; Ziarsolo, P.; Nuez, F.; Picó, B. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genom.* **2011**, *12*, 104. [[CrossRef](#)] [[PubMed](#)]
71. Hiremath, P.J.; Farmer, A.; Cannon, S.B.; Woodward, J.; Kudapa, H.; Tuteja, R.; Kumar, A.; BhanuPrakash, A.; Mulaosmanovic, B.; Gujaria, N. Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol. J.* **2011**, *9*, 922–931. [[CrossRef](#)] [[PubMed](#)]

72. Dutta, S.; Kumawat, G.; Singh, B.P.; Gupta, D.K.; Singh, S.; Dogra, V.; Gaikwad, K.; Sharma, T.R.; Raje, R.S.; Bandhopadhyay, T.K. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* **2011**, *11*, 17. [[CrossRef](#)] [[PubMed](#)]
73. Lu, F.H.; Yoon, M.Y.; Cho, Y.I.; Chung, J.W.; Kim, K.T.; Cho, M.C.; Cheong, S.R.; Park, Y.J. Transcriptome analysis and SNP/SSR marker information of red pepper variety YCM334 and Taeon. *Scientia Horticulturae* **2011**, *129*, 38–45. [[CrossRef](#)]
74. Severin, A.J.; Woody, J.L.; Bolon, Y.-T.; Joseph, B.; Diers, B.W.; Farmer, A.D.; Muehlbauer, G.J.; Nelson, R.T.; Grant, D.; Specht, J.E. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biol.* **2010**, *10*, 160. [[CrossRef](#)] [[PubMed](#)]
75. Zenoni, S.; Ferrarini, A.; Giacomelli, E.; Xumerle, L.; Fasoli, M.; Malerba, G.; Bellin, D.; Pezzotti, M.; Delledonne, M. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol.* **2010**, *152*, 1787–1795. [[CrossRef](#)] [[PubMed](#)]
76. Yates, S.A.; Swain, M.T.; Hegarty, M.J.; Chernukin, I.; Lowe, M.; Allison, G.G.; Ruttink, T.; Abberton, M.T.; Jenkins, G.; Skøt, L. De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genom.* **2014**, *15*, 453. [[CrossRef](#)] [[PubMed](#)]
77. Garg, R.; Patel, R.K.; Tyagi, A.K.; Jain, M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res.* **2011**, *18*, 53–63. [[CrossRef](#)] [[PubMed](#)]
78. Garg, R.; Patel, R.K.; Jhanwar, S.; Priya, P.; Bhattacharjee, A.; Yadav, G.; Bhatia, S.; Chattopadhyay, D.; Tyagi, A.K.; Jain, M. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* **2011**, *156*, 1661–1678. [[CrossRef](#)] [[PubMed](#)]
79. Zhang, J.; Liang, S.; Duan, J.; Wang, J.; Chen, S.; Cheng, Z.; Zhang, Q.; Liang, X.; Li, Y. De novo assembly and Characterization of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genom.* **2012**, *13*, 90. [[CrossRef](#)] [[PubMed](#)]
80. Wei, Z.; Sun, Z.; Cui, B.; Zhang, Q.; Xiong, M.; Wang, X.; Zhou, D. Transcriptome analysis of colored calla lily (*Zantedeschia rehmannii* Engl.) by Illumina sequencing: De novo assembly, annotation and EST-SSR marker development. *PeerJ* **2016**, *4*, e2378. [[CrossRef](#)] [[PubMed](#)]
81. Simsek, O.; Donmez, D.; Kacar, Y.A. RNA-Seq Analysis in Fruit Science: A Review. *Am. J. Plant Biol.* **2017**, *2*, 1–7.
82. Li, S.; Tighe, S.W.; Nicolet, C.M.; Grove, D.; Levy, S.; Farmerie, W.; Viale, A.; Wright, C.; Schweitzer, P.A.; Gao, Y. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat. Biotechnol.* **2014**, *32*, 915–925. [[CrossRef](#)] [[PubMed](#)]
83. Cloonan, N.; Forrest, A.R.; Kolle, G.; Gardiner, B.B.; Faulkner, G.J.; Brown, M.K.; Taylor, D.F.; Steptoe, A.L.; Wani, S.; Bethel, G. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **2008**, *5*, 613–619. [[CrossRef](#)] [[PubMed](#)]
84. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **2008**, *320*, 1344–1349. [[CrossRef](#)] [[PubMed](#)]
85. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
86. Wang, Z.; Fang, B.; Chen, J.; Zhang, X.; Luo, Z.; Huang, L.; Chen, X.; Li, Y. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genom.* **2010**, *11*, 726. [[CrossRef](#)] [[PubMed](#)]
87. Zhang, G.; Guo, G.; Hu, X.; Zhang, Y.; Li, Q.; Li, R.; Zhuang, R.; Lu, Z.; He, Z.; Fang, X. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* **2010**, *20*, 646–654. [[CrossRef](#)] [[PubMed](#)]
88. Vijay, N.; Poelstra, J.W.; Kunstner, A.; Wolf, J.B. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Mol. Ecol.* **2013**, *22*, 620–634. [[CrossRef](#)] [[PubMed](#)]
89. Huang, X.; Yan, H.-D.; Zhang, X.-Q.; Zhang, J.; Frazier, T.P.; Huang, D.-J.; Lu, L.; Huang, L.-K.; Liu, W.; Peng, Y. De novo Transcriptome Analysis and Molecular Marker Development of Two Hemarthria Species. *Front. Plant Sci.* **2016**, *7*, 496. [[CrossRef](#)] [[PubMed](#)]

90. Garcia-Seco, D.; Zhang, Y.; Gutierrez-Mañero, F.J.; Martin, C.; Ramos-Solano, B. RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus* sp. Var. Lochness) fruit. *BMC Genom.* **2015**, *16*, 5. [[CrossRef](#)] [[PubMed](#)]
91. Simon, S.A.; Zhai, J.; Nandety, R.S.; McCormick, K.P.; Zeng, J.; Mejia, D.; Meyers, B.C. Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.* **2009**, *60*, 305–333. [[CrossRef](#)] [[PubMed](#)]
92. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
93. Wolf, J.B. Principles of transcriptome analysis and gene expression quantification: An RNA-seq tutorial. *Mol. Ecol. Resour.* **2013**, *13*, 559–572. [[CrossRef](#)] [[PubMed](#)]
94. Varshney, R.; Grosse, I.; Hähnel, U.; Siefken, R.; Prasad, M.; Stein, N.; Langridge, P.; Altschmied, L.; Graner, A. Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *Theor. Appl. Genet.* **2006**, *113*, 239. [[CrossRef](#)] [[PubMed](#)]
95. Wang, Z.; Li, J.; Luo, Z.; Huang, L.; Chen, X.; Fang, B.; Li, Y.; Chen, J.; Zhang, X. Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). *BMC Plant Biol.* **2011**, *11*, 139. [[CrossRef](#)] [[PubMed](#)]
96. Iorizzo, M.; Senalik, D.A.; Grzebelus, D.; Bowman, M.; Cavagnaro, P.F.; Matvienko, M.; Ashrafi, H.; Van Deynze, A.; Simon, P.W. De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genom.* **2011**, *12*, 389. [[CrossRef](#)] [[PubMed](#)]
97. Gao, J.; Zhang, Y.; Zhang, C.; Qi, F.; Li, X.; Mu, S.; Peng, Z. Characterization of the floral transcriptome of Moso bamboo (*Phyllostachys edulis*) at different flowering developmental stages by transcriptome sequencing and RNA-seq analysis. *PLoS ONE* **2014**, *9*, e98910. [[CrossRef](#)] [[PubMed](#)]
98. Yin, D.; Wang, Y.; Zhang, X.; Li, H.; Lu, X.; Zhang, J.; Zhang, W.; Chen, S. De novo assembly of the peanut (*Arachis hypogaea* L.) seed transcriptome revealed candidate unigenes for oil accumulation pathways. *PLoS ONE* **2013**, *8*, e73767. [[CrossRef](#)] [[PubMed](#)]
99. Kaur, S.; Pembleton, L.W.; Cogan, N.O.; Savin, K.W.; Leonforte, T.; Paull, J.; Materne, M.; Forster, J.W. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genom.* **2012**, *13*, 104. [[CrossRef](#)] [[PubMed](#)]
100. Wu, J.; Wang, L.; Li, L.; Wang, S. De novo assembly of the common bean transcriptome using short reads for the discovery of drought-responsive genes. *PLoS ONE* **2014**, *9*, e109262. [[CrossRef](#)] [[PubMed](#)]
101. Liu, C.; Fan, B.; Cao, Z.; Su, Q.; Wang, Y.; Zhang, Z.; Wu, J.; Tian, J. A deep sequencing analysis of transcriptomes and the development of EST-SSR markers in mungbean (*Vigna radiata*). *J. Genet.* **2016**, *95*, 527–535. [[CrossRef](#)] [[PubMed](#)]
102. Tian, W.; Paudel, D.; Vendrame, W.; Wang, J. Enriching Genomic Resources and Marker Development from Transcript Sequences of *Jatropha curcas* for Microgravity Studies. *Int. J. Genom.* **2017**, *2017*. [[CrossRef](#)]
103. Kovi, M.R.; Amdahl, H.; Alsheikh, M.; Rognli, O.A. De novo and reference transcriptome assembly of transcripts expressed during flowering provide insight into seed setting in tetraploid red clover. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
104. Vatanparast, M.; Shetty, P.; Chopra, R.; Doyle, J.J.; Sathyanarayana, N.; Egan, A.N. Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). *Sci. Rep.* **2016**, *6*. [[CrossRef](#)] [[PubMed](#)]
105. Jia, H.; Yang, H.; Sun, P.; Li, J.; Zhang, J.; Guo, Y.; Han, X.; Zhang, G.; Lu, M.; Hu, J. De novo transcriptome assembly, development of EST-SSR markers and population genetic analyses for the desert biomass willow, *Salix psammophila*. *Sci. Rep.* **2016**, *6*. [[CrossRef](#)] [[PubMed](#)]
106. Mora-Ortiz, M.; Swain, M.T.; Vickers, M.J.; Hegarty, M.J.; Kelly, R.; Smith, L.M.; Sköt, L. De novo transcriptome assembly for gene identification, analysis, annotation, and molecular marker discovery in *Onobrychis vicifolia*. *BMC Genom.* **2016**, *17*, 756. [[CrossRef](#)] [[PubMed](#)]
107. An, M.; Deng, M.; Zheng, S.-S.; Song, Y.-G. De novo transcriptome assembly and development of SSR markers of oaks *Quercus austrocochinchinensis* and *Q. kerrii* (Fagaceae). *Tree Genet. Genom.* **2016**, *12*, 103. [[CrossRef](#)]
108. Zhou, T.; Li, Z.-H.; Bai, G.-Q.; Feng, L.; Chen, C.; Wei, Y.; Chang, Y.-X.; Zhao, G.-F. Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia Oliver* (Aceraceae). *Molecules* **2016**, *21*, 166. [[CrossRef](#)] [[PubMed](#)]

109. Zhou, Q.; Luo, D.; Ma, L.; Xie, W.; Wang, Y.; Wang, Y.; Liu, Z. Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. *Sci. Rep.* **2016**, *6*. [[CrossRef](#)] [[PubMed](#)]
110. White, O.W.; Doo, B.; Carine, M.A.; Chapman, M.A. Transcriptome sequencing and simple sequence repeat marker development for three Macaronesian endemic plant species. *Appl. Plant Sci.* **2016**, *4*. [[CrossRef](#)] [[PubMed](#)]
111. Wang, Y.; Liu, K.; Bi, D.; Zhou, B.S.; Shao, W.J. Characterization of the transcriptome and EST-SSR development in *Boea clarkeana*, a desiccation-tolerant plant endemic to China. *PeerJ* **2017**, *5*, e3422. [[CrossRef](#)] [[PubMed](#)]
112. Zhao, K.K.; Wang, H.F.; Sakaguchi, S.; Landrein, S.; Isagi, Y.; Maki, M.; Zhu, Z.X. Development and characterization of EST-SSR markers in an East Asian temperate plant genus *Diabelia* (Caprifoliaceae). *Plant Species Biol.* **2017**, *32*, 247–251. [[CrossRef](#)]
113. Wang, L.; Yang, Y.; Zhao, Y.; Yang, S.; Udikeri, S.; Liu, T. De Novo Characterization of the Root Transcriptome and Development of EST-SSR Markers in *Paris polyphylla* Smith var. *yunnanensis*, an Endangered Medical Plant. *J. Agric. Sci. Technol.* **2016**, *18*, 437–452.
114. Liang, M.; Yang, X.; Li, H.; Su, S.; Yi, H.; Chai, L.; Deng, X. De novo transcriptome assembly of pummelo and molecular marker development. *PLoS ONE* **2015**, *10*, e0120615. [[CrossRef](#)] [[PubMed](#)]
115. Dang, M.; Liu, Z.X.; Chen, X.; Zhang, T.; Zhou, H.J.; Hu, Y.H.; Zhao, P. Identification, development, and application of 12 polymorphic EST-SSR markers for an endemic Chinese walnut (*Juglans cathayensis* L.) using next-generation sequencing technology. *Biochem. Syst. Ecol.* **2015**, *60*, 74–80. [[CrossRef](#)]
116. Ding, Q.; Li, J.; Wang, F.; Zhang, Y.; Li, H.; Zhang, J.; Gao, J. Characterization and development of EST-SSRs by deep transcriptome sequencing in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Int. J. Genom.* **2015**, *2015*. [[CrossRef](#)]
117. Zheng, X.; You, Y.; Diao, Y.; Zheng, X.; Xie, K.; Zhou, M.; Hu, Z.; Wang, Y. Development and characterization of genic-SSR markers from different *Nelumbo nucifera* types by RNA-seq. *Gen. Mol. Res.* **2015**, *14*, 11171–11184. [[CrossRef](#)] [[PubMed](#)]
118. Ambreen, H.; Kumar, S.; Variath, M.T.; Joshi, G.; Bali, S.; Agarwal, M.; Kumar, A.; Jagannath, A.; Goel, S. Development of genomic microsatellite markers in *Carthamus tinctorius* L. (safflower) using next generation sequencing and assessment of their cross-species transferability and utility for diversity analysis. *PLoS ONE* **2015**, *10*, e0135443. [[CrossRef](#)] [[PubMed](#)]
119. Tsai, C.C.; Shih, H.C.; Wang, H.V.; Lin, Y.S.; Chang, C.H.; Chiang, Y.C.; Chou, C.H. RNA-seq SSRs of moth orchid and screening for molecular markers across genus *Phalaenopsis* (Orchidaceae). *PLoS ONE* **2015**, *10*, e0141761. [[CrossRef](#)] [[PubMed](#)]
120. Chen, L.Y.; Cao, Y.N.; Yuan, N.; Nakamura, K.; Wang, G.M.; Qiu, Y.X. Characterization of transcriptome and development of novel EST-SSR makers based on next-generation sequencing technology in *Neolitsea sericea* (Lauraceae) endemic to East Asian land-bridge islands. *Mol. Breed.* **2015**, *35*, 1–15. [[CrossRef](#)]
121. Ravishankar, K.; Dinesh, M.; Nischita, P.; Sandya, B. Development and characterization of microsatellite markers in mango (*Mangifera indica*) using next-generation sequencing technology and their transferability across species. *Mol. Breed.* **2015**, *35*, 1–13. [[CrossRef](#)]
122. Torre, S.; Tattini, M.; Brunetti, C.; Fineschi, S.; Fini, A.; Ferrini, F.; Sebastiani, F. RNA-seq analysis of *Quercus pubescens* Leaves: De novo transcriptome assembly, annotation and functional markers development. *PLoS ONE* **2014**, *9*, e112487. [[CrossRef](#)] [[PubMed](#)]
123. Izzah, N.K.; Lee, J.; Jayakodi, M.; Perumal, S.; Jin, M.; Park, B.-S.; Ahn, K.; Yang, T.-J. Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. *capitata* L.) and construction of an EST-based genetic map. *BMC Genom.* **2014**, *15*, 149. [[CrossRef](#)] [[PubMed](#)]
124. Salgado, L.R.; Koop, D.M.; Pinheiro, D.G.; Rivallan, R.; Le Guen, V.; Nicolás, M.F.; De Almeida, L.G.P.; Rocha, V.R.; Magalhães, M.; Gerber, A.L. De novo transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genom.* **2014**, *15*, 236. [[CrossRef](#)] [[PubMed](#)]
125. Wang, Z.; Yu, G.; Shi, B.; Wang, X.; Qiang, H.; Gao, H. Development and characterization of simple sequence repeat (SSR) markers based on RNA-sequencing of *Medicago sativa* and in silico mapping onto the *M. truncatula* genome. *PLoS ONE* **2014**, *9*, e92029. [[CrossRef](#)] [[PubMed](#)]

126. Giordano, A.; Cogan, N.O.; Kaur, S.; Drayton, M.; Mouradov, A.; Panter, S.; Schrauf, G.E.; Mason, J.G.; Spangenberg, G.C. Gene discovery and molecular marker development, based on high-throughput transcript sequencing of *Paspalum dilatatum* Poir. *PLoS ONE* **2014**, *9*, e85050. [[CrossRef](#)] [[PubMed](#)]
127. Zou, D.; Chen, X.; Zou, D. Sequencing, de novo assembly, annotation and SSR and SNP detection of sabaigrass (*Eulaliopsis binata*) transcriptome. *Genomics* **2013**, *102*, 57–62. [[CrossRef](#)] [[PubMed](#)]
128. Chung, J.W.; Kim, T.S.; Suresh, S.; Lee, S.Y.; Cho, G.T. Development of 65 novel polymorphic cDNA-SSR markers in common vetch (*Vicia sativa* subsp. *sativa*) using next generation sequencing. *Molecules* **2013**, *18*, 8376–8392. [[CrossRef](#)] [[PubMed](#)]
129. Suresh, S.; Park, J.H.; Cho, G.T.; Lee, H.S.; Baek, H.J.; Lee, S.Y.; Chung, J.W. Development and molecular characterization of 55 novel polymorphic cDNA-SSR markers in faba bean (*Vicia faba* L.) using 454 pyrosequencing. *Molecules* **2013**, *18*, 1844–1856. [[CrossRef](#)] [[PubMed](#)]
130. Verma, P.; Shah, N.; Bhatia, S. Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) transcriptome. *Plant Biotechnol. J.* **2013**, *11*, 894–905. [[CrossRef](#)] [[PubMed](#)]
131. Tan, L.-Q.; Wang, L.-Y.; Wei, K.; Zhang, C.-C.; Wu, L.-Y.; Qi, G.-N.; Cheng, H.; Zhang, Q.; Cui, Q.-M.; Liang, J.-B. Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). *PLoS ONE* **2013**, *8*, e81611. [[CrossRef](#)] [[PubMed](#)]
132. Wu, H.; Chen, D.; Li, J.; Yu, B.; Qiao, X.; Huang, H.; He, Y. De novo characterization of leaf transcriptome using 454 sequencing and development of EST-SSR markers in tea (*Camellia sinensis*). *Plant Mol. Biol. Rep.* **2013**, *31*, 524–538. [[CrossRef](#)]
133. Pazos-Navarro, M.; Dabauza, M.; Correal, E.; Hanson, K.; Teakle, N.; Real, D.; Nelson, M.N. Next generation DNA sequencing technology delivers valuable genetic markers for the genomic orphan legume species, *Bituminaria bituminosa*. *BMC Genet.* **2011**, *12*, 104. [[CrossRef](#)] [[PubMed](#)]
134. Kaur, S.; Cogan, N.O.; Pembleton, L.W.; Shinozuka, M.; Savin, K.W.; Materne, M.; Forster, J.W. Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genom.* **2011**, *12*, 265. [[CrossRef](#)] [[PubMed](#)]
135. Triwitayakorn, K.; Chatkulkawin, P.; Kanjanawattanawong, S.; Sraphet, S.; Yoocha, T.; Sangsrakru, D.; Chanprasert, J.; Ngamphiw, C.; Jomchai, N.; Therawattanasuk, K. Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. *DNA Res.* **2011**, *18*, 471–482. [[CrossRef](#)] [[PubMed](#)]
136. Cock, P.J.; Fields, C.J.; Goto, N.; Heuer, M.L.; Rice, P.M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* **2010**, *38*, 1767–1771. [[CrossRef](#)] [[PubMed](#)]
137. Surget-Groba, Y.; Montoya-Burgos, J.I. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res.* **2010**, *20*, 1432–1440. [[CrossRef](#)] [[PubMed](#)]
138. Martin, J.; Bruno, V.M.; Fang, Z.; Meng, X.; Blow, M.; Zhang, T.; Sherlock, G.; Snyder, M.; Wang, Z. Rnnotator: An automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genom.* **2010**, *11*, 663. [[CrossRef](#)] [[PubMed](#)]
139. Robertson, G.; Schein, J.; Chiu, R.; Corbett, R.; Field, M.; Jackman, S.D.; Mungall, K.; Lee, S.; Okada, H.M.; Qian, J.Q. De novo assembly and analysis of RNA-seq data. *Nat. Methods* **2010**, *7*, 909–912. [[CrossRef](#)] [[PubMed](#)]
140. Schulz, M.H.; Zerbino, D.R.; Vingron, M.; Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **2012**, *28*, 1086–1092. [[CrossRef](#)] [[PubMed](#)]
141. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2011**, *29*, 644. [[CrossRef](#)] [[PubMed](#)]
142. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
143. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]

144. Perteua, G.; Huang, X.; Liang, F.; Antonescu, V.; Sultana, R.; Karamycheva, S.; Lee, Y.; White, J.; Cheung, F.; Parvizi, B. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* **2003**, *19*, 651–652. [[CrossRef](#)] [[PubMed](#)]
145. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
146. Cameron, M.; Williams, H.E.; Cannane, A. Improved gapped alignment in BLAST. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2004**, *1*, 116–129. [[CrossRef](#)] [[PubMed](#)]
147. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
148. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)] [[PubMed](#)]
149. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S.; Group, W.P.W. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2009**, *25*, 288–289. [[CrossRef](#)] [[PubMed](#)]
150. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
151. Da Maia, L.C.; Palmieri, D.A.; De Souza, V.Q.; Kopp, M.M.; de Carvalho, F.I.F.; Costa de Oliveira, A. SSR locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int. J. Plant Genom.* **2008**, *2008*. [[CrossRef](#)] [[PubMed](#)]
152. Wang, X.; Lu, P.; Luo, Z. GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformatics* **2013**, *9*, 541. [[CrossRef](#)] [[PubMed](#)]
153. Wang, X.; Wang, L. GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Front. Plant Sci.* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
154. Pandey, M.; Kumar, R.; Srivastava, P.; Agarwal, S.; Srivastava, S.; Nagpure, N.S.; Jena, J.K.; Kushwaha, B. WGSSAT: A High-Throughput Computational Pipeline for Mining and Annotation of SSR Markers From Whole Genomes. *J. Hered.* **2017**. [[CrossRef](#)] [[PubMed](#)]
155. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)] [[PubMed](#)]
156. Verstrepen, K.J.; Jansen, A.; Lewitter, F.; Fink, G.R. Intragenic tandem repeats generate functional variability. *Nat. Genet.* **2005**, *37*, 986. [[CrossRef](#)] [[PubMed](#)]
157. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat. Rev. Genet.* **2012**, *13*, 36–46. [[CrossRef](#)] [[PubMed](#)]
158. Functamman, A.; Ananda, G.; Hile, S.E.; Su, M.S.-W.; Sun, C.; Harris, R.; Medvedev, P.; Eckert, K.; Makova, K.D. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* **2015**, *25*, 736–749. [[CrossRef](#)] [[PubMed](#)]
159. Gymrek, M.; Golan, D.; Rosset, S.; Erlich, Y. lobSTR: A short tandem repeat profiler for personal genomes. *Genome Res.* **2012**, *22*, 1154–1162. [[CrossRef](#)] [[PubMed](#)]
160. Highnam, G.; Franck, C.; Martin, A.; Stephens, C.; Puthige, A.; Mittelman, D. Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles. *Nucleic Acids Res.* **2012**, *41*, e32. [[CrossRef](#)] [[PubMed](#)]
161. Cao, M.D.; Tasker, E.; Willadsen, K.; Imelfort, M.; Vishwanathan, S.; Sureshkumar, S.; Balasubramanian, S.; Bodén, M. Inferring short tandem repeat variation from paired-end short reads. *Nucleic Acids Res.* **2013**, *42*, e16. [[CrossRef](#)] [[PubMed](#)]
162. Cantarella, C.; D’Agostino, N. PSR: Polymorphic SSR retrieval. *BMC Res. Notes* **2015**, *8*, 525. [[CrossRef](#)] [[PubMed](#)]
163. Buckler, E.S.; Ilut, D.C.; Wang, X.; Kretzschmar, T.; Gore, M.A.; Mitchell, S.E. rAmpSeq: Using repetitive sequences for robust genotyping. *BioRxiv* **2016**. [[CrossRef](#)]
164. Tang, H.; Nzabarushimana, E. STRScan: Targeted profiling of short tandem repeats in whole-genome sequencing data. *BMC Bioinform.* **2017**, *18*, 398. [[CrossRef](#)] [[PubMed](#)]
165. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]

166. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
167. Levy, S.; Sutton, G.; Ng, P.C.; Feuk, L.; Halpern, A.L.; Walenz, B.P.; Axelrod, N.; Huang, J.; Kirkness, E.F.; Denisov, G. The diploid genome sequence of an individual human. *PLoS Biol.* **2007**, *5*, e254. [[CrossRef](#)] [[PubMed](#)]
168. Consortium, G.P. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467*, 1061–1073.



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).