



Review

# Progress in Plant Genome Sequencing

Robert J. Henry <sup>1,2</sup>

<sup>1</sup> Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, QLD 4072, Australia; robert.henry@uq.edu.au

<sup>2</sup> ARC Centre of Excellence for Plant Success in Nature and Agriculture, University of Queensland, Brisbane, QLD 4072, Australia

**Abstract:** The genome sequence of any organism is key to understanding the biology and utility of that organism. Plants have diverse, complex and sometimes very large nuclear genomes, mitochondrial genomes and much smaller and more highly conserved chloroplast genomes. Plant genome sequences underpin our understanding of plant biology and serve as a key platform for the genetic selection and improvement of crop plants to achieve food security. The development of technology that can capture large volumes of sequence data at low costs and with high accuracy has driven the acceleration of plant genome sequencing advancements. More recently, the development of long read sequencing technology has been a key advance for supporting the accurate sequencing and assembly of chromosome-level plant genomes. This review explored the progress in the sequencing and assembly of plant genomes and the outcomes of plant genome sequencing to date. The outcomes support the conservation of biodiversity, adaptations to climate change and improvements in the sustainability of agriculture, which support food and nutritional security.

**Keywords:** DNA sequencing; plant genome; long read sequencing; chromosome assembly



**Citation:** Henry, R.J. Progress in Plant Genome Sequencing. *Appl. Biosci.* **2022**, *1*, 113–128. <https://doi.org/10.3390/applbiosci1020008>

Academic Editor: Adriana Basile

Received: 15 May 2022

Accepted: 1 July 2022

Published: 4 July 2022

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



**Copyright:** © 2022 by the author. 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

Advances in the analysis of DNA sequences have been a key driver of enhanced biological understanding and the application of biological knowledge [1]. DNA sequencing in the 20th century was largely based on Sanger sequencing, which limited both the quality (accuracy) and volume of data that could be generated relative to the next generation sequencing that we have today [2]. The introduction and rapid development of next generation sequencing has resulted in an acceleration in the development of plant genome sequencing, especially over the last decade [3]. This technology has evolved rapidly, resulting in continuous major changes to the strategies that are used to sequence and assemble genomes. For example, when only short-read sequences were available, physical mapping was a key strategy. Large fragments of the genomes were cloned in bacterial artificial chromosomes (BACs) [4]. The BACs were then sequenced and the genomes were assembled by covering the genetic maps with BAC tiles [5]. The availability of accurate long read sequencing has made these approaches largely redundant [6]. A review in 2018 [7] reported that 236 angiosperm genome sequences had been reported. Since then, many more genomes have been sequenced and the quality of the genome sequences has increased significantly. The NCBI database (<https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/flowering%20plants>; accessed 6 June 2022) includes 831 flowering plant genomes, with 373 at the chromosome level. The de novo assembly of long read sequences allows very large contigs to be assembled, sometimes representing a complete plant chromosome [8].

## 2. Diversity of Plant Genomes

Plant genomes vary enormously in size, even within closely related groups of plants [9]. The nuclear genomes of flowering plants (angiosperms) vary more than 1000-fold, from

less than 100 kb to more than 100 Gb [10]. The genomes of gymnosperms are generally large and complex and represent an even greater challenge for genome sequencing [11]. The large (10 Gb) genome of *Ginkgo biloba* has recently been reported [12], which provides the first reference genome for gymnosperms. Genomes also vary greatly in terms of their content of repetitive sequences, the level of gene duplication, their ploidy and their heterozygosity, providing a range of challenges and degrees of difficulty within genome sequencing and assembly.

### 3. Applications of Plant Genome Sequencing

#### 3.1. Model Genomes

The challenge of sequencing plant genomes using early technologies made it necessary to focus on sequencing model genomes that could be used to study related, but more complex, species. The first plant to have a sequenced genome was *Arabidopsis thaliana* [13], which was chosen because it is a small plant with a rapid generation time and a very small genome, thereby making it an ideal model plant for research use. The first crop plant with a sequenced genome [14] was rice (*Oryza sativa*), which was chosen because it is a major food crop plant with a relatively small genome. This became a model for cereal and grass genomes. Similarly, *Brachypodium distachyon* was sequenced [15] as a model grass genome, which is especially relevant for the wheat genome. Recent advances in genome sequencing technology have greatly reduced the need for models as it is now possible to sequence most species easily.

#### 3.2. Crop Plant Genomes

The sequencing of the genomes of crop species has become a key enabling tool for plant improvement. Most major crops now have reference genome sequences [16] and as the technology becomes more powerful and the costs reduce, genomes are also being generated for many other minor crops. This usually involves the production of a reference genome sequence for a species and the re-sequencing of many individuals to define allelic variations within that species. Current efforts recognize that a single reference genome cannot always serve the needs of plant breeders, so pan-genomes that capture the variations in many diverse genomes within the gene pool are being produced as breeding platforms.

#### 3.3. Sequencing Plant Biodiversity

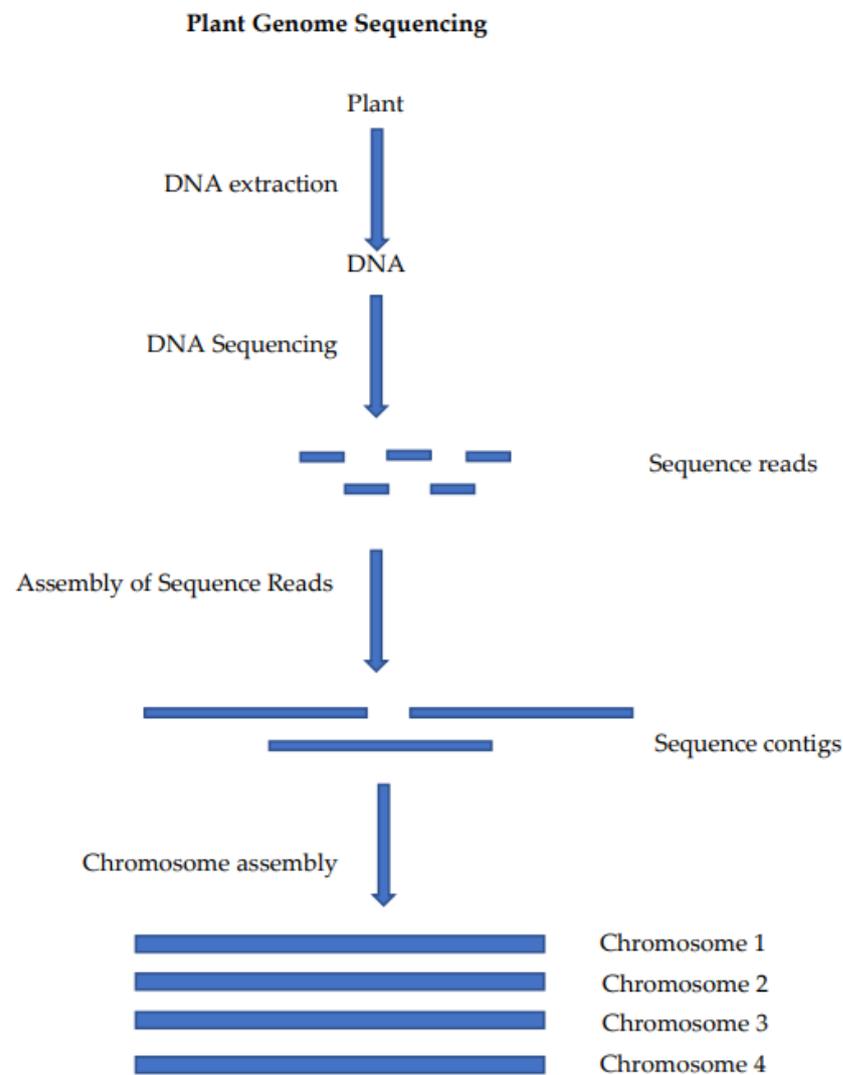
Many diverse plant genomes have now been sequenced with an increasing coverage of the major groups, especially among flowering plants. The coverage of plant orders is high and the genomes from many plant families have now been reported; however, coverage at the genus level is still very low for most plant groups. Systematic efforts to obtain plant genome sequences may take a top-down approach to sequencing a member of each plant family, then each genus and, finally, each species would become available as resources. Ultimately, the re-sequencing of the diversity within each species is of value. A knowledge of the diversity within plant populations is a fundamental tool that can guide the effective conservation of the diversity within species.

#### 3.4. Sequencing Rare and Threatened Species

Targeted efforts are now being made to sequence rare and threatened species of plants as a tool to aid conservation, both in situ [17] and ex situ [18]. This is more urgent among critically endangered species, for which a genome sequence may be all we can retain as the species are lost to extinction. Efforts to sequence biodiversity often focus on rare species as the highest priority.

The critically endangered wild crop relative *Macadamia janseni* has been used to compare plant genome sequencing and assembly methods [19]. This has allowed for the comparison of sequencing platforms and bioinformatics tools for genome assembly using a common sample. The generation of a chromosome-level genome sequence for a plant involves the preparation of a DNA sample, the sequencing of that DNA, the assembly of

the sequence reads into contigs and, finally, the assembly of the sequence contigs into a chromosome-level assembly (Figure 1).



**Figure 1.** Steps in the sequencing and assembly of a plant genome: DNA extraction is used to produce a DNA sample that is suitable for sequencing, the sequencing of the DNA produces long read sequences, the reads are self-assembled into contigs (often at or near chromosome length) and these contigs are then assembled at the chromosome level using chromatin mapping or genetic mapping.

## 4. Sequencing Technology

### 4.1. DNA Isolation

The starting point for the sequencing of plant genomes is obtaining a sample of DNA to sequence. The ease of obtaining a DNA sample of suitable quality for sequencing varies greatly between species. Plants contain many secondary metabolites, proteins and polysaccharides that may interfere with DNA extraction and become a source of contaminants that often reduce the efficiency of DNA extraction. Current technologies require a minimum amount of DNA and the DNA must be pure (free from contaminants that may inhibit sequencing) to facilitate efficient sequencing and the generation of large volumes of data. The amount of DNA required for long read sequencing has been greater (usually  $\mu\text{g}$  quantities) than that needed for short read sequencing (<https://dnatech.genomecenter.ucdavis.edu/pacbio-library-prep-sequencing>; accessed on 6 June 2022). For long read sequencing, the DNA must be intact (not degraded) so that long sequences are present in the sample and can be sequenced to extract long reads.

Simple methods that were suitable for DNA extraction in the past for purposes such as PCR analysis [20] are no longer adequate, so the development of species- [21] or tissue-specific methods that can support next generation sequencing are often required [22]. Some species are especially difficult and require the isolation of the nucleus [23] first as a source of DNA that may be free from contaminants from other parts of the plant cell. The isolation of nuclei in plants is challenging for some of the same reasons that DNA isolation is difficult. The disruption of the plant cell wall requires forces that may damage organelles, such as nuclei and shear DNA.

#### 4.2. Short Read Sequences

The first set of next generation sequencing technologies provided larger volumes of short DNA sequences. The accuracy of these short sequences and the volume of data have since increased dramatically. The length of these sequences started at around 30 bp and has rapidly advanced to 100–150 bp. Paired-end sequencing has extended this technology to allow for the production of sequences of around 400 bp, but most applications currently deliver sequences of around 150 bp. Illumina sequencing platforms are the dominant technology used for short read sequencing. This technology conducts sequencing by synthesis in a very large number of parallel reactions. The incorporation of nucleotides is monitored as the DNA is copied. Other techniques (e.g., solid [24] and 454 sequencing [25]) have been replaced by new technologies because they generally offered lower accuracy or data volumes, resulting in relatively higher costs. Ion Torrent sequencing is used for the rapid determination of a sequence of large numbers of small sequences, such as amplicon sequencing and 16 S metagenomic sequencing. In plants, this has been used for chloroplast sequencing [26].

#### 4.3. Long Read Sequences

The assembly of plant genomes with large numbers of repetitive sequences is not possible with only short read sequences. Therefore, technology that allows much longer sequences to be generated has been key to simplifying genome assembly. The length of these sequences and their accuracy have improved greatly since the technology was first introduced.

##### 4.3.1. PacBio

Pacific Biosciences (PacBio) has developed a long-read sequencing platform that provides accurate long read sequencing. The single-molecule real-time (SMRT) sequencing involves monitoring the incorporation of fluorescent-labeled nucleotides [27]. Recently, single long reads (also known as continuous long reads (CLR)) have largely been replaced by HiFi reads, which provide a consensus sequence based on sequencing a long fragment of DNA (approximately 15,000 bp) multiple times by first circularizing the DNA and reading around the circle many times [28]. The repeated sequencing of the same molecule allows a highly accurate sequence to be generated as the circular consensus sequence (CCS) is read. The quality of the genomes that are generated by the assembly of these reads into contigs has been improved by the application of optimized assembly tools, such as those provided by hifiasm.

##### 4.3.2. ONT

Oxford Nanopore Technologies (ONT) provides a long read sequencing technology that delivers accurate sequence data quickly. The sequence is determined by measuring the changes in electrical currents as the DNA is passed through a pore. The ONT platform generates very long reads and has the advantage of very low instrumentation costs. This platform has continued to improve and deliver very long read sequences with increasing accuracy [29]. ONT sequencing has been widely applied to very rapid sequencing, such as that required for diagnostics [30], due to the advantage of having portable instruments.

The chromosome-level assemblies of plant genomes can be achieved in combination with methods, such as optical mapping [31].

#### 4.3.3. Other

Several technology providers have developed pseudo-long reads that are created by linking short reads. These techniques may produce long reads at lower costs, but the long reads that are generated often do not match the accuracy of the current long read methods [19]. These technologies have been developed by Universal Sequencing Technology [32], MGI [33] and 10× genomics [34]. Despite the great contribution that long read sequencing technology has made to the efficient production of high-quality plant genomes, the emergence of further advances in long read sequencing technologies remains one of the key areas that may contribute to future advances.

#### 4.3.4. Advances

Genome sequencing and assembly requires an adequate depth of sequencing. The size of contigs that can be assembled as long read sequence data has been shown to increase in an almost linear way [8]. The size of the assembled genomes reduces slightly with more contiguous assemblies, probably due to the joining of homologous contig ends, as does the completeness [8]. Improved software has also enabled improvements in the assembly of long read sequences [35]. The use of hifiasm has been shown to allow the haplotype-resolved assembly of the large (30 Gb) genome of the Californian redwood (*Sequoia sempervirens*) [36]. These advances are illustrated by the quality of the early plant genomes relative to those that are being generated by the latest technology. The first rice genome, which was reported in 2002, was highly fragmented while current technology delivers sequence contigs that are often full-length chromosomes [35].

#### 4.4. Chromosome-Level Assembly

The ultimate aim of genome sequencing is to obtain a complete genome sequence of each chromosome, from telomere to telomere. This relies on evidence from beyond the DNA sequence data. Physical and genetic mapping methods have been used to achieve the chromosome-level assembly of contigs that were generated from sequencing data [37]. Recently, the advances in sequencing technology have made it possible to generate many full-length chromosomes from the sequence data alone [35]. The complete assembly of sequence contigs into whole chromosomes has been widely achieved using genetic mapping data, chromatin mapping (Hi-C) or optical mapping. Hi-C [38] involves the mapping of chromatin by crosslinking the DNA in the intact chromatin, digesting the DNA and then sequencing (short reads) the DNA fragments at the ends of the crosslinks. These are used to position the sequence contigs along the chromosome. Optical mapping (Bionano) can also be used to locate sequences along the DNA sequence and to scaffold the sequence contigs [39]. Many projects have combined these technologies to support the generation of high-quality genomes. Recent advances in long read sequencing have enabled the generation of long contigs of highly accurate sequences, reducing reliance on these techniques for high-level assembly. They remain essential for the de novo assembly of most chromosome-level genomes. High-quality sequence contigs in combination with genetic mapping data, Bionano optical data or Hi-C chromatin mapping have generally succeeded in achieving chromosome-level assemblies of plant genomes. A report on more than 100 chromosome-level assemblies in 2021 [40] found only a 73% coverage of the pseudomolecules that represent the chromosomes. The combination of long read sequencing and the use of these tools has resulted in the recent reporting of many high-quality chromosome-level genome sequences (Table 1), with the quality improving greatly along with the most recent technology.

**Table 1.** Some recent chromosome-level assemblies of plant genomes.

	Species	Technique *	Reference
Amorphophallus	<i>Amorphophallus konjac</i>	Hi-C	[41]
Apple	<i>Malus domestica</i>	Genetic Map	[42]
Avocado	<i>Persea americana</i>	Genetic Map	[43]
Banana	<i>Musa balbisiana</i>	Hi-C	[44]
Camphor	<i>Cinnamomum camphora</i>	Hi-C	[45]
Carrot	<i>Daucus carota</i>	Genetic Map	[46]
Chinese Skullcap	<i>Scutellaria baicalensis</i>	Hi-C	[47]
Crucihimalaya	<i>Crucihimalaya lasiocarpa</i>	Hi-C	[48]
Cucumber	<i>Cucumis metuliferus</i>	Hi-C	[49]
Eucalypt	<i>Corymbia citriodora</i>	Genetic Map	[37]
Field Pennycress	<i>Thlaspi arvense</i>	Genetic Map/Hi-C/Bionano	[50]
Ginger	<i>Zingiber officinale</i>	Hi-C	[51]
Ginkgo	<i>Ginkgo biloba</i>	Hi-C	[52]
Jojoba	<i>Simmondsia chinensis</i>	Hi-C	[53]
Macadamia	<i>Macadamia jansonii</i>	Hi-C	[54]
	<i>Macadamia integrifolia</i>	Genetic Map	[55]
Paper Mulberry	<i>Broussonetia papyrifera</i>	Hi-C	[56]
Peach	<i>Prunus persica</i>	Hi-C	[57]
Peanut	<i>Arachis hypogaea</i>	Hi-C	[58]
Speranskia	<i>Speranskia yunnanensis</i>	Hi-C	[59]
Taxus	<i>Taxus chinensis</i>	Hi-C	[60]
Tea	<i>Camellia sinensis</i>	Hi-C	[61]
Water Caltrop	<i>Trapa spp.</i>	Hi-C	[62]

\* The technique used to achieve the chromosome-level assembly of the sequence contigs.

#### 4.5. Haplotype-Resolved Genomes

Most published plant genomes are collapsed representations of the diploid genome as a single sequence, with a random inclusion of one of the two alleles at each heterozygous position. Only recently has it become possible to assemble each haplotype separately [63]. This has been the result of advances in both sequencing technology and sequence assembly tools. Current technology suggests that most genomes can now be sequenced at the haplotype level, thereby replacing the reporting of collapsed genomes with the sequences of the two haplotypes.

#### 4.6. Pan-Genomes

The sequencing of plant genomes has shown that significant differences may be found within a plant species, which means that more than one reference genome is required to represent the species. The sequencing of plant genomes has also demonstrated that many genes are variably present in different individuals within a species. These presence/absence differences have led to the construction of pan-genomes, which represent the complete set of genes found within a population. A genome that includes all of the variations within a group of plants is known as a pan-genome. The pan-genome concept is a powerful tool for plant breeders for the analysis of gene pools [64]. Pan-genomes can be generated at different levels to represent the diversity that is found within, for example, domesticated gene pools, species or genera.

#### 4.7. Transcriptomes

Transcriptome sequencing is an important tool for the analysis of the expressed regions of a genome. This is key to understanding gene functions and the determination of the genetic basis of important plant traits [65–70]. Transcriptome sequencing complements genome sequencing in genome characterization. Transcripts provide physical evidence that the sequence is formed of the expressed and complementary predictions of genes, based on the sequence alone. The comparison of the transcriptomes of different genotypes

from different tissues or cell types at different stages of development and under different environmental conditions allows for the discovery of the genes that control plant traits and has become a key approach in plant biology and the discovery of genes for selection in plant breeding. Single-cell transcriptomics has become a powerful tool for understanding gene expression at the cell and tissue level but has had limited application in plants [71], partly due to the difficulty in isolating specific plant cells without disrupting expression.

#### 4.7.1. RNAseq

The quantitative analysis of the levels of expression of genes in any specific cell, tissue, organ, genotype or development stage is widely determined by RNA sequencing (RNAseq) [72,73]. RNAseq has largely replaced earlier array-based or gene by gene analysis tools as it provided a more unbiased analysis of the whole transcriptome.

An analysis of the gene expression in the highly polyploid sugarcane genome revealed that while the different alleles of most genes are expressed in direct proportion to their abundance in the genome, some genes show highly biased patterns of expression [74]. In hexaploid wheat, subgenome-specific responses to diseases have also been reported [75].

#### 4.7.2. Long Read Transcriptomes

Long read sequencing is a method that has been applied to the analysis of plant transcriptomes, which reveals the diversity of full-length transcripts and defines the variations in splicing and intron retention in gene expression [76]. The long read sequencing of transcriptomes avoids the challenge of the assembly of many closely related transcripts from short reads. Unique 3' and 5' sequences may be separated by common intervening sequences, which creates the risk of incorrectly combining the ends of the transcripts when using short reads.

Some examples of the application of long read sequencing to the analysis of plant transcriptomes of increasing complexity can be found for polyploid species in Table 2.

**Table 2.** The long read sequencing of polyploid transcriptomes.

Species	Ploidy	Reference
Coffee	4X	[76]
Wheat	6X	[77]
Strawberry	8X	[78]
Sugarcane	12X	[79]

### 4.8. Organelle Genome Sequencing

Plant cells usually contain a single nucleus and many organelles, probably hundreds of mitochondria and thousands of chloroplasts. Sequencing the organelle genomes is complicated by the transfer of genes between these genomes. The nuclear genome often contains many insertions of large and small sequences of organellar genomes. Many early methods struggled to distinguish organellar gene sequences from those of copies that were inserted into the nuclear genome because they relied on PCR amplification or organelle separation [80]. Nuclear inserts may represent versions of organellar genomes that were transferred in the past and that have diverged since insertion.

#### 4.8.1. Chloroplast Genomes

The chloroplast genomes of plants are highly conserved sequences of 100–150 Kb, containing around 100 genes [81]. The structure of most chloroplasts is similar, with four components including inverted repeats that separate large and small single-copy regions. Chloroplasts have been widely used in plant identification due to their presence in all green plants and the high copy numbers in the cell that simplify the detection of chloroplast sequences. Early approaches that relied on chloroplast isolation or PCR amplification were plagued by confusion due to the copies of chloroplast sequences in the nuclear

and mitochondrial genomes. Recent approaches [82] rely on the higher abundance of chloroplast genome sequence reads in short read sequence data to clearly distinguish the correct sequence of the relevant chloroplast [83]. The development of software tools now allows for the efficient extraction of accurate whole chloroplast genome sequences from even low (nuclear) coverage sequencing datasets. The annotation of chloroplast genomes that were generated in this way has resulted in the identification of around 100 genes with increasingly well-defined functions [84].

The sequencing of the maternal (e.g., chloroplast) and nuclear genomes of plants has frequently revealed discordant phylogenies [85–87], suggesting widespread reticulate evolution in plant populations (Table 3). Chloroplast transfers between species during rare events results in “chloroplast capture” by closely related species.

**Table 3.** Discordant phylogenies for chloroplast and nuclear genome sequences.

Species	Reference
Apple Tribe	[88]
Eucalypts	[87]
Osmorhiza spp.	[89]
<i>Pedicularis</i> spp.	[90]
Sorghum	[85]
Rice ( <i>Oryza</i> )	[86]

#### 4.8.2. Plant Mitochondrial Genomes

The mitochondrial genomes of plants [91] are much larger and less conserved than the chloroplast genomes and as a result, they have been much less studied than chloroplasts. The mitochondrial genome, as with the nuclear genome, may include sequences that were derived from the chloroplast that have been inserted into the genome at various times throughout its evolutionary history. Due to the relatively higher number of mitochondrial genomes in cells, these sequences are even more likely to be confused with chloroplast genome sequences than chloroplast sequences that were inserted into nuclear genomes.

### 5. Biological Understanding

Sequencing plant genomes provides an enhanced understanding of the biology of plant species. This knowledge informs the better conservation of biodiversity and sustainable use in agriculture and food production. Plant genomes may often explain the response of plants to the environment and may assist in improving the management of crops.

#### 5.1. Whole Genome Duplications

Genome sequencing and assembly defines the presence of duplicated regions of the genome that are often the result of whole genome duplications during the evolutionary history of the species. In many species, evidence can be found for more than one duplication event. The analysis of these events can aid in the determination of evolutionary relationships. The selection of key genomes for sequencing allows evolutionary relationships to be defined. For example, the sequencing of the *Aristolochia fimbriata* genome revealed that this plant lacked the whole genome duplication that was reported in other magnoliid plants, placing it at a basal position in the angiosperm phylogenies [92].

#### 5.2. Polyploid Challenges

The sequencing and assembly of a polyploid plant genome is complicated by the presence of many similar sequences that can be difficult to assemble. Sugarcane is an important global crop, but the high degree of ploidy (12X) in this species has resulted in it being one of the last major crop species to have a sequenced genome. Instead, a monoploid equivalent (based on BAC clones from sugarcane that cover the sorghum genome) has been widely used in sugarcane genomes due to the lack of a polyploid genome [5].

While a genome for the diploid Robusta coffee was reported in 2014 [93], the sequencing of the tetraploid Arabica coffee has been more difficult due to a genome that was based on a doubled diploid currently being used to characterize the origins of Arabica coffee (Arabica Genome Sequencing Consortium).

### 5.3. Genomics of Plants with Diverse Reproductive Biology

While most flowering plants are hermaphrodites with both male and female reproductive organs, separate male and female plants are found in dioecious species. There are dioecious plants in many plant families and they represent around 6% of all plant species. The differences between the genomes of male and female dioecious plants have not been fully characterized. The recent sequencing of the male and female genomes of jojoba revealed large differences between the sex chromosomes, with the presence of many sex-specific genes [53]. The chromosome-level assembly of male and female genomes has provided a perspective on the basis of sexual dimorphism that was not possible with more limited genomic information. The sequencing of the genomes of more dioecious plants may define the diversity of dimorphisms that have evolved into the many separate lineages of dioecious members. The genomes of other plants with unusual reproductive biology, such as apomictic plants, have not yet been studied but may explain the adaptive value of these modes of reproduction. An analysis of transcriptomes defined the conserved genes for organogenesis that are associated with reproduction in flowering plants [94].

### 5.4. Evolutionary Insights

The comparison of the genome sequences of plant species is the key basis that we have for defining evolutionary relationships. The phylogenetic comparison of plant genomes has often been based on the analysis of one or a few genes. The availability of whole genome sequence data has allowed the use of much larger numbers of genes. A common approach has been to compare the sequences of many conserved single-copy genes. This approach has been used to define relationships between species in the rice (*Oryza*) [86,95] and sorghum (*Sorghum*) [85] genera.

Genome sequencing provides an opportunity to better understand the process of domestication, through which human selection has resulted in plants that are better suited to survival in agricultural environments rather than through natural selection in nature. The sequencing of the coffee genome revealed that human attraction to caffeine might have resulted in the domestication of plants for coffee, tea and chocolate production from diverse species in which caffeine had evolved separately [93].

### 5.5. Maternal Genome Inheritance

In most plants, the cytoplasmic organelles (chloroplasts and mitochondria) are maternally inherited. In some rare cases, other patterns of inheritance are observed; for example, in cucumber (*Cucumis sativus*), the chloroplasts are inherited maternally while the mitochondria are paternally inherited [96]. Although some plant groups display paternal inheritance, the maternal inheritance of organellar DNA and the paternal contribution to nuclear DNA can result in discordant patterns of evolution. Reticulate evolution, in which there is a gene flow between different lineages, may result in the transfer (or “capture”) of organellar genomes from closely related species. Many examples have been reported via chloroplast genome analyses that suggest phylogenies that differ from those indicated by the nuclear genomes [87]. The potential for the nuclear, chloroplast and mitochondrial genomes of plants to have evolved along separate paths makes it important to focus phylogenetic analysis on nuclear genomes. Improved methods for applying phylogenetic analysis to whole nuclear genomes are needed. Because nuclear genomes show so many variations at the sequence level, the current approaches rely on aligning a small subset of highly conserved genes from nuclear genomes that can be reliably aligned. As whole genome sequence data become more widely available, it may be possible to develop tools that take advantage of the larger volumes of data to better determine relationships through phylogenetic analysis.

### 5.6. Importance of Genome Size

Plant genomes vary widely in size, with more than a 1000-fold variation among flowering plants. Even within a single plant family, genome sizes may vary more than 100-fold. The biological significance of these variations remains poorly understood. Differences may be due to gene content, genome duplications, polyploidy or differences in repetitive DNA content. Improvements in genome sequencing technology are likely to allow larger numbers of diverse genomes to be sequenced to facilitate our understanding of genome size diversity.

## 6. Enabling Plant Breeding

The availability of plant genomes facilitates the breeding and selection of plants, which is essential to support ongoing food security. The need for an accelerated genetic improvement of plants has been made more urgent by climate change, which is demanding new plant genotypes that are adapted to new and more difficult environments. Climate change is altering the physical environment, with higher temperatures and greater water stress, and it is also changing the biological environment, with a wider range of pests and diseases [97].

Plant genomics is critical for the management of plant genetic resources in seed banks [18] and the conservation of the wild crop relatives that provide the genetic resources that are required for sustainable food production.

Plant genomes also enable an understanding of plant biology and the molecular and genetic basis of plant traits.

The analysis of plant genome sequences has facilitated the rapid identification of the key genes that determine the traits of great importance in food crops. This has allowed for more efficient crop breeding by simplifying the selection of these critical traits. The completion of the rice genome allowed for the discovery of the identity of the recessive gene for fragrance in rice [98], which is a trait that can double the value of the rice and can now be detected with a perfect marker [99]. The cooking temperature (gelatinization temperature) of rice was also shown to be determined by a major gene [100]. In wheat, transcriptome analysis has revealed the genetic basis of flour yield [101] and loaf volume for bread [102].

### 6.1. Molecular Markers and Plant Selection

Sequencing has changed the approaches to marker development and applications in plant breeding. Traditional molecular markers [103], which were linked to traits by being close to the genes that control the phenotype, are now able to be replaced by sequence-based markers [104] for the differences in genomes that are actually responsible for the traits under selection. This greatly increases the reliability of the selection. As the costs for whole genome sequencing reduce, this becomes an alternative to the analysis of large numbers of markers. In a sense, genome sequences are the ultimate genetic marker tool as they capture all of the variations within the genome that may explain the phenotype. They also avoid decisions on which genetic markers to select for analysis needing to be made in advance by capturing all of the possibilities.

### 6.2. Genetic Manipulation

Genome sequencing is the ultimate method for the characterization of genetically modified plants, revealing the exact changes in genomes that have been produced. This may be a requirement for release into the environment in some jurisdictions. The sequencing of transgenic genotypes defines both the exact point of insertion and the sequence of the added genes, but it also defines the number of copies that have been added and may reveal any other genetic changes that might have been a result of cell culture [105].

### 6.3. Editing Plant Genomes

The growing application of genome editing is being aided by the use of genome sequencing [106] to support the better targeting of gene editing and to ensure the avoidance of off-target effects. The routine sequencing of the genomes of transgenic- or gene-edited plants can also be used to confirm that the intended changes have been made and that no other unintended changes have occurred. Often, only a single nucleotide needs to be edited or many different loci need to be targeted simultaneously to make the required change in a phenotype; however, in each case, it may be necessary to sequence the genome to confirm that other changes have not occurred.

### 6.4. Biotechnology Applications (Food, Medicinal and Industrial Crops)

Genome sequencing is a key tool that enables the rapid production of plant varieties with higher nutritional value, enhanced levels of bioactivity, improved biomass composition [107] or the expression of high-value molecules. The availability of sequencing technology allows for the identification of novel genotypes with the desired traits and supports the manipulation of plant genomes to produce plants with novel traits. Genome sequencing can be used to confirm the results of gene editing or any other changes resulting from plant transformation, mutagenesis or conventional crossbreeding. Sequencing may be critical for ensuring compliance with regulations that govern plant genetic manipulation. This may become a major application for plant genome sequencing as gene editing becomes more routine. This application for plant genome sequencing may require the development of standard protocols that can be applied routinely, both in research and in industry, for quality assurance and to protect intellectual property.

## 7. IP Issues

An issue of growing importance is the ownership and control of genome sequence data [108]. Modern biological science has been built upon the widespread dissemination of sequence data by providing public access to large sequence databases. The Convention on Biological Diversity empowered countries to claim ownership of their biological resources and the more recent Nagoya protocol requires the consideration of access and benefit sharing. These rights may extend to DNA sequence data that were derived from genetic resources. These issues are especially difficult for historical collections because prior informed consent cannot be obtained for the sample collection [109]. International efforts to resolve these issues are urgently required to balance the rights of traditional owners with the need for open access to sequence data to advance biological science. The protection of plant varieties through the use of plant breeders' rights (PBRs) may be supported by the use of genome sequence data to confirm the identity of genotypes. Establishing the distinctness of genotypes to secure PBRs may also become more dependent on genome sequence data. Genome sequence data may be critical for determining whether a new variety was essentially derived from an earlier variety, which is a question that becomes more important as genome editing becomes more widely used.

## 8. Future Prospects

Many visionary projects aim to ultimately sequence all of the planet's biodiversity as a long-term goal. The sequencing of rare and endangered species can be considered as a priority in this process. Advances in technology have made it difficult to determine when to start such a project as the costs continue to drop and the quality of the data continues to improve. The achievement of plant genome sequences for all species on the planet could accelerate as sequencing technology advances and data storage and handling become more effective.

The difficulty in obtaining a suitable sample of DNA from a plant is one of the remaining challenges in the widespread application of the sequencing of plant genomes. The development of more general methods that can be applied to a wide range of plant

samples would represent a major advance, unless DNA sequencing methods become more robust and can cope with poorer quality DNA preparations.

The technology that is now available for plant genome sequencing and assembly make this an increasingly cost-effective strategy for improving our understanding of the biology of all plant species and a key tool for the conservation of plant biodiversity and the use of plants in agriculture and food production. The sequencing of all plant species is a long-term goal that may become key to effectively supporting life on Earth through the improved management of plants in wild populations and their selection and genetic enhancement for use in agriculture and food production. Threats to food security from human conflicts and pandemics [110] have created more interest in food supply from local sources. Plant genome sequencing provides a platform for innovation in plant breeding to deliver a diverse and balanced diet regionally. Adaptations to climate change require the development of plant varieties that can support the adaptation and relocation of agriculture and the development of plant varieties for production in vertical farming [111]. Genomics is a key tool for tackling climate change [112] and for capturing a wider range of diversity from wild crop relatives [113] and other plants to support food production [114].

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The author acknowledges the support of the ARC Centre of Excellence for Plant Success in Nature and Agriculture.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

- Henry, R.J. Applied Biosciences: Application of Biological Science and Technology. *Appl. Biosci.* **2022**, *1*, 38–39. [[CrossRef](#)]
- Shendure, J.; Balasubramanian, S.; Church, G.M.; Gilbert, W.; Rogers, J.; Schloss, J.; Waterston, R.H. DNA sequencing at 40: Past, present and future. *Nature* **2017**, *550*, 345–353. [[CrossRef](#)] [[PubMed](#)]
- Marks, R.A.; Hotaling, S.; Frandsen, P.B.; VanBuren, R. Representation and participation across 20 years of plant genome sequencing. *Nat. Plants* **2021**, *7*, 1571–1578. [[CrossRef](#)]
- Yüksel, B.; Paterson, A.H. Construction and characterization of a peanut HindIII BAC library. *Theor. Appl. Genet.* **2005**, *111*, 630–639. [[CrossRef](#)]
- Garsmeur, O.; Droc, G.; Antonise, R.; Grimwood, J.; Potier, B.; Aitken, K.; Jenkins, J.; Martin, G.; Charron, C.; Hervouet, C.; et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* **2018**, *9*, 2638. [[CrossRef](#)] [[PubMed](#)]
- Pucker, B.; Irisarri, I.; de Vries, J.; Xu, B. Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quant. Plant Biol.* **2021**, *3*, 1–14. [[CrossRef](#)]
- Chen, F.; Dong, W.; Zhang, J.; Guo, X.; Chen, J.; Wang, Z.; Lin, Z.; Tang, H.; Zhang, L. The Sequenced Angiosperm Genomes and Genome Databases. *Front. Plant Sci.* **2018**, *9*, 418. [[CrossRef](#)]
- Sharma, P.A.O.; Alsubaie, B.; Al-Mssallem, I.; Nath, O.; Mitter, N.; Margarido, G.R.A.; Topp, B.; Murigneux, V.; Masouleh, A.K.; Furtado, A.; et al. Improvements in The Sequencing and Assembly of Plant Genomes. *Gigabyte* **2021**. [[CrossRef](#)]
- Wendel, J.F.; Jackson, S.A.; Meyers, B.C.; Wing, R.A. Evolution of plant genome architecture. *Genome Biol.* **2016**, *17*, 37. [[CrossRef](#)]
- Pellicer, J.; Hidalgo, O.; Dodsworth, S.; Leitch, I.J. Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes* **2018**, *9*, 88. [[CrossRef](#)]
- Uddenberg, D.; Akhter, S.; Ramachandran, P.; Sundström, J.F.; Carlsbecker, A. Sequenced genomes and rapidly emerging technologies pave the way for conifer evolutionary developmental biology. *Front. Plant Sci.* **2015**, *6*, 970. [[CrossRef](#)] [[PubMed](#)]
- Liu, H.; Wang, X.; Wang, G.; Cui, P.; Wu, S.; Ai, C.; Hu, N.; Li, A.; He, B.; Shao, X.; et al. The nearly complete genome of Ginkgo biloba illuminates gymnosperm evolution. *Nat. Plants* **2021**, *7*, 748–756. [[CrossRef](#)] [[PubMed](#)]
- Kaul, S.; Koo, H.L.; Jenkins, J.; Rizzo, M.; Rooney, T.; Tallon, L.J.; Feldblyum, T.; Nierman, W.; Benito, M.I.; Lin, X.Y.; et al. Analysis of The Genome Sequence of The Flowering Plant *Arabidopsis Thaliana*. *Nature* **2000**, *408*, 796–815.
- Jackson, S.A. Rice: The First Crop Genome. *Rice* **2016**, *9*, 14. [[CrossRef](#)] [[PubMed](#)]
- Vogel, J.P.; Garvin, D.F.; Mockler, T.C.; Schmutz, J.; Rokhsar, D.; Bevan, M.W.; Barry, K.; Lucas, S.; Harmon-Smith, M.; Lail, K.; et al. Genome Sequencing and Analysis of the Model Grass *Brachypodium distachyon*. *Nature* **2010**, *463*, 763–768.
- Kersey, P.J. Plant genome sequences: Past, present, future. *Curr. Opin. Plant Biol.* **2018**, *48*, 1–8. [[CrossRef](#)]

17. Wambugu, P.W.; Henry, R.; Browne, L. Supporting in situ conservation of the genetic diversity of crop wild relatives using genomic technologies. *Mol. Ecol.* **2022**, *31*, 2207–2222. [[CrossRef](#)]
18. Wambugu, P.W.; Ndjiondjop, M.-N.; Henry, R.J. Role of genomics in promoting the utilization of plant genetic resources in genebanks. *Brief. Funct. Genom.* **2018**, *17*, 198–206. [[CrossRef](#)]
19. Murigneux, V.; Rai, S.K.; Furtado, A.; Bruxner, T.J.C.; Tian, W.; Harliwong, I.; Wei, H.; Yang, B.; Ye, Q.; Anderson, E.; et al. Comparison of long-read methods for sequencing and assembly of a plant genome. *GigaScience* **2020**, *9*, gaa146. [[CrossRef](#)]
20. Graham, G.C.; Mayers, P.; Henry, R.J. A simplified method for the preparation of fungal genomic DNA for PCR and RAPD analysis. *BioTechniques* **1994**, *16*, 48–50.
21. Nath, O.; Fletcher, S.J.; Hayward, A.; Shaw, L.M.; Agarwal, R.; Furtado, A.; Henry, R.J.; Mitter, N. A Comprehensive High-Quality DNA and RNA Extraction Protocol for a Range of Cultivars and Tissue Types of the Woody Crop Avocado. *Plants* **2022**, *11*, 242. [[CrossRef](#)] [[PubMed](#)]
22. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **2014**, *10*, 21. [[CrossRef](#)] [[PubMed](#)]
23. Guilfoyle, T.J. Chapter 8 Isolation and Characterization of Plant Nuclei. *Method Cell Biol.* **1995**, *50*, 101–112. [[CrossRef](#)]
24. Garrido-Cardenas, J.A.; Garcia-Maroto, F.; Alvarez-Bermejo, J.A.; Manzano-Agugliaro, F. DNA Sequencing Sensors: An Overview. *Sensors* **2017**, *17*, 588. [[CrossRef](#)] [[PubMed](#)]
25. Bundock, P.C.; Elliott, F.G.; Ablett, G.; Benson, A.D.; Casu, R.E.; Aitken, K.S.; Henry, R.J. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnol. J.* **2009**, *7*, 347–354. [[CrossRef](#)] [[PubMed](#)]
26. Brozynska, M.; Furtado, A.; Henry, R.J. Direct Chloroplast Sequencing: Comparison of Sequencing Platforms and Analysis Tools for Whole Chloroplast Barcoding. *PLoS ONE* **2014**, *9*, e110387. [[CrossRef](#)]
27. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinf.* **2015**, *13*, 278–289. [[CrossRef](#)]
28. Hon, T.; Mars, K.; Young, G.; Tsai, Y.-C.; Karalius, J.W.; Landolin, J.M.; Maurer, N.; Kudrna, D.; Hardigan, M.A.; Steiner, C.C.; et al. Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci. Data* **2020**, *7*, 399. [[CrossRef](#)]
29. Wang, Y.H.; Zhao, Y.; Bollas, A.; Wang, Y.R.; Au, K.F. Nanopore Sequencing Technology, Bioinformatics and Applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [[CrossRef](#)]
30. Wang, J.; Bhakta, N.; Miller, V.A.; Revsine, M.; Litzow, M.; Paietta, E.; Roberts, K.; Gu, Z.; Mullighan, C.; Jones, C.; et al. Nanopore RNA Sequencing as A Diagnostic Tool for Acute Leukemia in Low Resource Settings. *Pediatr. Blood Cancer* **2021**, *68*, S107.
31. Belsler, C.; Istace, B.; Denis, E.; Dubarry, M.; Baurens, F.-C.; Falentin, C.; Genete, M.; Berrabah, W.; Chèvre, A.-M.; Delourme, R.; et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* **2018**, *4*, 879–887. [[CrossRef](#)] [[PubMed](#)]
32. Chen, Z.; Pham, L.; Wu, T.-C.; Mo, G.; Xia, Y.; Chang, P.L.; Porter, D.; Phan, T.; Che, H.; Tran, H.; et al. Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. *Genome Res.* **2020**, *30*, 898–909. [[CrossRef](#)] [[PubMed](#)]
33. Wang, O.; Chin, R.; Cheng, X.; Wu, M.K.Y.; Mao, Q.; Tang, J.; Sun, Y.; Anderson, E.; Lam, H.K.; Chen, D.; et al. Efficient and unique cobarcodeing of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* **2019**, *29*, 798–808. [[CrossRef](#)] [[PubMed](#)]
34. Singh, M.; Al-Eryani, G.; Carswell, S.; Ferguson, J.M.; Blackburn, J.; Barton, K.; Roden, D.; Luciani, F.; Phan, T.G.; Junankar, S.; et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* **2019**, *10*, 3120. [[CrossRef](#)] [[PubMed](#)]
35. Sharma, P.; Masouleh, A.K.; Topp, B.; Furtado, A.; Henry, R.J. De novo chromosome level assembly of a plant genome from long read sequence data. *Plant J.* **2021**, *109*, 727–736. [[CrossRef](#)]
36. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [[CrossRef](#)]
37. Healey, A.L.; Shepherd, M.; King, G.J.; Butler, J.B.; Freeman, J.S.; Lee, D.J.; Potts, B.M.; Silva-Junior, O.B.; Baten, A.; Jenkins, J.; et al. Pests, diseases, and aridity have shaped the genome of *Corymbia citriodora*. *Commun. Biol.* **2021**, *4*, 537. [[CrossRef](#)]
38. Harewood, L.; Kishore, K.; Eldridge, M.; Wingett, S.; Pearson, D.; Schoenfelder, S.; Collins, V.P.; Fraser, P. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* **2017**, *18*, 125. [[CrossRef](#)]
39. Chen, P.; Jing, X.; Ren, J.; Cao, H.; Hao, P.; Li, X. Modelling BioNano optical data and simulation study of genome map assembly. *Bioinformatics* **2018**, *34*, 3966–3974. [[CrossRef](#)]
40. Shirasawa, K.; Harada, D.; Hirakawa, H.; Isobe, S.; Kole, C. Chromosome-level de novo genome assemblies of over 100 plant species. *Breed. Sci.* **2021**, *71*, 117–124. [[CrossRef](#)]
41. Gao, Y.; Zhang, Y.; Feng, C.; Chu, H.; Feng, C.; Wang, H.; Wu, L.; Yin, S.; Liu, C.; Chen, H.; et al. A chromosome-level genome assembly of *Amorphophallus konjac* provides insights into konjac glucomannan biosynthesis. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1002–1011. [[CrossRef](#)] [[PubMed](#)]
42. Sun, X.; Jiao, C.; Schwaninger, H.; Chao, C.T.; Ma, Y.; Duan, N.; Khan, A.; Ban, S.; Xu, K.; Cheng, L.; et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* **2020**, *52*, 1423–1432. [[CrossRef](#)] [[PubMed](#)]

43. Rendón-Anaya, M.; Ibarra-Laclette, E.; Méndez-Bravo, A.; Lan, T.; Zheng, C.; Carretero-Paulet, L.; Perez-Torres, C.A.; Chacón-López, A.; Hernandez-Guzmán, G.; Chang, T.-H.; et al. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17081–17089. [[CrossRef](#)] [[PubMed](#)]
44. Wang, Z.; Miao, H.; Liu, J.; Xu, B.; Yao, X.; Xu, C.; Zhao, S.; Fang, X.; Jia, C.; Wang, J.; et al. Musa balbisiana genome reveals subgenome evolution and functional divergence. *Nat. Plants* **2019**, *5*, 810–821. [[CrossRef](#)] [[PubMed](#)]
45. Shen, T.; Qi, H.; Luan, X.; Xu, W.; Yu, F.; Zhong, Y.; Xu, M. The chromosome-level genome sequence of the camphor tree provides insights into Lauraceae evolution and terpene biosynthesis. *Plant Biotechnol. J.* **2021**, *20*, 244–246. [[CrossRef](#)] [[PubMed](#)]
46. Iorizzo, M.; Ellison, S.; Senalik, D.; Zeng, P.; Satapoomin, P.; Huang, J.; Bowman, M.; Iovene, M.; Sanseverino, W.; Cavagnaro, P.; et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **2016**, *48*, 657–666. [[CrossRef](#)] [[PubMed](#)]
47. Zhao, Q.; Yang, J.; Cui, M.-Y.; Liu, J.; Fang, Y.; Yan, M.; Qiu, W.; Shang, H.; Xu, Z.; Yidiresi, R.; et al. The Reference Genome Sequence of *Scutellaria baicalensis* Provides Insights into the Evolution of Wogonin Biosynthesis. *Mol. Plant* **2019**, *12*, 935–950. [[CrossRef](#)]
48. Feng, L.; Lin, H.; Kang, M.; Ren, Y.; Yu, X.; Xu, Z.; Wang, S.; Li, T.; Yang, W.; Hu, Q. A chromosome-level genome assembly of an alpine plant *Crucihimalaya lasiocarpa* provides insights into high-altitude adaptation. *DNA Res.* **2022**, *29*, dsac004. [[CrossRef](#)]
49. Ling, J.; Xie, X.X.; Gu, X.F.; Zhao, J.L.; Ping, X.X.; Li, Y.; Yang, Y.H.; Mao, Z.C.; Xie, B.Y. High-quality chromosome-level genomes of *Cucumis metuliferus* and *Cucumis melo* provide insight into *Cucumis* genome evolution. *Plant J.* **2021**, *107*, 136–148. [[CrossRef](#)]
50. Nunn, A.; Rodríguez-Arévalo, I.; Tandukar, Z.; Frels, K.; Contreras-Garrido, A.; Carbonell-Bejerano, P.; Zhang, P.; Cruz, D.R.; Jandrasits, K.; Lanz, C.; et al. Chromosome-level *Thlaspi arvense* genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. *Plant Biotechnol. J.* **2022**, *20*, 944–963. [[CrossRef](#)]
51. Li, H.L.; Wu, L.; Dong, Z.M.; Jiang, Y.S.; Jiang, S.J.; Xing, H.T.; Li, Q.; Liu, G.C.; Tian, S.M.; Wu, Z.Y.; et al. Haplotype-Resolved Genome of Diploid Ginger (*Zingiber Officinale*) and Its Unique Gingerol Biosynthetic Pathway. *Hortic. Res.* **2021**, *8*, 189. [[CrossRef](#)] [[PubMed](#)]
52. Shan, H.; Kong, H. The genome of *Ginkgo biloba* refined. *Nat. Plants* **2021**, *7*, 714–715. [[CrossRef](#)] [[PubMed](#)]
53. Al-Dossary, O.; Alsubaie, B.; Kharabian-Masouleh, A.; Al-Mssallem, I.; Furtado, A.; Henry, R.J. The jojoba genome reveals wide divergence of the sex chromosomes in a dioecious plant. *Plant J.* **2021**, *108*, 1283–1294. [[CrossRef](#)] [[PubMed](#)]
54. Sharma, P.; Murigneux, V.; Haimovitz, J.; Nock, C.J.; Tian, W.; Masouleh, A.K.; Topp, B.; Alam, M.; Furtado, A.; Henry, R.J. The genome of the endangered *Macadamia jansonii* displays little diversity but represents an important genetic resource for plant breeding. *Plant Direct* **2021**, *5*, e364. [[CrossRef](#)] [[PubMed](#)]
55. Nock, C.J.; Baten, A.; Mauleon, R.; Langdon, K.S.; Topp, B.; Hardner, C.; Furtado, A.; Henry, R.J.; King, G.J. Chromosome-Scale Assembly and Annotation of the Macadamia Genome (*Macadamia integrifolia* HAES 741). *G3-Genes Genomes Genet.* **2020**, *10*, 3497–3504. [[CrossRef](#)]
56. Peng, X.; Liu, H.; Chen, P.; Tang, F.; Hu, Y.; Wang, F.; Pi, Z.; Zhao, M.; Chen, N.; Chen, H.; et al. A Chromosome-Scale Genome Assembly of Paper Mulberry (*Broussonetia papyrifera*) Provides New Insights into Its Forage and Papermaking Usage. *Mol. Plant* **2019**, *12*, 661–677. [[CrossRef](#)] [[PubMed](#)]
57. Lian, X.D.; Zhang, H.P.; Jiang, C.; Gao, F.; Yan, L.; Zheng, X.B.; Cheng, J.; Wang, W.; Wang, X.B.; Ye, X.; et al. De Novo Chromosome-Level Genome of A Semi-Dwarf Cultivar of Prunus Persica Identifies The Aquaporin Pptip2 as Responsible for Temperature-Sensitive Semi-Dwarf Trait And Ppb3-1 for Flower Type And Size. *Plant Biotechnol. J.* **2022**, *20*, 886–902. [[CrossRef](#)]
58. Zhuang, W.; Chen, H.; Yang, M.; Wang, J.; Pandey, M.K.; Zhang, C.; Chang, W.-C.; Zhang, L.; Zhang, X.; Tang, R.; et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* **2019**, *51*, 865–876. [[CrossRef](#)]
59. Yuan, G.; Tan, S.; Wang, D.; Yang, Y.; Tian, B. Chromosome-Level Genome Assembly of the Rare and Endangered Tropical Plant *Speranskia yunnanensis* (Euphorbiaceae). *Front. Genet.* **2022**, *12*, 755564. [[CrossRef](#)]
60. Xiong, X.; Gou, J.; Liao, Q.; Li, Y.; Zhou, Q.; Bi, G.; Li, C.; Du, R.; Wang, X.; Sun, T.; et al. The Taxus genome provides insights into paclitaxel biosynthesis. *Nat. Plants* **2021**, *7*, 1026–1036. [[CrossRef](#)]
61. Zhang, X.; Chen, S.; Shi, L.; Gong, D.; Zhang, S.; Zhao, Q.; Zhan, D.; Vasseur, L.; Wang, Y.; Yu, J.; et al. Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat. Genet.* **2021**, *53*, 1250–1259. [[CrossRef](#)] [[PubMed](#)]
62. Lu, R.; Chen, Y.; Zhang, X.; Feng, Y.; Comes, H.P.; Li, Z.; Zheng, Z.; Yuan, Y.; Wang, L.; Huang, Z.; et al. Genome sequencing and transcriptome analyses provide insights into the origin and domestication of water caltrop (*Trapa* spp., Lythraceae). *Plant Biotechnol. J.* **2022**, *20*, 761–776. [[CrossRef](#)] [[PubMed](#)]
63. Low, W.Y.; Tearle, R.; Liu, R.; Koren, S.; Rhie, A.; Bickhart, D.; Rosen, B.D.; Kronenberg, Z.N.; Kingan, S.B.; Tseng, E.; et al. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat. Commun.* **2020**, *11*, 2071. [[CrossRef](#)] [[PubMed](#)]
64. Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and Exploiting Pan-genomics for Crop Improvement. *Mol. Plant* **2018**, *12*, 156–169. [[CrossRef](#)] [[PubMed](#)]
65. Cheng, B.; Furtado, A.; Henry, R.J. The coffee bean transcriptome explains the accumulation of the major bean components through ripening. *Sci. Rep.* **2018**, *8*, 11414. [[CrossRef](#)]

66. Gillies, S.A.; Futardo, A.; Henry, R.J. Gene expression in the developing aleurone and starchy endosperm of wheat. *Plant Biotechnol. J.* **2012**, *10*, 668–679. [[CrossRef](#)]
67. Kasirajan, L.; Hoang, N.V.; Furtado, A.; Botha, F.C.; Henry, R.J. Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. *Sci. Rep.* **2018**, *8*, 11612. [[CrossRef](#)]
68. Nirmal, R.C.; Furtado, A.; Wrigley, C.; Henry, R.J. Influence of Gene Expression on Hardness in Wheat. *PLoS ONE* **2016**, *11*, e0164746. [[CrossRef](#)]
69. Rangan, P.; Furtado, A.; Henry, R. Differential response of wheat genotypes to heat stress during grain filling. *Exp. Agric.* **2019**, *55*, 818–827. [[CrossRef](#)]
70. Rangan, P.; Furtado, A.; Henry, R.J. The transcriptome of the developing grain: A resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat. *BMC Genom.* **2017**, *18*, 766. [[CrossRef](#)]
71. Cuperus, J.T. Single-cell genomics in plants: Current state, future directions, and hurdles to overcome. *Plant Physiol.* **2021**, *188*, 749–755. [[CrossRef](#)] [[PubMed](#)]
72. Hoang, N.V.; Furtado, A.; Thirugnanasambandam, P.P.; Botha, F.C.; Henry, R.J. De novo assembly and characterizing of the culm-derived meta-transcriptome from the polyploid sugarcane genome based on coding transcripts. *Heliyon* **2018**, *4*, e00583. [[CrossRef](#)] [[PubMed](#)]
73. Tirumala, R.K.; Subbaiyan, G.K.; Singh, A.K.; Furtado, A.; Henry, R.J. RNA-Seq to Understand Transcriptomes and Application in Improving Crop Quality. In *Comprehensive Foodomics*, 1st ed.; Cifuentes, A., Ed.; Elsevier: Amsterdam, The Netherlands, 2021; Volume 1, pp. 472–485. [[CrossRef](#)]
74. Margarido, G.R.A.; Correr, F.H.; Furtado, A.; Botha, F.C.; Henry, R.J. Limited allele-specific gene expression in highly polyploid sugarcane. *Genome Res.* **2022**, *32*, 297–308. [[CrossRef](#)] [[PubMed](#)]
75. Powell, J.J.; Fitzgerald, T.L.; Stiller, J.; Berkman, P.J.; Gardiner, D.M.; Manners, J.M.; Henry, R.J.; Kazan, K. The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. *Plant Biotechnol. J.* **2016**, *15*, 533–543. [[CrossRef](#)] [[PubMed](#)]
76. Cheng, B.; Furtado, A.; Henry, R.J. Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *GigaScience* **2017**, *6*, gix086. [[CrossRef](#)] [[PubMed](#)]
77. Zhou, S.; Zhang, J.; Han, H.; Zhang, J.; Ma, H.; Zhang, Z.; Lu, Y.; Liu, W.; Yang, X.; Li, X.; et al. Full-length transcriptome sequences of *Agropyron cristatum* facilitate the prediction of putative genes for thousand-grain weight in a wheat-A. *cristatum* translocation line. *BMC Genom.* **2019**, *20*, 1025. [[CrossRef](#)] [[PubMed](#)]
78. Yuan, H.Z.; Yu, H.M.; Huang, T.; Shen, X.J.; Xia, J.; Pang, F.H.; Wang, J.; Zhao, M.Z. The Complexity of The *Fragaria X Ananassa* (Octoploid) Transcriptome by Single-Molecule Long-Read Sequencing. *Hortic. Res.* **2019**, *6*, 46. [[CrossRef](#)] [[PubMed](#)]
79. Hoang, N.V.; Furtado, A.; Mason, P.J.; Marquardt, A.; Kasirajan, L.; Thirugnanasambandam, P.P.; Botha, F.C.; Henry, R.J. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genom.* **2017**, *18*, 395. [[CrossRef](#)]
80. Hoang, N.V.; Furtado, A.; McQualter, R.B.; Henry, R.J. Next generation sequencing of total DNA from sugarcane provides no evidence for chloroplast heteroplasmy. *New Negat. Plant Sci.* **2015**, *1–2*, 33–45. [[CrossRef](#)]
81. Dobrogojski, J.; Adamiec, M.; Luciński, R. The chloroplast genome: A review. *Acta Physiol. Plant.* **2020**, *42*, 98. [[CrossRef](#)]
82. Henry, R.; Rice, N.; Edwards, M.; Nock, C. Next-Generation Technologies to Determine Plastid Genome Sequences. In *Chloroplast Biotechnology*; Maliga, P., Ed.; Humana Press: Totova, NJ, USA, 2014; Volume 1132, pp. 39–46.
83. Nock, C.J.; Waters, D.L.; Edwards, M.A.; Bowen, S.G.; Rice, N.; Cordeiro, G.M.; Henry, R.J. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* **2010**, *9*, 328–333. [[CrossRef](#)] [[PubMed](#)]
84. Guyeux, C.; Charr, J.-C.; Tran, H.T.M.; Furtado, A.; Henry, R.J.; Cruzillat, D.; Guyot, R.; Hamon, P. Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. *PLoS ONE* **2019**, *14*, e0216347. [[CrossRef](#)] [[PubMed](#)]
85. Ananda, G.; Norton, S.; Blomstedt, C.; Furtado, A.; Møller, B.; Gleadow, R.; Henry, R. Phylogenetic relationships in the *Sorghum* genus based on sequencing of the chloroplast and nuclear genes. *Plant Genome* **2021**, *14*, e20123. [[CrossRef](#)] [[PubMed](#)]
86. Brozynska, M.; Copetti, D.; Furtado, A.; Wing, R.A.; Crayn, D.; Fox, G.; Ishikawa, R.; Henry, R.J. Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. *Plant Biotechnol. J.* **2016**, *15*, 765–774. [[CrossRef](#)]
87. Healey, A.; Lee, D.J.; Furtado, A.; Henry, R.J. Evidence of inter-sectional chloroplast capture in *Corymbia* among sections *Torellianae* and *Maculatae*. *Aust. J. Bot.* **2018**, *66*, 369. [[CrossRef](#)]
88. Hodel, R.G.J.; Zimmer, E.A.; Liu, B.-B.; Wen, J. Synthesis of Nuclear and Chloroplast Data Combined With Network Analyses Supports the Polyploid Origin of the Apple Tribe and the Hybrid Origin of the Maleae—Gillenieae Clade. *Front. Plant Sci.* **2022**, *12*, 820997. [[CrossRef](#)] [[PubMed](#)]
89. Yoo, K.; Lowry, P.P.; Wen, J. Discordance of chloroplast and nuclear ribosomal DNA data in *Osmorhiza* (Apiaceae). *Am. J. Bot.* **2002**, *89*, 966–971. [[CrossRef](#)]
90. Yu, W.-B.; Huang, P.-H.; Li, D.-Z.; Wang, H. Incongruence between Nuclear and Chloroplast DNA Phylogenies in *Pedicularis* Section *Cyathophora* (Orobanchaceae). *PLoS ONE* **2013**, *8*, e74828. [[CrossRef](#)]
91. Gualberto, J.M.; Mileshina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **2014**, *100*, 107–120. [[CrossRef](#)]

92. Qin, L.; Hu, Y.; Wang, J.; Wang, X.; Zhao, R.; Shan, H.; Li, K.; Xu, P.; Wu, H.; Yan, X.; et al. Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. *Nat. Plants* **2021**, *7*, 1239–1253. [[CrossRef](#)]
93. Denoed, F.; Carretero-Paulet, L.; Dereeper, A.; Droc, G.; Guyot, R.; Pietrella, M.; Zheng, C.; Alberti, A.; Anthony, F.; Aprea, G.; et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **2014**, *345*, 1181–1184. [[CrossRef](#)] [[PubMed](#)]
94. Julca, I.; Ferrari, C.; Flores-Tornero, M.; Proost, S.; Lindner, A.-C.; Hackenberg, D.; Steinbachová, L.; Michaelidis, C.; Gomes Pereira, S.; Misra, C.S.; et al. Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nat. Plants* **2021**, *7*, 1143–1159. [[CrossRef](#)] [[PubMed](#)]
95. Stein, J.C.; Yu, Y.; Copetti, D.; Zwickl, D.J.; Zhang, L.; Zhang, C.; Chougule, K.; Gao, N.; Iwata, A.; Goicoechea, J.L.; et al. Publisher Correction: Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* **2018**, *50*, 285–296. [[CrossRef](#)] [[PubMed](#)]
96. Park, H.-S.; Lee, W.K.; Lee, S.-C.; Lee, H.O.; Joh, H.J.; Park, J.Y.; Kim, S.; Song, K.; Yang, T.-J. Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F1 hybrid combinations. *Sci. Rep.* **2021**, *11*, 2506. [[CrossRef](#)] [[PubMed](#)]
97. Henry, R.J. Genomics strategies for germplasm characterization and the development of climate resilient crops. *Front. Plant Sci.* **2014**, *5*, 68. [[CrossRef](#)]
98. Bradbury, L.M.T.; Fitzgerald, T.L.; Henry, R.J.; Jin, Q.; Waters, D.L.E. The gene for fragrance in rice. *Plant Biotechnol. J.* **2005**, *3*, 363–370. [[CrossRef](#)]
99. Bradbury, L.M.T.; Henry, R.J.; Jin, Q.; Reinke, R.F.; Waters, D. A Perfect Marker for Fragrance Genotyping in Rice. *Mol. Breed.* **2005**, *16*, 279–283. [[CrossRef](#)]
100. Waters, D.L.E.; Henry, R.J.; Reinke, R.F.; Fitzgerald, M.A. Gelatinization temperature of rice explained by polymorphisms in starch synthase. *Plant Biotechnol. J.* **2006**, *4*, 115–122. [[CrossRef](#)]
101. Nirmal, R.C.; Furtado, A.; Rangan, P.; Henry, R.J. Fasciclin-like arabinogalactan protein gene expression is associated with yield of flour in the milling of wheat. *Sci. Rep.* **2017**, *7*, 12539. [[CrossRef](#)]
102. Furtado, A.; Bundock, P.; Banks, P.M.; Fox, G.; Yin, X.; Henry, R.J. A novel highly differentially expressed gene in wheat endosperm associated with bread quality. *Sci. Rep.* **2015**, *5*, 10446. [[CrossRef](#)]
103. Henry, R.J. *Plant Genotyping: The DNA Fingerprinting of Plants*; CABI Publishing: Oxford, UK, 2001; p. 325.
104. Henry, R.J.; Edwards, M.; Waters, D.L.E.; Krishnan, S.G.; Bundock, P.; Sexton, T.R.; Kharabian-Masouleh, A.; Nock, C.; Pattemore, J. Application of large-scale sequencing to marker discovery in plants. *J. Biosci.* **2012**, *37*, 829–841. [[CrossRef](#)] [[PubMed](#)]
105. Krishna, H.; Alizadeh, M.; Singh, D.; Singh, U.; Chauhan, N.; Eftekhari, M.; Sadh, R.K. Somaclonal variations and their applications in horticultural crops improvement. *3 Biotech* **2016**, *6*, 54. [[CrossRef](#)] [[PubMed](#)]
106. Henry, R.J. Genomics and Gene-Editing Technologies Accelerating Grain Product Innovation. *Cereal Foods World* **2019**, *64*. [[CrossRef](#)]
107. Henry, R.J. Advances in DNA sequencing enabling more rapid development of improved biomass and biofuel conversion technologies. *Biofuels* **2012**, *3*, 507–509. [[CrossRef](#)]
108. Sherman, B.; Henry, R.J. Access to biodiversity for food production: Reconciling open access digital sequence information with access and benefit sharing. *Mol. Plant* **2021**, *14*, 701–704. [[CrossRef](#)]
109. Sherman, B.; Henry, R.J. The Nagoya Protocol and historical collections of plants. *Nat. Plants* **2020**, *6*, 430–432. [[CrossRef](#)]
110. Henry, R. Innovations in Agriculture and Food Supply in Response to the COVID-19 Pandemic. *Mol. Plant* **2020**, *13*, 1095–1097. [[CrossRef](#)]
111. Henry, R.J. Innovations in plant genetics adapting agriculture to climate change. *Curr. Opin. Plant Biol.* **2019**, *56*, 168–173. [[CrossRef](#)]
112. Abberton, M.; Batley, J.; Bentley, A.; Bryant, J.; Cai, H.; Cockram, J.; De Oliveira, A.C.; Cseke, L.J.; Dempewolf, H.; De Pace, C.; et al. Global agricultural intensification during climate change: A role for genomics. *Plant Biotechnol. J.* **2015**, *14*, 1095–1098. [[CrossRef](#)]
113. Henry, R.J. Sequencing of wild crop relatives to support the conservation and utilization of plant genetic resources. *Plant Genet. Resour.* **2014**, *12*, S9–S11. [[CrossRef](#)]
114. McCouch, S.; Navabi, Z.K.; Abberton, M.; Anglin, N.L.; Barbieri, R.L.; Baum, M.; Bett, K.; Booker, H.; Brown, G.L.; Bryan, G.J.; et al. Mobilizing Crop Biodiversity. *Mol. Plant* **2020**, *13*, 1341–1344. [[CrossRef](#)] [[PubMed](#)]