Current status of whole-genome sequences of Korean angiosperms

Jongsun PARK*, Yunho YUN, Hong XI, Woochan KWON and Janghyuk SON

Infoboss Inc. and InfoBoss Research Center, Seoul 06278, Korea

Corresponding author
Jongsun PARK
E-mail: starflr@infoboss.co.kr

ABSTRACT: Owing to the rapid development of sequencing technologies, more than 1,000 plant genomes have been sequenced and released. Among them, 69 Korean plant taxa (85 genome sequences) contain at least one whole-genome sequence despite the fact that some samples were not collected in Korea. The sequencing-by-synthesis method (next-generation sequencing) and the PacBio (third-generation sequencing) method were the most commonly used in studies appearing in 65 publications. Several scaffolding methods, such as the Hi-C and 10x types, have also been used for pseudo-chromosomal assembly. The most abundant families among the 69 taxa are Rosaceae (10 taxa), Brassicaceae (7 taxa), Fabaceae (7 taxa), and Poaceae (7 taxa). Due to the rapid release of plant genomes, it is necessary to assemble the current understanding of Korean plant species not only to understand their whole genomes as our own plant resources but also to establish new tools for utilizing plant resources efficiently with various analysis pipelines, including AI-based engines.

INTRODUCTION

A genome sequence consisting of the complete list of nucleotides contains all of the genetic information of an organism (Alberts et al., 2002). Once the genome is deciphered properly, theoretically we may find all information pertaining to an organism, thus providing strong motivation for several early genome projects (Liggett, 2001), including those focusing on *Arabidopsis thaliana* L. (Martienssen and McCombie 2001; Mitchell-Olds, 2001). In 2000, the first plant whole-genome sequence, five chromosomes of *A. thaliana* L., was fully assembled and published (The Arabidopsis Genome Initiative, 2000), marking a relatively late start in comparison to those of bacteriophages (Sanger et al., 1977), bacteria (Fleischmann et al., 1995), and fungi (Goffeau et al., 1996). At that time, the *A. thaliana* genome had 25,498 genes (The Arabidopsis Genome Initiative, 2000), considered to be the most important element to understanding plants (Hutchison et al., 2016). Over time, researchers realized that additional elements, such as non-coding RNAs, also played important roles (Jones-Rhoades et al., 2006; Song et al., 2019) as other aspects by which we could understand plant whole-genome sequences.

Whole-genome sequences have been assembled from small pieces of DNA from organisms due to two major limitations of sequencing: (1) the start position of sequencing cannot be determined in a whole-genome project, and (2) the length of sequences we can obtain is limited, e.g., 500 bp to 1,000 bp when using the Sanger sequencing method. To decipher whole-genome sequences with sufficiently long lengths (e.g., *A. thaliana* is 119 Mbp (The Arabidopsis Genome Initiative, 2000) and *Zea mays* L. is 2.3 Gbp (Schnable et al., 2009)), an approach known as the whole-shotgun strategy was adopted after a method based on bacterial artificial chromosome (BAC) sequences (Zhang and Wu, 2001) given the large-scale computing power for genome assembly (Weber and Myers, 1997). After the commercialization of next-generation sequencing (NGS) technologies, which have critically reduced sequencing costs, to promote whole-genome sequencing projects, many plant genome sequences were released.

The National Center for Biotechnology Information (NCBI) has served as a data depository for many plant genome sequences (e.g., *Cymbidium goeringii* (Rchb.f.) Rchb.f. (Chung et al., 2021)). A centralized database in which to archive plant genome sequences is useful to provide access

for comparative genomic analyses. The Plant Genome Database (https://www.plantgenome.info/) is such an example of a centralized database for plant whole-genome sequences (Park et al., 2021a), containing 3,509 genomes originating from 1,431 species (Release 3.0 to be available in March 2023). These genomic resources are important to conduct further various comparative genomic analyses to understand the genomes.

The accumulation of genome data for eukaryotes on a global scale has already begun (Lewin et al., 2018). Synthesizing the current status of whole-genome data on a local scale is necessary to set the stage for future research to ascertain the biodiversity and evolution of Korean plants and to discern applications of genomic resources ultimately to conserve endangered species. In this review, we examine the historical background of genome sequencing for plants, emphasizing the development of various sequencing technologies and the current status of whole genomes for angiosperms in Korea.

HISTORY OF SEQUENCING TECHNOLOGIES UTILIZED FOR DECIPHERING PLANT GENOMES

Sequencing technologies have been improved based on the requirements of whole-genome projects. First, it is now possible to increase the number of reads when conducting a whole-genome assembly given the greater computing power. As an example, the human genome project initiated by Craig Venter ended with the generation of 14.8 billion base-pairs (~5× coverage) for *de novo* assembly (Venter et al., 2001). Second, improved *de novo* assembly results are now possible by increasing the length of NGS raw reads. The first advance triggered the development of NGS technologies, including pyrosequencing (Fakhrai-Rad et al., 2002), sequencing by synthesis (Fuller et al., 2009), and sequencing by ligation (Smith et al., 2010). These technologies produced a great many reads in comparison to the Sanger sequencing method, though the reads are shorter than those generated by the Sanger sequencing method, apart from those that use pyrosequencing technology. After several plant genome sequences, including *A. thaliana* (The Arabidopsis Genome Initiative, 2000), *Populus trichocarpa* Torr. & A. Gray ex. Hook. (Tuskan et al., 2006), *Vitis vinifera* L. (Jaillon et al., 2007), and *Z. mays* (Schnable et al., 2009), were deciphered by the Sanger sequencing method, the cucumber (*Cucumis sativus* L.) genome was successfully assembled mainly based on sequencing by a synthesis method (commonly known as the Illumina method) (Huang et al., 2009). Although the NGS read length is relatively short, it has been shown that plant genome sequences can be assembled based on early short-read sequences (36-bp reads) generated by NGS technology.

On the other hand, the 22-Gbp plant genome of *Pinus taeda* L. was successfully assembled (Zimin et al., 2014) based on pyrosequencing, implying that the *de novo* assembly of very large plant genomes is also possible with read lengths similar to those by the Sanger sequencing method.

As more plant genome sequences continue to be sequenced based on NGS technologies at a much lower cost in comparison to the Sanger sequencing method, the short length has become a critical hurdle hindering the realization of high-

---

Terminology 1

Next-generation sequencing (NGS) technologies: NGS technologies refer to major sequencing technologies which have been developed to overcoming the limitations of Sanger sequencing technologies addressed in the late 1990s. The first commercialized NGS technology is a pyrosequencing method known as the 454 technology method initially. In the early phase, it generated less than one million 150-bp reads but eventually could generate several million reads of which the length exceeds 1 kb (GS-Flx). This technology was officially ended in 2017. The sequencing-by-synthesis (SBS) technology was commercialized in 2007 with several million 36-bp reads per lane. This approach provided low-cost data, but each read was too short in comparison to those by the pyrosequencing method. Human genomes were successfully re-sequenced with data from this method in 2008 and 2009 (Bentley et al., 2008; Wang et al., 2008; Kim et al., 2009) and the cucumber genome was successfully assembled *de novo* with short-read data together with reads by the Sanger sequencing method (Huang et al., 2009). Currently this method provides a hundred million 151-bp reads per lane with the latest version of SBS technology (NovaSeq-6000). Other technologies such as sequencing-by-ligation developed by Applied Bioscience have also been commercialized, known as SOLiD (Miles et al., 2013), but disappeared after several years due to failures to address quality control issues.

Third-generation sequencing (TGS) technologies: TGS technologies were developed to meet long-read requirements (more than that by the Sanger sequencing method) in a high-throughput manner (similar to NGS technologies). From the long period of stabilization of TGS technologies, the single-molecular real-time (SMRT) sequencing method by Pacific Biosciences (PacBio) and the method developed by Oxford Nanopore Technology (ONT) have been successfully commercialized, providing a large number of long reads (>10 kb). One weak point of TGS is its low base-pair accuracy; HiFi reads provided by PacBio increased this accuracy so that additional polishing is not required now (Hon et al., 2020); while Nanopore (ONT) still requires a polishing process based on NGS sequences (Amarasinghe et al., 2020) but provides much longer reads than that by PacBio.

Table 1. List of plant whole genome sequences of Korean angiosperms.

No.	Order	Family	Species	Cultivar/Strain	Method	Genome size (bp)	No. of scaffolds	N50 (bp)	GC ratio (%)	No. of genes	Reference
1	Alismatales	Araceae	*Lemna minor*	-	HiSeq MiSeq	763,364,415	6,868	335,126	44.81	N/A	Van Hoeck et al. (2015)
2	Alismatales	Araceae	*Spirodela polyrhiza*	7498	454 BAC-end	132,009,443	207	5,765,642	42.45	N/A	Wang et al. (2014)
3	Alismatales	Araceae	*Spirodela polyrhiza*	9509	HiSeq Irys	136,591,703	20	7,641,483	42.30	N/A	Michael et al. (2017)
4	Alismatales	Araceae	*Spirodela polyrhiza*	Sp9504	N/A	137,175,874	16,051	14,533	42.21	N/A	Unpublished
5	Alismatales	Zosteraceae	*Zostera marina*	-	HiSeq	203,914,448	2,228	485,578	38.34	20,450	Olsen et al. (2016)
6	Alismatales	Zosteraceae	*Zoysia japonica*	Nagirizaki	HiSeq	334,384,427	11,786	2,370,062	44.08	N/A	Tanaka et al. (2016)
7	Apiales	Apiaceae	*Centella asiatica*	BB-174	10x Hi-C	430,216,894	8,739	50,798,654	34.17	N/A	Pootakham et al. (2021)
8	Apiales	Apiaceae	*Oenanthe javanica*	-	HiSeq	1,650,714,995	150,335	32,514	35.44	N/A	Liu et al. (2021a)
9	Apiales	Araliaceae	*Panax ginseng*	-	HiSeq	3,414,349,854	83,074	108,708	34.94	N/A	Xu et al. (2017)
10	Apiales	Araliaceae	*Panax ginseng*	-	HiSeq	2,984,993,682	9,845	3,641,815	34.84	59,352	Kim et al. (2018)
11	Asparagales	Orchidaceae	*Gastrodia elata*	EJP_2020-GE1012	HiSeq PacBio Hi-C	1,046,143,939	514	50,595,616	34.27	N/A	Bae et al. (2022)
12	Asparagales	Orchidaceae	*Gastrodia elata* f. *glauca*	-	HiSeq	1,060,984,162	3,768	4,911,943	34.52	N/A	Yuan et al. (2018)
13	Asparagales	Orchidaceae	*Cymbidium goeringii*	-	HiSeq PacBio Hi-C	3,990,519,457	19,377	178,198,413	33.77	29,556	Chung et al. (2021)
14	Asterales	Asteraceae	*Artemisia annua*	-	HiSeq 454 PacBio	1,792,856,094	39,400	104,891	35.34	N/A	Shen et al. (2018)
15	Asterales	Campanulaceae	*Codonopsis lanceolata*	NIHHS 239928	Nanopore HiSeq	1,347,489,827	22,630	82,893	37.21	N/A	Jang et al. (2023)
16	Asterales	Campanulaceae	*Codonopsis pilosula*	-	N/A	937,709,907	1,154	2,556,440	37.13	N/A	Unpublished
17	Boraginales	Boraginaceae	*Lithospermum erythrorhizon*	-	Nanopore HiSeq	366,684,367	2,451	315,765	35.15	32,360	Auber et al. (2020)
18	Brassicales	Brassicaceae	*Arabidopsis lyrata*	MN47	Sanger	206,667,935	695	24,464,547	36.08	32,550	Hu et al. (2011)
19	Brassicales	Brassicaceae	*Arabidopsis thaliana*	Col-0	Sanger	119,145,879	5	23,459,830	36.03	48,113	Kaul et al. (2000)
20	Brassicales	Brassicaceae	*Arabis glabra*	-	N/A	171,129,789	250	5,712,653	36.01	N/A	Unpublished
21	Brassicales	Brassicaceae	*Capsella bursa-pastoris*	-	HiSeq MiSeq	268,430,517	8,186	627,605	35.74	52,528	Kasianov et al. (2017)
22	Brassicales	Brassicaceae	*Erysimum cheiranthoides*	-	HiSeq PacBio Hi-C	177,180,559	223	22,409,365	36.29	N/A	Züst et al. (2020)

Table 1. Continued.

No.	Order	Family	Species	Cultivar/Strain	Method	Genome size (bp)	No. of scaffolds	N50 (bp)	GC ratio (%)	No. of genes	Reference
23	Brassicales	Brassicaceae	*Isatis tinctoria*	-	N/A	756,237,117	37,877	85,498	36.89	116,152	Unpublished
24	Brassicales	Brassicaceae	*Rorippa islandica*	-	N/A	390,742,222	430	3,103,162	35.67	65,406	Unpublished
25	Caryophyllales	Caryophyllaceae	*Silene noctiflora*	OPL-1.1	N/A	2,598,026,552	79,767	59,004	37.95	N/A	Unpublished
26	Cucurbitales	Cucurbitaceae	*Gynostemma pentaphyllum*	JGL-2020	HiSeq BGISeq PacBio Hi-C	582,948,444	578	50,780,587	32.86	N/A	Huang et al. (2021)
27	Ericales	Actinidiaceae	*Actinidia rufa*	Fuchu	HiSeq PacBio	677,239,177	501	15,941,708	35.44	52,342	Tahir et al. (2022)
28	Ericales	Ebenaceae	*Diospyros lotus*	Kunsenshi	N/A	945,758,782	8,975	653,513	36.35	N/A	Unpublished
29	Ericales	Ebenaceae	*Diospyros lotus*	W01	N/A	617,726,390	142	40,720,603	36.52	N/A	Unpublished
30	Ericales	Ebenaceae	*Diospyros lotus*	Yz01	N/A	630,098,508	42	42,671,757	36.57	N/A	Unpublished
31	Fabales	Fabaceae	*Amphicarpaea edgeworthii*	AE-2020	BGISeq Nanopore	299,059,313	24	27,224,559	32.06	N/A	Liu et al. (2021b)
32	Fabales	Fabaceae	*Glycine soja*	F	N/A	975,918,537	320	48,777,390	34.78	57,501	Unpublished
33	Fabales	Fabaceae	*Glycine soja*	USDA-GRIN-PI 483463	HiSeq PacBio Bionano	985,259,865	306	48,820,272	34.94	N/A	Valliyodan et al. (2019)
34	Fabales	Fabaceae	*Glycine soja*	W05	HiSeq PacBio Hi-C	863,568,428	33,170	404,776	33.56	50,337	Xie et al. (2019)
35	Fabales	Fabaceae	*Medicago ruthenica*	Xinghe	HiSeq PacBio Hi-C	904,130,090	650	99,386,155	35.89	N/A	Yin et al. (2021)
36	Fabales	Fabaceae	*Vicia sativa*	KSR5	HiSeq PacBio	1,541,180,487	54,083	90,105	34.96	N/A	Shirasawa et al. (2021b)
37	Fabales	Fabaceae	*Vigna minima*	-	HiSeq PacBio	486,496,346	2,940	25,346,430	34.19	N/A	Naito et al. (2022)
38	Fagales	Betulaceae	*Betula pendula*	-	HiSeq SOLiD	435,914,794	5,644	239,696	35.74	86,599	Salojärvi et al. (2017)
39	Fagales	Betulaceae	*Corylus heterophylla*	CHET03	HiSeq PacBio Hi-C	346,499,040	381	2,025,119	35.99	N/A	Unpublished
40	Fagales	Betulaceae	*Corylus heterophylla*	R026	Nanopore Hi-C	370,750,808	951	31,328,411	35.84	N/A	Zhao et al. (2021)
41	Fagales	Fagaceae	*Castanea crenata*	Ginyose	HiSeq PacBio	721,168,657	781	1,595,543	35.14	N/A	Shirasawa et al. (2021c)
42	Fagales	Fagaceae	*Castanea mollissima*	N11-1	HiSeq PacBio Hi-C	833,240,550	112	57,343,431	35.11	N/A	Wang et al. (2020)
43	Fagales	Fagaceae	*Castanea mollissima*	Vanuxem	454 MiSeq	725,180,808	14,110	101,575	35.06	34,314	Staton et al. (2020)
44	Fagales	Fagaceae	*Castanea mollissima*	Hubei	HiSeq PacBio	785,529,252	2,707	944,461	35.21	36,479	Xing et al. (2019)

Table 1. Continued.

No.	Order	Family	Species	Cultivar/Strain	Method	Genome size (bp)	No. of scaffolds	N50 (bp)	GC ratio (%)	No. of genes	Reference
45	Fagales	Fagaceae	*Quercus mongolica*	Qm_2020SYAU	HiSeq PacBio Hi-C	809,993,317	321	66,735,633	35.84	N/A	Ai et al. (2022)
46	Fagales	Juglandaceae	*Juglans mandshurica*	-	HiSeq	558,070,702	13,809	496,923	36.26	N/A	Bai et al. (2018)
47	Lamiales	Lamiaceae	*Perilla citriodora*	-	N/A	618,797,171	29,924	2,413,193	35.38	155,867	Unpublished
48	Lamiales	Oleaceae	*Fraxinus mandshurica*	-	HiSeq MiSeq 454	830,473,688	297,505	30,144	34.96	N/A	Sollars et al. (2017)
49	Lamiales	Oleaceae	*Fraxinus sieboldiana*	-	HiSeq MiSeq 454	744,777,403	709,960	1,987	36.21	N/A	Sollars et al. (2017)
50	Lamiales	Orobanchaceae	*Phtheirospermum japonicum*	Okayama	N/A	1,226,606,249	10,559	1,109,341	36.03	30,299	Unpublished
51	Malpighiales	Euphorbiaceae	*Euphorbia esula*	-	N/A	1,124,886,465	1,633,094	1,035	34.80	N/A	Unpublished
52	Malpighiales	Salicaceae	*Populus davidiana*	NFU_1	N/A	417,659,603	1,565	562,472	33.56	N/A	Unpublished
53	Malpighiales	Salicaceae	*Populus simonii*	-	HiSeq PacBio	441,407,051	369	19,598,675	33.65	N/A	Wu et al. (2020)
54	Poales	Poaceae	*Brachypodium sylvaticum*	-	N/A	358,283,154	629	38,764,466	46.38	50,263	Unpublished
55	Poales	Poaceae	*Echinochloa crus-galli*	STB08	HiSeq PacBio	1,486,609,408	4,534	1,802,240	45.70	N/A	Guo et al. (2017)
56	Poales	Poaceae	*Eleusine indica*	HZ-2018	HiSeq	492,270,386	24,072	233,459	44.42	N/A	Zhang et al. (2019)
57	Poales	Poaceae	*Miscanthus sacchariflorus*	-	HiSeq	2,074,815,175	105,321	37,711	45.94	N/A	De Vega et al. (2021)
58	Poales	Poaceae	*Miscanthus sinensis*	-	HiSeq Hi-C	2,079,430,866	14,431	88,510,541	45.77	89,486	Mitros et al. (2020)
59	Poales	Poaceae	*Setaria viridis*	A10	HiSeq pacBio	395,731,502	14	46,702,114	46.17	52,459	Mamidi et al. (2020)
60	Poales	Poaceae	*Setaria viridis*	ME034V	HiSeq Nanopore	397,031,521	9	46,382,547	45.99	N/A	Thielen et al. (2020)
61	Poales	Poaceae	*Themeda triandra*	sf006	N/A	889,201,528	3,179	600,859	46.24	N/A	Unpublished
62	Poales	Poaceae	*Zizania latifolia*	HSD2	HiSeq	603,989,347	4,522	604,864	42.72	N/A	Guo et al. (2015)
63	Poales	Typhaceae	*Typha latifolia*	L0001	N/A	214,326,387	81	14,472,745	37.76	N/A	Unpublished
64	Poales	Typhaceae	*Typha latifolia*	SDW-2020	HiSeq PacBio	287,194,721	1,158	8,705,712	38.07	N/A	Widanagama et al. (2022)
65	Rosales	Moraceae	*Ficus erecta*	-	HiSeq PacBio Genetic map	595,834,738	2,455	697,200	34.32	111,921	Shirasawa et al. (2020)

Table 1. Continued.

No.	Order	Family	Species	Cultivar/ Strain	Method	Genome size (bp)	No. of scaffolds	N50 (bp)	GC ratio (%)	No. of genes	Reference
66	Rosales	Rosaceae	Fragaria nipponica	-	454 HiSeq	206,414,979	215,024	1,275	38.44	87,803	Hirakawa et al. (2014)
67	Rosales	Rosaceae	Fragaria orientalis	-	454 HiSeq	214,184,046	323,163	722	38.11	99,674	Hirakawa et al. (2014)
68	Rosales	Rosaceae	Malus baccata	-	HiSeq	703,023,564	47,474	743,357	38.13	45,900	Chen et al. (2019)
69	Rosales	Rosaceae	Prunus yedoensis var. nudiflora*	-	HiSeq PacBio Fosmid-end	319,209,792	4,016	145,140	37.65	41,294	Baek et al. (2018)
70	Rosales	Rosaceae	Prunus yeonesis	-	HiSeq PacBio	690,105,700	4,571	918,183	37.93	N/A	Shirasawa et al. (2019)
71	Rosales	Rosaceae	Prunus davidiana	ST	HiSeq PacBio Hi-C	243,913,910	127	28,110,464	37.55	N/A	Tan et al. (2021)
72	Rosales	Rosaceae	Pyrus pyrifolia	Nijisseiki	BGISeq PacBio	503,888,133	114	7,676,629	37.34	N/A	Shirasawa et al. (2021a)
73	Rosales	Rosaceae	Pyrus pyrifolia	Cuiguan	HiSeq PacBio	541,340,367	428	27,968,327	37.36	42,559	Gao et al. (2021)
74	Rosales	Rosaceae	Rosa lucieae	-	N/A	786,105,075	500,476	10,695	38.73	N/A	Unpublished
75	Rosales	Rosaceae	Rosa multiflora	-	HiSeq MiSeq	739,637,845	83,189	90,830	38.91	67,380	Nakamura et al. (2018)
76	Rosales	Rosaceae	Rosa rugosa	-	PacBio 10x Hi-C	401,223,225	43	61,014,187	39.31	39,704	Chen et al. (2021)
77	Rosales	Rhamnaceae	Ziziphus jujuba	Dongzao	HiSeq BAC	437,753,511	4,789	25,259,912	33.40	37,526	Liu et al. (2014)
78	Rosales	Rhamnaceae	Ziziphus jujuba	Junzao	HiSeq	362,583,438	47	27,412,306	32.95	N/A	Huang et al. (2016)
79	Rosales	Rhamnaceae	Ziziphus jujuba var. spinosa	AT0	HiSeq PacBio Hi-C	405,636,987	538	30,278,369	33.07	24,146	Shen et al. (2021)
80	Rosales	Urticaceae	Boehmeria nivea	-		344,616,830	12,775	1,094,501	35.25	N/A	Unpublished
81	Rosales	Urticaceae	Boehmeria nivea	ZZ1	HiSeq	316,026,324	330	17,838,734	35.20	N/A	Liu et al. (2018)
82	Rosales	Urticaceae	Boehmeria nivea	Zhongsizhu	HiSeq Nanopore PacBio	266,599,076	154,955	48,874	36.94	N/A	Wang et al. (2021a)
83	Rosales	Urticaceae	Boehmeria nivea var. tenacissima	-	HiSeq Nanopore PacBio	270,213,153	135	19,552,154	35.19	N/A	Wang et al. (2021a)
84	Solanales	Convolvulaceae	Cuscuta australis	-	HiSeq PacBio	262,630,465	218	3,625,894	36.40	18,157	Sun et al. (2018)
85	Vitales	Vitaceae	Vitis amurensis	IBCAS1988	HiSeq PacBio Hi-C Bionano	603,559,200	3,040	26,085,413	34.36	N/A	Wang et al. (2021b)

Asterisks indicate Korean endemic species (Chung et al., 2023).
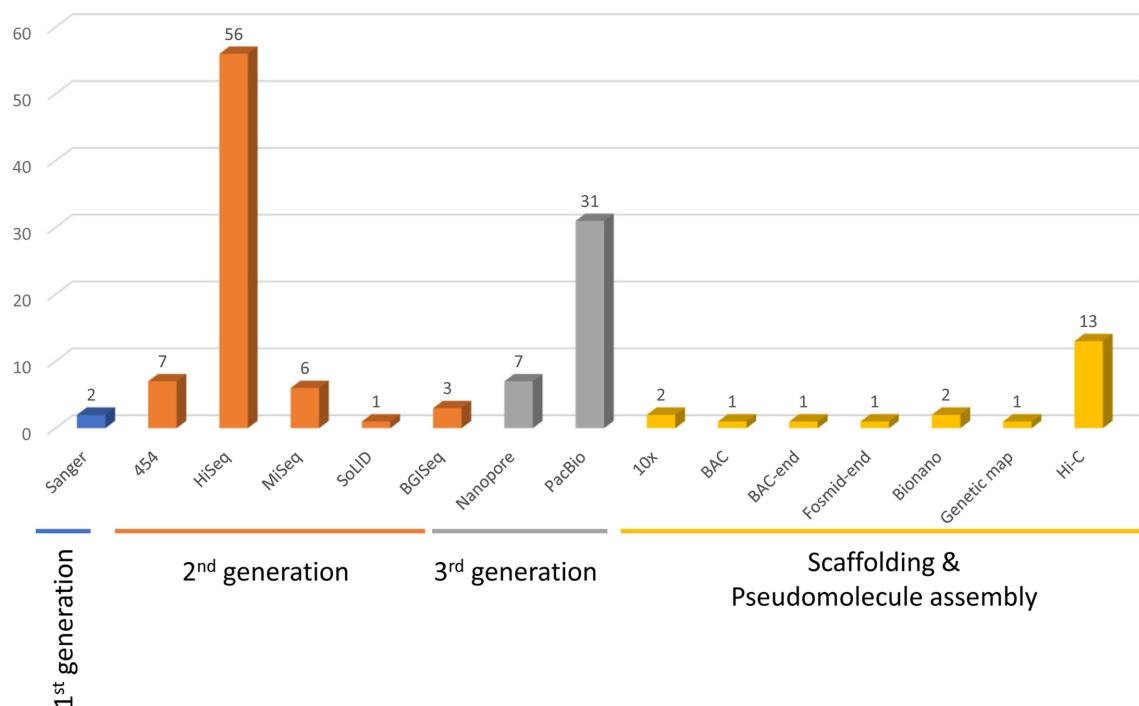N/A, not available.

Fig. 1. Sequencing methods used in whole-genome projects focusing on native and naturalized Korean plant species: the X-axis presents the sequencing methods used in genome projects with four-color tagged legends denoting their classification. The Y-axis displays the number of genomes for each method.

quality genome assemblies. New sequencing technologies after NGS technologies, referred to as third-generation sequencing (TGS) technologies, have been developed, with high-throughput sequencing reads longer than 10 kb, much longer than those by pyrosequencing and even the Sanger sequencing method (Terminology 1). Currently, the average length of the most recent single-molecular real-time (SMRT) sequencing method by Pacific Biosciences (PacBio) is around 10-25 kb with high accuracy (Hon et al., 2020) and the nanopore sequencing method by Oxford Nanopore Technologies (ONT) (referred to as Nanopore hereafter) produces ultra-long reads (>100 kb) (Amarasinghe et al., 2020).

Based on 65 genome publications covering 72 plant genomes (Table 1), the corresponding sequencing methods were investigated (Fig. 1). Most of the investigated genome projects utilized multiple methods, especially for those that adopted TGS methods requiring a base pair polishing process supported by NGS technologies. Only two genomes (2.77%), *A. thaliana* (The Arabidopsis Genome Initiative, 2000) and *Arabidopsis lyrata* (L.) O'Kane & Al-Shehbaz (Hu et al., 2011), were sequenced with only the Sanger sequencing method (Fig. 1, Table 1). Thirteen genomes were sequenced together with pyrosequencing and MiSeq technologies, which provide longer sequences than the typical NGS technologies; however, these methods were selected not due to the large

genome (i.e., *P. taeda* (Zimin et al., 2014)) but because of available optimized sequencing technologies when these projects were conducted. Since the commercialization of TGS technologies, most genomes have been sequenced with PacBio or Nanopore (Fig. 1). Because Nanopore was commercialized later than PacBio, there have been fewer genomes deciphered using this method compared to those by PacBio (Fig. 1).

After assembling whole-genome sequences from raw reads successfully, additional processes for chromosomal-level assembly, including Hi-C (Lieberman-Aiden et al., 2009), 10X Genomics Chromium (termed 10× hereafter) (Weisenfeld et al., 2017), and Bionano (Bocklandt et al., 2019), referred to as the scaffolding method hereafter, were conducted for 20 genomes (23.53%; the *Rosa rugosa* Thunb. genome project used both Hi-C and 10× technologies) (Fig. 1) (Chen et al., 2021). The most frequently used technology is Hi-C (65.00%) (Fig. 1), which has been utilized in a wide range of genomic studies (Kong and Zhang, 2019). The remaining scaffolding methods, including old-fashioned techniques, i.e., the BAC-end (Osoegawa et al., 2001) and Fosmid-end (Williams et al., 2012) sequencing methods, showed to be rarely used (Fig. 1). Based on the ratio between N50 and the total length, recent scaffolding methods including Hi-C, 10×, and Bionano presented better efficiency of assembly with some exceptional cases, i.e., *Erysimum cheiranthoides* L. and *Prunus davidiana*
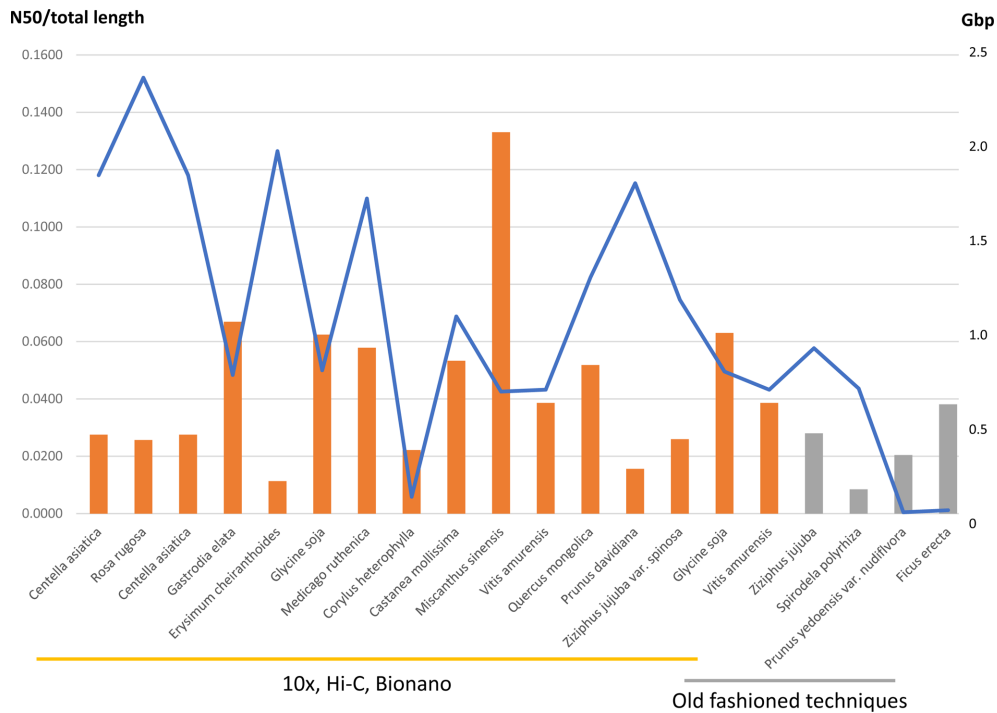
Fig. 2. Genome assembly properties, the ratio of N50 and total lengths and total lengths of genomes using scaffolding technologies the X-axis displays species names in a genome project with two-color tags denoting the classification of the scaffolding technologies. Bars indicate the genome length of each species and the line graph shows the ratio of N50 and the total lengths.

Terminology 2

N50 length: The genome assembly displayed various lengths of contigs or scaffolds due to the limitation of *de novo* assembly, particularly the assembly of repetitive sequences. Hence, the quality of genome assembly can be estimated for the N50 length, defined as the length of the contig or scaffold for which longer contigs or scaffolds explain half of the assembled sequence. This reflects how well long sequences were assembled in comparison to average or even median lengths by focusing on relatively long assembled sequences. This approach is more effective for the assembled sequences based on long reads generated by TGS.

Carrière (Fig. 2). In addition, the ratio does not present a correlation with the total genome length ($R^2 = 0.0397$), strongly suggesting that the N50 lengths of plant genomes depend on the species (Terminology 2).

CURRENT STATUS OF WHOLE GENOMES OF KOREAN ANGIOSPERM SPECIES

There are official lists of native and naturalized Korean plant species, including (1) the Database of National Species List of Korea (Park et al., 2020a), (2) the Standard List of Korean plant species (http://www.nature.go.kr/kpni), and (3) The

Northeastern Asia Biodiversity Institute's List of Korean Vascular Plants. All three lists suggest approximately 4,500 native and naturalized plant species in Korea, also presenting utility as a filter to select published plant genome sequences to investigate the current status of plant genomes of native and naturalized Korean plant species. We used the Northeastern Asia Biodiversity Institute's List of Korean Vascular Plants, selecting 85 plant genomes originating from 69 taxa as Korean angiosperm species genomes from the Plant Genome Database (https://www.plantgenome.info/), which contains available plant whole genome-sequences (Park et al., 2021a) (Table 1). The difference between the numbers of plant genome sequences and the taxa indicates that multiple genomes of some plant taxa were sequenced, making this a good resource to understand intraspecific variations at the level of the genome. We excluded more than 1,700 *A. thaliana* genomes (Ossowski et al., 2008; Ashelford et al., 2011; Cao et al., 2011; Gan et al., 2011; Long et al., 2013; Schmitz et al., 2013; The 1001 Genomes Consortium, 2016; Zou et al., 2017) due to the extremely large number of genomes in this case.

Despite the time gap between the releases of genome sequences and the corresponding publication dates, only 15.29% (13 out of 85 genomes) of genomes do not have any publication. Moreover, five species, *Spirodela polyrhiza* (L.)
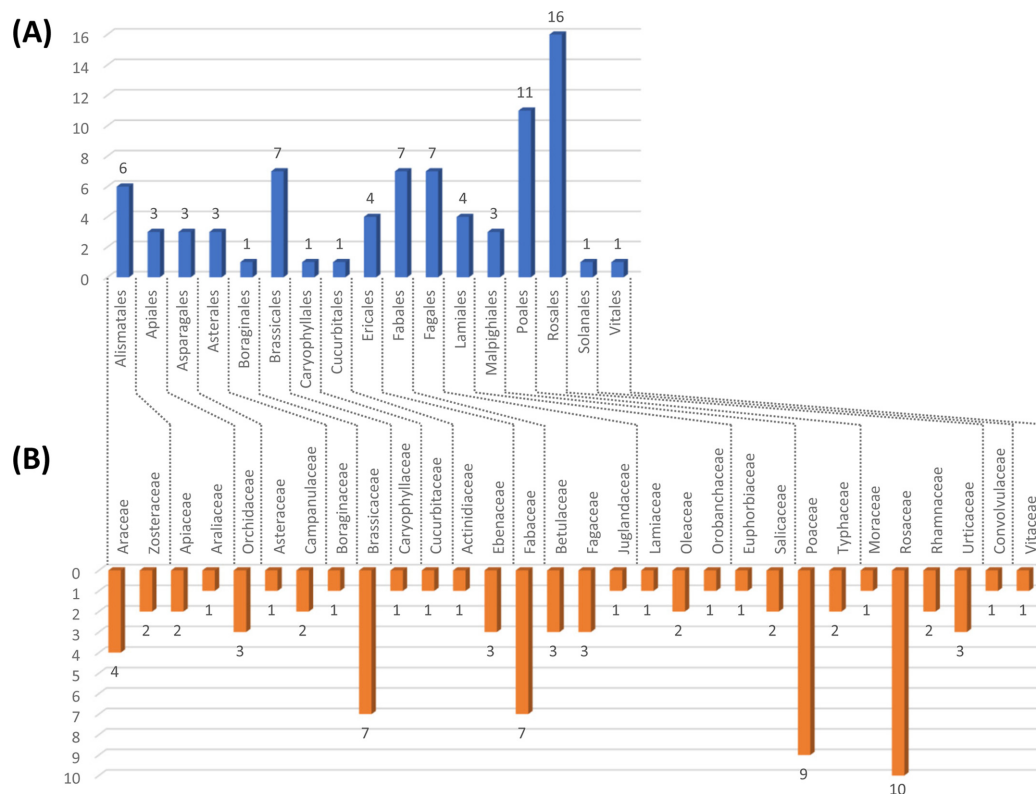
Fig. 3. Taxonomical distribution of 69 native and naturalized Korean plant taxa containing whole-genome sequences: (A) graph displaying the order-level distribution, and (B) graph showing the family-level distribution. Dotted lines represent the relationship between families and orders.

Schleid., *Diospyros lotus* L., *Glycine soja* Sieb. and Zucc., *Castanea mollissima* Blume, and *Boehmeria nivea* (L.) Gaudich., contained three genomes while six species, *Panax ginseng* C. A. Mey., *Corylus heterophylla* Fisch. ex Trautv., *Setaria viridis* (L.) P. Beauv., *Typha latifolia* L., *Pyrus pyrifolia* (Burm. f.) Nakai., and *Ziziphus jujuba* Mill., covered two genomes (Table 1), making them good candidates for understanding intraspecific variations at the genome-wide level. Remarkably, only two genome publications with accessible genome sequences, *Gastrodia elata* Blume and *Codonopsis lanceolata* (Siebold. & Zucc.) Trautv., described the whole genome of Korean angiosperm species (Table 1), indicating that most genomes filtered by the Korean angiosperm species list described in this paper were sequenced based on samples outside of Korea. As the importance of biological resources increases, as exemplified in the Nagoya Protocol (Buck and Hamilton, 2011), additional efforts to obtain whole-genome sequences of these species using the samples isolated in Korea and to prepare assembled sequences (i.e., unfortunately, while the *G. soja* genome was sequenced in Korea (Kim et al., 2010), no assembled sequence is available.) are required to maximize the benefits of the low sequencing costs and feasible bioinformatic analyses.

Sixty-nine taxa were classified into 17 orders and 30 families. Rosales covering four families (Moraceae, Rosaceae, Rhamnaceae, and Utricaceae) contained the largest number of genomes, and Poales covered the second largest from Poaceae to Typhaceae, while Boraginales, Caryophyllales, Cucurbitales, Solanales, and Vitales had only one genome (Fig. 3). At the family level, Rosaceae (10 taxa), Brassicaceae (7 taxa), Fabaceae (7 taxa), and Poaceae (7 taxa) contained a large number of genomes among the 30 families (Fig. 3). This taxonomical bias was expected due to various technical factors associated with genome sequencing projects, including the genome size and ploidy. Most taxa are economically important species. Based on current sequencing technologies, especially TGS technologies, many of these problems have already been overcome; e.g., the hexaploid large genome of wheat (*Triticum aestivum* L.) was sequenced and successfully assembled at the pseudo-chromosomal level (International Wheat Genome Sequencing Consortium, 2018). Rosaceae contained ten genomes originating from five genera, specifically *Prunus*, *Rosa*, *Fragaria*, *Malus*, and *Pyrus* and Poaceae, had nine genomes from the six genera, displaying large taxonomic coverage (Fig. 3, Table 1). These genomes are good candidates for understanding the genomic features of the aforementioned families.
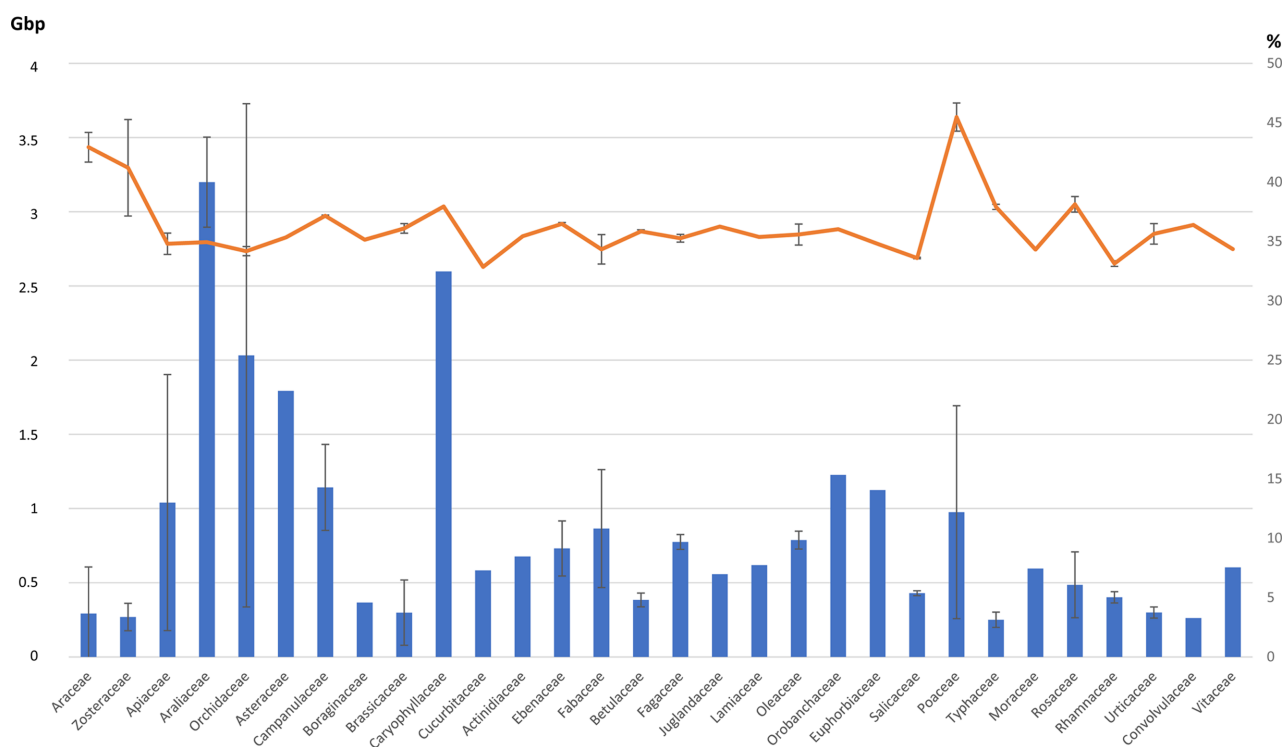
Fig. 4. Family-level genomic properties, total length and GC ratio: the X-axis shows families containing at least one whole genome used in this paper. Bars show the average genome length with the error bars of the standard deviation. The line represents the average GC ratio of genomes with the error bars of the standard deviation.

GENOMIC STATISTICS OF 85 NATIVE AND NATURALIZED KOREAN PLANT GENOMES

The genome length and GC ratio of 85 native and naturalized Korean plant genomes were investigated, indicating average genome lengths of Araliaceae, Caryophyllaceae, and Orchidaceae of 3.20, 2.60, and 2.03 Gbp, respectively; the standard deviation of the Orchidaceae genome length was largest at 1.70 Gbp, while that of the Salicaceae genome is smallest at 16.79 Mbp (Fig. 4). Three genomes from Orchidaceae showed that two of the three genomes were around 1 Gbp; while one was 3 Gbp in length, with Salicaceae covering two genomes from the same genus, *Populus*. The trend of genome size variations along with the families is congruent with the findings of earlier work (Šmarda et al., 2014). In addition, the standard deviation of the GC ratio in Zosteraceae is extremely large, 4.05%, despite the fact that the two genomes are from the same genus, *Zostera* (Table 1). Araceae, Fabaceae, and Poaceae displayed that the standard deviation of the GC ratio ranges from 1.18 to 1.25% (Fig. 4), which is a large value in comparison to the other families. These variations in GC ratio are similar to those in a previous study that investigated variations of the GC ratio of the whole genome in a monocot species (Šmarda et al., 2014). The

number of genomes in each family is small to present a corresponding trend, therefore they provide a glimpse of family-specific genomic features through these numbers as an indicator. This indicates that further genomic studies of Korean plants are necessary.

UTILIZATION OF KOREAN ANGIOSPERM GENOME SEQUENCES

More than 1,000 plant whole genomes have been sequenced, meaning that we can investigate their characteristics in detail. However, grants and proper human resources are still required for us to present the potential or direct economic value of this resource. Genomics-assisted breeding, a good example of the utilization of whole genomes, can be conducted based on genome-wide association studies to target useful phenotypes for breeding (Ahmar et al., 2020, 2021). To shorten the breeding time and increase the efficiency of the process, genomic editing is a viable upcoming strategy, with modifications to genetic elements to achieve helpful characteristics for humans using CRISPR-associated protein 9 (Lee et al., 2019). Legal regulations pertaining to genetically modified organisms, considered a main suppression factor, have become positive, accepting genome-edited plants in comparison to genetically

modified organisms (Sprink et al., 2022). This development will promote the potential usage of plant whole-genome sequences because the techniques mentioned above commonly require whole genomes.

Native plant resources have been utilized to develop a range of useful products, including medicines over many years due to effective compounds such as aspirin, an acetyl salicylic acid from willow bark (Norn et al., 2009), and paclitaxel extracted from the bark of *Taxus brevifolia* Nutt. (Bose et al., 2020). In addition, various natural product medicines have also been developed based on plant extracts (Ngo et al., 2013) as well as traditional medicines, which have been utilized for several thousand years (Ansari and Inamdar, 2010), reflecting the commercial usage of plant resources. Whole-genome sequences can be analyzed to predict useful phytocompounds because they contain all enzymes involved in biochemical synthesis theoretically (Kang et al., 2020; Park et al., 2020b), suggesting another useful feature of Korean native plant genomes.

FURTHER DIRECTIONS FOR KOREAN ANGIOSPERM GENOME SEQUENCES

Currently, more than a hundred plant whole genomes have been sequenced per year, a rate much faster than that of ten years ago, when NGS technologies were merely utilized for the *de novo* assembly of plant genomes. Hence, more Korean angiosperm genomes will also be available from our own genome projects and genome projects conducted outside of Korea. We can consider two main strategies for a database for archive and utilization of Korean angiosperm genomes. First, to expand the coverage of Korean angiosperm genomes, additional Korean angiosperm genomes, especially endemic species which can be utilized commercially or which may be valuable in research, can be sequenced. Second, additional individuals or populations of Korean angiosperm species can be sequenced to understand genome-wide intraspecific variations (Slavov et al., 2012; The 3,000 Rice Genomes Project, 2014; Gulyaev et al., 2022) as well as functional gene families (Kim et al., 2021a, 2021b), as useful phenotypes of plant resources show differences in an intraspecific manner (Moore et al., 2014; Aspinwall et al., 2015; Ren et al., 2020). In addition, all of these genomes can be managed under the environment of a standardized integrated platform to analyze them further, such as the web-based genomic analysis platform Galaxy (Giardine et al., 2005; Blankenberg et al., 2010) or the Genome Information System (GeIS; https://geis.infoboss.co.kr/), which have been utilized in various genomics studies (Lee et al., 2020; Park et al., 2020c, 2021b).

ORCID: Jongsun PARK http://orcid.org/0000-0003-0786-4701; Yunho YUN https://orcid.org/0000-0002-2250-386X; Hong XI https://orcid.org/0000-0003-0731-2210; Woochan KWON https://orcid.org/0000-0001-8665-8097; Janghyuk SON http://orcid.org/0000-0003-3525-8274

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

LITERATURE CITED

Ahmar, S., P. Ballesta, M. Ali and F. Mora-Poblete. 2021. Achievements and challenges of genomics-assisted breeding in forest trees: From marker-assisted selection to genome editing. International Journal of Molecular Sciences 22: 10583.

Ahmar, S., R. A. Gill, K.-H. Jung, A. Faheem, M. U. Qasim, M. Mubeen and W. Zhou. 2020. Conventional and molecular techniques from simple breeding to speed breeding in crop plants: Recent advances and future outlook. International Journal of Molecular Sciences 21: 2590.

Ai, W., Y. Liu, M. Mei, X. Zhang, E. Tan, H. Liu, X. Han, H. Zhan and X. Lu. 2022. A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*). Molecular Ecology Resources 22: 2396–2410.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter. 2002. Molecular Biology of the Cell. 4th ed. Garland Science, New York, 1616 pp.

Amarasinghe, S. L., S. Su, X. Dong, L. Zappia, M. E. Ritchie and Q. Gouil. 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biology 21: 30.

Ansari, J. A. and N. N. Inamdar. 2010. The promise of traditional medicines. International Journal of Pharmacology 6: 808–812.

Ashelford, K., M. E. Eriksson, C. M. Allen, R. D'Amore, M. Johansson, P. Gould, S. Kay, A. J. Millar, N. Hall and A. Hall. 2011. Full genome re-sequencing reveals a novel circadian clock mutation in *Arabidopsis*. Genome Biology 12: R28.

Aspinwall, M. J., M. E. Loik, V. Resco de Dios, M. G. Tjoelker, P. R. Payton and D. T. Tissue. 2015. Utilizing intraspecific variation in phenotypic plasticity to bolster agricultural and forest productivity under climate change. Plant, Cell & Environment 38: 1752–1764.

Auber, R. P., T. Suttiyut, R. M. McCoy, M. Ghaste, J. W. Crook, A. L. Pendleton, J. R. Widhalm and J. H. Wisecaver. 2020.

Hybrid de novo genome assembly of red gromwell (*Lithospermum erythrorhizon*) reveals evolutionary insight into shikonin biosynthesis. Horticulture Research 7: 82.

Bae, E.-K., C. An, M.-J. Kang, S.-A. Lee, S. J. Lee, K.-T. Kim and E.-J. Park. 2022. Chromosome-level genome assembly of the fully mycoheterotrophic orchid *Gastrodia elata*. G3 Genes|Genomes|Genetics 12: jkab433.

Baek, S., K. Choi, G.-B. Kim, H.-J. Yu, A. Cho, H. Jang, C. Kim, H.-J. Kim, K. S. Chang, J.-H. Kim and J.-H. Mun. 2018. Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. Genome Biology 19: 127.

Bai, W.-N., P.-C. Yan, B.-W. Zhang, K. E. Woeste, K. Lin and D.-Y. Zhang. 2018. Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus *Juglans* (Juglandaceae) revealed by whole-genome sequences. New Phytologist 217: 1726–1736.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. C. E. Catenazzi, S. Chang, R. N. Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G.-D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, Taksina Newington, Zemin Ning, Bee Ling Ng, Sonia M. Novo, Michael J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Par-

aschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. C. Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. E. S. Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin and A. J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59.

Blankenberg, D., G. V. Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko and J. Taylor. 2010. Galaxy: A web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology 89: 19.10. 11–19.10. 21.

Bocklandt, S., A. Hastie and H. Cao. 2019. Bionano genome mapping: high-throughput, ultra-long molecule genome analysis system for precision genome assembly and haploid-resolved structural variation discovery. *In* Single Molecule and Single Cell Sequencing. Advances in Experimental Medicine and Biology, Vol. 1129. Suzuki, Y. (ed.), Springer, Singapore. Pp. 97–118.

Bose, S., S. Banerjee, A. Mondal, U. Chakraborty, J. Pumarol, C. R. Croley and A. Bishayee. 2020. Targeting the JAK/STAT signaling pathway using phytocompounds for cancer prevention and therapy. Cells 9: 1451.

Buck, M. and C. Hamilton. 2011. The Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity. Review of European Community & International Environmental Law 20: 47–61.

Cao, J., K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, X. Wang, F. Ott, J. Müller, C. Alonso-Blanco, K. Borgwardt, K. J. Schmid and D. Weigel. 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nature Genetics 43: 956–963.

Chen, F., L. Su, S. Hu, J.-Y. Xue, H. Liu, G. Liu, Y. Jiang, J. Du, Y. Qiao, Y. Fan, H. Liu, Q. Yang, W. Lu, Z.-Q. Shao, J. Zhang, L. Zhang, F. Chen and Z.-M. (Max) Cheng. 2021. A chromosome-level genome assembly of rugged rose (*Rosa rugosa*) provides insights into its evolution, ecology, and floral characteristics. Horticulture Research 8: 141.

Chen, X., S. Li, D. Zhang, M. Han, X. Jin, C. Zhao, S. Wang, L. Xing, J. Ma, J. Ji and N. An. 2019. Sequencing of a wild apple (*Malus baccata*) genome unravels the differences between cultivated and wild apple species regarding disease resistance and

cold tolerance. G3: Genes, Genomes, Genetics 9: 2051–2060.

Chung, G. Y., H.-D. Jang, K. S. Chang, H. J. Choi, Y.-S. Kim, H.-J. Kim and D. C. Son. 2023. A checklist of endemic plants on the Korean Peninsula II. Korean Journal of Plant Taxonomy 53: 79–101.

Chung, O., J. Kim, D. Bolser, H.-M. Kim, J. H. Jun, J.-P. Choi, H.-D. Jang, Y. S. Cho, J. Bhak and M. Kwak. 2021. A chromosome-scale genome assembly and annotation of the spring orchid (*Cymbidium goeringii*). Molecular Ecology Resources 22: 1168–1177.

De Vega, J., I. Donnison, S. Dyer and K. Farrar. 2021. Draft genome assembly of the biofuel grass crop *Miscanthus sacchariflorus*. F1000Research 10: 29.

Fakhrai-Rad, H., N. Pourmand and M. Ronaghi. 2002. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. Human Mutation 19: 479–485.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. Fitzhugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-L. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. Fraser, H. O. Smith and J. Craig Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496–512.

Fuller, C. W., L. R. Middendorf, S. A. Benner, G. M. Church, T. Harris, X. Huang, S. B. Jovanovich, J. R. Nelson, J. A. Schloss, D. C. Schwartz and D. V. Vezenov. 2009. The challenges of sequencing by synthesis. Nature Biotechnology 27: 1013–1023.

Gan, X., O. Stegle, J. Behr, J. G. Steffen, P. Drewe, K. L. Hildebrand, R. Lyngsoe, S. J. Schultheiss, E. J. Osborne, V. T. Sreedharan, A. Kahles, R. Bohnert, G. Jean, P. Derwent, P. Kersey, E. J. Belfield, N. P. Harberd, E. Kemen, C. Toomajian, P. X. Kover, R. M. Clark, G. Rätsch and R. Mott. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature 477: 419–423.

Gao, Y., Q. Yang, X. Yan, X. Wu, F. Yang, J. Li, J. Wei, J. Ni, M. Ahmad, S. Bai and Y. Teng. 2021. High-quality genome assembly of 'Cuiguan' pear (*Pyrus pyrifolia*) as a reference genome for identifying regulatory genes and epigenetic modifications responsible for bud dormancy. Horticulture Research 8: 197.

Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent and A. Nekrutenko. 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Research 15: 1451–1455.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin and S. G. Oliver. 1996. Life with 6000 genes. Science 274: 546–567.

Gulyaev, S., X.-J. Cai, F.-Y. Guo, S. Kikuchi, W. L. Applequist, Z.-X. Zhang, E. Hörandl and L. He. 2022. The phylogeny of *Salix* revealed by whole genome re-sequencing suggests different sex-determination systems in major groups of the genus. Annals of Botany 129: 485–498.

Guo, L., J. Qiu, Z. Han, Z. Ye, C. Chen, C. Liu, X. Xin, C.-Y. Ye, Y.-Y. Wang, H. Xie, Y. Wang, J. Bao, S. Tang, J. Xu, Y. Gui, F. Fu, W. Wang, X. Zhang, Q. Zhu, X. Guang, C. Wang, H. Cui, D. Cai, S. Ge, G. A. Tuskan, X. Yang, Q. Qian, S. Y. He, J. Wang, X.-P. Zhou and L. Fan. 2015. A host plant genome (*Zizania latifolia*) after a century-long endophyte infection. The Plant Journal 83: 600–609.

Guo, L., J. Qiu, C. Ye, G. Jin, L. Mao, H. Zhang, X. Yang, Q. Peng, Y. Wang, L. Jia, Z. Lin, G. Li, F. Fu, C. Liu, L. Chen, E. Shen, W. Wang, Q. Chu, D. Wu, S. Wu, C. Xia, Y. Zhang, X. Zhou, L. Wang, L. Wu, W. Song, Y. Wang, Q. Shu, D. Aoki, E. Yumoto, T. Yokota, K. Miyamoto, K. Okada, D.-S. Kim, D. Cai, C. Zhang, Y. Lou, Q. Qian, H. Yamaguchi, H. Yamane, C.-H. Kong, M. P. Timko, L. Bai and L. Fan. 2017. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. Nature Communications 8: 1031.

Hirakawa, H., K. Shirasawa, S. Kosugi, K. Tashiro, S. Nakayama, M. Yamada, M. Kohara, A. Watanabe, Y. Kishida, T. Fujishiro, H. Tsuruoka, C. Minami, S. Sasamoto, M. Kato, K. Nanri, A. Komaki, T. Yanagi, Q. Guoxin, F. Maeda, M. Ishikawa, S. Kuhara, S. Sato, S. Tabata and S. N. Isobe. 2014. Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. DNA Research 21: 169–181.

Hon, T., K. Mars, G. Young, Y.-C. Tsai, J. W. Karalius, J. M. Landolin, N. Maurer, D. Kudrna, M. A. Hardigan, C. C. Steiner, S. J. Knapp, D. Ware, B. Shapiro, P. Peluso and D. R. Rank. 2020. Highly accurate long-read HiFi sequencing data for five complex genomes. Scientific Data 7: 399.

Hu, T. T., P. Pattyn, E. G. Bakker, J. Cao, J.-F. Cheng, R. M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haberer, J. D. Hollister, S. Ossowski, R. P. Ottilar, A. A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J. C. Carrington, B. S. Gaut, J. Schmutz, K. F. X. Mayer, Y. Van de Peer, I. V Grigoriev, M. Nordborg, D. Weigel and Y.-L. Guo. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size

change. Nature Genetics 43: 476–481.

Huang, D., R. Ming, S. Xu, J. Wang, S. Yao, L. Li, R. Huang and Y. Tan. 2021. Chromosome-level genome assembly of *Gynostemma pentaphyllum* provides insights into gypenoside biosynthesis. DNA Research 28: dsab018.

Huang, J., C. Zhang, X. Zhao, Z. Fei, K. Wan, Z. Zhang, X. Pang, X. Yin, Y. Bai, X. Sun, L. Gao, R. Li, J. Zhang and X. Li. 2016. The jujube genome provides insights into genome evolution and the domestication of sweetness/acidity taste in fruit trees. PLoS Genetics 12: e1006433.

Huang, S., R. Li, Z. Zhang, L. Li, X. Gu, W. Fan, W. J. Lucas, X. Wang, B. Xie, P. Ni, Y. Ren, H. Zhu, J. Li, K. Lin, W. Jin, Z. Fei, G. Li, J. Staub, A. Kilian, E. A. G. van der Vossen, Y. Wu, J. Guo, J. He, Z. Jia, Y. Ren, G. Tian, Y. Lu, J. Ruan, W. Qian, M. Wang, Q. Huang, B. Li, Z. Xuan, J. Cao, Asan, Z. Wu, J. Zhang, Q. Cai, Y. Bai, B. Zhao, Y. Han, Y. Li, X. Li, S. Wang, Q. Shi, S. Liu, W. K. Cho, J.-Y. Kim, Y. Xu, K. Heller-Uszynska, H. Miao, Z. Cheng, S. Zhang, J. Wu, Y. Yang, H. Kang, M. Li, H. Liang, X. Ren, Z. Shi, M. Wen, M. Jian, H. Yang, G. Zhang, Z. Yang, R. Chen, S. Liu, J. Li, L. Ma, H. Liu, Y. Zhou, J. Zhao, X. Fang, G. Li, L.Fang, Y. Li, D. Liu, H. Zheng, Y. Zhang, N. Qin, Z. Li, G. Yang, S. Yang, L. Bolund, K. Kristiansen, H. Zheng, S. Li, X. Zhang, H. Yang, J. Wang, R. Sun, B.Zhang, S. Jiang, J. Wang, Y. Du and S. Li. 2009. The genome of the cucumber, *Cucumis sativus* L. Nature Genetics 41: 1275–1281.

Hutchison, C. A. III, R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Qi, R. Alexander Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson and J. Craig Venter. 2016. Design and synthesis of a minimal bacterial genome. Science 351: aad6253.

International Wheat Genome Sequencing Consortium (IWGSC). 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361: eaar7191.

Jaillon, O., J.-M. Aury, B. Noel, A. Policriti, C. Clepet, A. Cassagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyère, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. Enrico Pè, G. Valle, M. Morgante, M. Caboche, A.-F. Adam-Blondon, J. Weissenbach, F. Quétier, P. Wincker and French-Italian Public Consortium for Grapevine Genome Characterization. 2007.

The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–467.

Jang, W., J.-N. Kang, I.-H. Jo, S.-M. Lee, G.-H. Park and C.-K. Kim. 2023. The chromosome-level genome assembly of lance asiabell (*Codonopsis lanceolata*), a medicinal and vegetable plant of the Campanulaceae family. Frontiers in Genetics 14: 1100819.

Jones-Rhoades, M. W., D. P. Bartel and B. Bartel. 2006. MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology 57: 19–53.

Kang, S.-H., R. P. Pandey, C.-M. Lee, J.-S. Sim, J.-T. Jeong, B.-S. Choi, M. Jung, D. Ginzburg, K. Zhao, S. Y. Won, T.-J. Oh, Y. Yu, N.-H. Kim, O. R. Lee, T.-H. Lee, P. Bashyal, T.-S. Kim, W.-H. Lee, C. Hawkins, C.-K. Kim, J. S. Kim, B. O. Ahn, S. Y. Rhee and J. K. Sohng. 2020. Genome-enabled discovery of anthraquinone biosynthesis in *Senna tora*. Nature Communications 11: 5875.

Kasianov, A. S., A. V. Klepikova, I. V. Kulakovskiy, E. S. Gerasimov, A. V. Fedotova, E. G. Besedina, A. S. Kondrashov, M. D. Logacheva and A. A. Penin. 2017. High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution. The Plant Journal 91: 278–291.

Kim, J.-I., Y. S. Ju, H. Park, S. Kim, S. Lee, J.-H. Yi, J. Mudge, N. A. Miller, D. Hong, C. J. Bell, H.-S. Kim, I.-S. Chung, W.-C. Lee, J.-S. Lee, S.-H. Seo, J.-Y. Yun, H. N. Woo, H. Lee, D. Suh, S. Lee, H.-J. Kim, M. Yavartanoo, M. Kwak, Y. Zheng, M. K. Lee, H. Park, J. Y. Kim, O. Gokcumen, R. E. Mills, A. W. Zaranek, J. Thakuria, X. Wu, R. W. Kim, J. J. Huntley, S. Luo, G. P. Schroth, T. D. Wu, H. Kim, K.-S. Yang, W.-Y. Park, H. Kim, G. M. Church, C. Lee, S. F. Kingsmore and J.-S. Seo. 2009. A highly annotated whole-genome sequence of a Korean individual. Nature 460: 1011–1015.

Kim, M. Y., S. Lee, K. Van, T.-H. Kim, S.-C. Jeong, I.-Y. Choi, D.-S. Kim, Y.-S. Lee, D. Park, J. Ma, W.-Y. Kim, B.-C. Kim, S. Park, K.-A. Lee, D. H. Kim, K. H. Kim, J. H. Shin, Y. E. Jang, K. D. Kim, W. X. Liu, T. Chaisan, Y. J. Kang, Y.-H. Lee, K.-H. Kim, J.-K. Moon, J. Schmutz, S. A. Jackson, J. Bhak and S.-H. Lee. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences 107: 22032–22037.

Kim, M., H. Xi and J. Park. 2021a. Genome-wide comparative analyses of GATA transcription factors among 19 Arabidopsis ecotype genomes: Intraspecific characteristics of GATA transcription factors. PLoS ONE 16: e0252181.

Kim, M., H. Xi, S. Park, Y. Yun and J. Park. 2021b. Genome-wide comparative analyses of GATA transcription factors among seven *Populus* genomes. Scientific Reports 11: 16578.

Kim, N.-H., M. Jayakodi, S.-C. Lee, B.-S. Choi, W. Jang, J. Lee,

H. H. Kim, N. E. Waminal, M. Lakshmanan, B. van Nguyen, Y. S. Lee, H.-S. Park, H. J. Koo, J. Y. Park, S. Perumal, H. J. Joh, H. Lee, J. Kim, I. S. Kim, K. Kim, L. Koduru, K. B. Kang, S. H. Sung, Y. Yu, D. S. Park, D. Choi, E. Seo, S. Kim, Y.-C. Kim, D. Y. Hyun, Y.-I. Park, C. Kim, T.-H. Lee, H. U. Kim, M. S. Soh, Y. Lee, J. G. In, H.-S. Kim, Y.-M. Kim, D.-C. Yang, R. A. Wing, D.-Y. Lee, A. H. Paterson and T.-J. Yang. 2018. Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. Plant Biotechnology Journal 16: 1904–1917.

Kong, S. and Y. Zhang. 2019. Deciphering Hi-C: From 3D genome to function. Cell Biology and Toxicology 35: 15–32.

Lee, J. E., M. Neumann, D. I. Duro and M. Schmid. 2019. CRISPR-based tools for targeted transcriptional and epigenetic regulation in plants. PLoS ONE 14: e0222778.

Lee, J., J. Park, H. Xi and J. Park. 2020. Comprehensive analyses of the complete mitochondrial genome of *Figulus binodulus* (Coleoptera: Lucanidae). Journal of Insect Science 20: 10.

Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, K. A. Crandall, R. Durbin, S. V. Edwards, F. Forest, M. T. P. Gilbert, M. M. Goldstein, I. V. Grigoriev, K. J. Hackett, D. Haussler, E. D. Jarvis, W. E. Johnson, A. Patrinos, S. Richards, J. C. Castilla-Rubio, M.-A. van Sluys, P. S. Soltis, X. Xu, H. Yang and G. Zhang. 2018. Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences of the United States of America 115: 4325–4333.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326: 289–293.

Liggett, S. B. 2001. Pharmacogenetic applications of the Human Genome project. Nature Medicine 7: 281–283.

Liu, C., L. Zeng, S. Zhu, L. Wu, Y. Wang, S. Tang, H. Wang, X. Zheng, J. Zhao, X. Chen, Q. Dai and T. Liu. 2018. Draft genome analysis provides insights into the fiber yield, crude protein biosynthesis, and vegetative growth of domesticated ramie (*Boehmeria nivea* L. Gaud). DNA Research 25: 173–181.

Liu, J.-X., Q. Jiang, J.-P. Tao, K. Feng, T. Li, A.-Q. Duan, H. Wang, Z.-S. Xu, H. Liu and A.-S. Xiong. 2021a. Integrative genome, transcriptome, microRNA, and degradome analysis of water dropwort (*Oenanthe javanica*) in response to water stress. Horticulture Research 8: 262.

Liu, M.-J., J. Zhao, Q.-L. Cai, G-C. Liu, J.-R. Wang, Z.-H. Zhao, P. Liu, L. Dai, G. Yan, W.-J. Wang, X.-S. Li, Y. Chen, Y.-D. Sun, Z.-G. Liu, M.-J. Lin, J. Xiao, Y.-Y. Chen, X.-F. Li, B. Wu, Y. Ma, J.-B. Jian, W. Yang, Z. Yuan, X.-C. Sun, Y.-L. Wei, L.-

L. Yu, C. Zhang, S.-G. Liao, R.-J. He, X.-M. Guang, Z. Wang, Y.-Y. Zhang and L.-H. Luo. 2014. The complex jujube genome provides insights into fruit tree biology. Nature Communications 5: 5315.

Liu, Y., X. Zhang, K. Han, R. Li, G. Xu, Y. Han, F. Cui, S. Fan, I. Seim, G. Fan, G. Li and S. Wan. 2021b. Insights into amphicarpy from the compact genome of the legume *Amphicarpaea edgeworthii*. Plant Biotechnology Journal 19: 952–965.

Long, Q., F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow, A. Platzer, Q. Zhang, B. J. Vilhjálmsson, A. Korte, V. Nizhynska, V. Voronin, P. Korte, L. Sedman, T. Mandáková, M. A. Lysak, Ü. Seren, I. Hellmann and M. Nordborg. 2013. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. Nature Genetics 45: 884–890.

Mamidi, S., A. Healey, P. Huang, J. Grimwood, J. Jenkins, K. Barry, A. Sreedasyam, S. Shu, J. T. Lovell, M. Feldman, J. Wu, Y. Yu, C. Chen, J. Johnson, H. Sakakibara, T. Kiba, T. Sakurai, R. Tavares, D. A. Nusinow, I. Baxter, J. Schmutz, T. P. Brutnell and E. A. Kellogg. 2020. A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. Nature Biotechnology 38: 1203–1210.

Martienssen, R. and W. R. McCombie. 2001. The first plant genome. Cell 105: 571–574.

Michael, T. P., D. Bryant, R. Gutierrez, N. Borisjuk, P. Chu, H. Zhang, J. Xia, J. Zhou, H. Peng, M. El Baidouri, B. T. Hallers, A. R. Hastie, T. Liang, K. Acosta, S. Gilbert, C. McEntee, S. A. Jackson, T. C. Mockler, W. Zhang and E. Lam. 2017. Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. The Plant Journal 89: 617–635.

Miles, B. N., A. P. Ivanov, K. A. Wilson, F. Doğan, D. Japrung and J. B. Edel. 2013. Single molecule sensing with solid-state nanopores: Novel materials, methods, and applications. Chemical Society Reviews 42: 15–28.

Mitchell-Olds, T. 2001. *Arabidopsis thaliana* and its wild relatives: A model system for ecology and evolution. Trends in Ecology and Evolution 16: 693–700.

Mitros, T., A. M. Session, B. T. James, G. A. Wu, M. B. Belaffif, L. V. Clark, S. Shu, H. Dong, A. Barling, J. R. Holmes, J. E. Mattick, J. V. Bredeson, S. Liu, K. Farrar, K. Głowacka, S. Jeżowski, K. Barry, W. B. Chae, J. A. Juvik, J. Gifford, A. Oladeinde, T. Yamada, J. Grimwood, N. H. Putnam, J. De Vega, S. Barth, M. Klaas, T. Hodkinson, L. Li, X. Jin, J. Peng, C. Y. Yu, K. Heo, J. H. Yoo, B. K. Ghimire, I. S. Donnison, J. Schmutz, M. E. Hudson, E. J. Sacks, S. P. Moose, K. Swaminathan and D. S. Rokhsar. 2020. Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. Nature Communications 11: 5442.

Moore, B. D., R. L. Andrew, C. Külheim and W. J. Foley. 2014.

Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytologist 201: 733–750.

Naito, K., T. Wakatake, T. F. Shibata, K. Iseki, S. Shigenobu, Y. Takahashi, E. Ogiso-Tanaka, C. Muto, K. Teruya, A. Shiroma, M. Shimoji, K. Satou, T. Hirano, A. J. Nagano, N. Tomooka, M. Hasebe, K. Fukushima and H. Sakai. 2022. Genome sequence of 12 *Vigna* species as a knowledge base of stress tolerance and resistance. bioRxiv https://doi.org/10.1101/2022.03.28.486085.

Nakamura, N., H. Hirakawa, S. Sato, S. Otagaki, S. Matsumoto, S. Tabata and Y. Tanaka. 2018. Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. DNA Research 25: 113–121.

Ngo, L. T., J. I. Okogun and W. R. Folk. 2013. 21st century natural product research and drug development and traditional medicines. Natural Product Reports 30: 584–592.

Norn, S., H. Permin, P. R. Kruse and E. Kruse. 2009. From willow bark to acetylsalicylic acid. Dansk Medicinhistorisk Arbog 37: 79–98.

Olsen, J. L., P. Rouzé, B. Verhelst, Y.-C. Lin, T. Bayer, J. Collen, E. Dattolo, E. De Paoli, S. Dittami, F. Maumus, G. Michel, A. Kersting, C. Lauritano, R. Lohaus, M. Töpel, T. Tonon, K. Vanneste, M. Amirebrahimi, J. Brakel, C. Boström, M. Chovatia, J. Grimwood, J. W. Jenkins, A. Jueterbock, A. Mraz, W. T. Stam, H. Tice, E. Bornberg-Bauer, P. J. Green, G. A. Pearson, G. Procaccini, C. M. Duarte, J. Schmutz, T. B. H. Reusch and Y. Van de Peer. 2016. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. Nature 530: 331–335.

Osoegawa, K., A. G. Mammoser, C. Wu, E. Frengen, C. Zeng, J. J. Catanese and P. J. de Jong. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. Genome Research 11: 483–496.

Ossowski, S., K. Schneeberger, R. M. Clark, C. Lanz, N. Warthmann and D. Weigel. 2008. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Research 18: 2024–2033.

Park, J., J.-H. An, Y. Kim, D. Kim, B.-G. Yang and T. Kim. 2020a. Database of National Species List of Korea: The taxonomical systematics platform for managing scientific names of Korean native species. Journal of Species Research 9: 233–246.

Park, J., M. Kim, H. Xi, S. Park and Y. Yun. 2021a. Plant Genome Database Release 2.7: One tera base pairs plant genomes era. *In* The 17th KOGO Winter Symposium; 2021 Feb 1-2; Seoul, Korea.

Park, J., S. H. Lee and J. H. Kim. 2021b. Complete genome sequence of the endosymbiotic bacterium "*Candidatus* Riesia pediculicola". Microbiology Resource Announcements 10: e01181–01120.

Park, J., H. Xi, J. Han, J. Lee, Y. Kim, J.-M. Lee, J. Son, J. Ahn, T. Jang, J. Choi and J. Park. 2020b. Prediction and identification of biochemical pathway of acteoside from whole genome sequences of *Abeliophyllum distichum* Nakai, cultivar Ok Hwang 1ho. Journal of Convergence for Information Technology 10: 76–91.

Park, J., H. Xi and Y. Kim. 2020c. The complete chloroplast genome of *Arabidopsis thaliana* isolated in Korea (Brassicaceae): An investigation of intraspecific variations of the chloroplast genome of Korean *A. thaliana*. International Journal of Genomics 2020: 3236461.

Pootakham, W., C. Naktang, W. Kongkachana, C. Sonthirod, T. Yoocha, D. Sangsrakru, N. Jomchai, S. U-thoomporn, K. Romyanon, T. Toojinda and S. Tangphatsornruang. 2021. *De novo* chromosome-level assembly of the *Centella asiatica* genome. Genomics 113: 2221–2228.

Ren, L., X. Guo, S. Liu, T. Yu, W. Guo, R. Wang, S. Ye, C. Lambertini, H. Brix and F. Eller. 2020. Intraspecific variation in *Phragmites australis*: Clinal adaption of functional traits and phenotypic plasticity vary with latitude of origin. Journal of Ecology 108: 2531–2543.

Salojärvi, J., O.-P. Smolander, K. Nieminen, S. Rajaraman, O. Safronov, P. Safdari, A. Lamminmäki, J. Immanen, T. Lan, J. Tanskanen, P. Rastas, A. Amiryousefi, B. Jayaprakash, J. I. Kammonen, R. Hagqvist, G. Eswaran, V. H. Ahonen, J. A. Serra, F. O. Asiegbu, J. de Dios Barajas-Lopez, D. Blande, O. Blokhina, T. Blomster, S. Broholm, M. Brosché, F. Cui, C. Dardick, S. E. Ehonen, P. Elomaa, S. Escamez, K. V. Fagerstedt, H. Fujii, A. Gauthier, P. J. Gollan, P. Halimaa, P. I. Heino, K. Himanen, C. Hollender, S. Kangasjärvi, L. Kauppinen, C. T. Kelleher, S. Kontunen-Soppela, J. P. Koskinen, A. Kovalchuk, S. O. Kärenlampi, A. K. Kärkönen, K.-J. Lim, J. Leppälä, L. Macpherson, J. Mikola, K. Mouhu, A. P. Mähönen, Ü. Niinemets, E. Oksanen, K. Overmyer, E. T. Palva, L. Pazouki, V. Pennanen, T. Puhakainen, P. Poczai, B. J. H. M. Possen, M. Punkkinen, M. M. Rahikainen, M. Rousi, R. Ruonala, C. van der Schoot, A. Shapiguzov, M. Sierla, T. P. Sipilä, S. Sutela, T. H. Teeri, Arja. I. Tervahauta, A. Vaattovaara, J. Vahala, L. Vetchinnikova, A. Welling, M. Wrzaczek, E. Xu, L. G. Paulin, A. H. Schulman, M. Lascoux, V. A. Albert, P. Auvinen, Y. Helariutta and J. Kangasjärvi. 2017. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. Nature Genetics 49: 904–912.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe and M. Smith. 1977. Nucleotide sequence of bacteriophage φX174 DNA. Nature 265: 687–695.

Schmitz, R. J., M. D. Schultz, M. A. Urich, J. R. Nery, M. Peliz-

zola, O. Libiger, A. Alix, R. B. McCosh, H. Chen, N. J. Schork and J. R. Ecker. 2013. Patterns of population epigenomic diversity. Nature 495: 193–198.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T. A. Graves, P. Minx, A. D. Reily, L. Courtney, S. S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S. M. Rock, E. Belter, F. Du, K. Kim, R. M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S. M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M. J. Levy, L. McMahan, P. Van Buren, M. W. Vaughn, K. Ying, C.-T. Yeh, S. J. Emrich, Y. Jia, A. Kalyanaraman, A.-P. Hsia, W. B. Barbazuk, R. S. Baucom, T. P. Brutnell, N. C. Carpita, C. Chaparro, J.-M. Chia, J.-M. Deragon, J. C. Estill, Y. Fu, J. A. Jeddeloh, Y. Han, H. Lee, P. Li, D. R. Lisch, S. Liu, Z. Liu, D. H. Nagel, M. C. McCann, P. SanMiguel, A. M. Myers, D. Nettleton, J. Nguyen, B. W. Penning, L. Ponnala, K. L. Schneider, D. C. Schwartz, A. Sharma, C. Soderlund, N. M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T. K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J. L. Bennetzen, R. K. Dawe, J. Jiang, N. Jiang, G. G. Presting, S. R. Wessler, S. Aluru, R. A. Martienssen, S. W. Clifton, W. R. McCombie, R. A. Wing and R. K. Wilson. 2009. The B73 maize genome: Complexity, diversity, and dynamics. Science 326: 1112–1115.

Shen, L.-Y., H. Luo, X.-L. Wang, X.-M. Wang, X.-J. Qiu, H. Liu, S.-S. Zhou, K.-H. Jia, S. Nie, Y.-T. Bao, R.-G. Zhang, Q.-Z. Yun, Y.-H. Chai, J.-Y. Lu, Y. Li, S.-W. Zhao, J.-F. Mao, S.-G. Jia, and Y.-M. Mao. 2021. Chromosome-scale genome assembly for Chinese sour jujube and insights into its genome evolution and domestication signature. Frontiers in Plant Science 12: 773090.

Shen, Q., L. Zhang, Z. Liao, S. Wang, T. Yan, P. Shi, M. Liu, X. Fu, Q. Pan, Y. Wang, Z. Lv, X. Lu, F. Zhang, W. Jiang, Y. Ma, M. Chen, X. Hao, L. Li, Y. Tang, G. Lv, Y. Zhou, X. Sun, P. E. Brodelius, J. K.C. Rose and K. Tang. 2018. The genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. Molecular Plant 11: 776–788.

Shirasawa, K., T. Esumi, H. Hirakawa, H. Tanaka, A. Itai, A. Ghelfi, H. Nagasaki and S. Isobe. 2019. Phased genome sequence of an interspecific hybrid flowering cherry, 'Somei-Yoshino' (Cerasus× yedoensis). DNA Research 26: 379–389.

Shirasawa, K., A. Itai and S. Isobe. 2021a. Chromosome-scale genome assembly of Japanese pear (*Pyrus pyrifolia*) variety 'Nijisseiki'. DNA Research 28: dsab001.

Shirasawa, K., S. Kosugi, K. Sasaki, A. Ghelfi, K. Okazaki, A. Toyoda, H. Hirakawa and S. Isobe. 2021b. Genome features of common vetch (*Vicia sativa*) in natural habitats. Plant Direct 5: e352.

Shirasawa, K., S. Nishio, S. Terakami, R. Botta, D. T. Marinoni and S. Isobe. 2021c. Chromosome-level genome assembly of Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) reveals conserved chromosomal segments in woody rosids. DNA Research 28: dsab016.

Shirasawa, K., H. Yakushiji, R. Nishimura, T. Morita, S. Jikumaru, H. Ikegami, A. Toyoda, H. Hirakawa and S. Isobe. 2020. The *Ficus erecta* genome aids Ceratocystis canker resistance breeding in common fig (*F. carica*). The Plant Journal 102: 1313–1322.

Slavov, G. T., S. P. DiFazio, J. Martin, W. Schackwitz, W. Muchero, E. Rodgers-Melnick, M. F. Lipphardt, C. P. Pennacchio, U. Hellsten, L. A. Pennacchio, L. E. Gunter, P. Ranjan, K. Vining, K. R. Pomraning, L. J. Wilhelm, M. Pellegrini, T. C. Mockler, M. Freitag, A. Geraldes, Y. A. El-Kassaby, S. D. Mansfield, Q. C. B. Cronk, C. J. Douglas, S. H. Strauss, D. Rokhsar and G. A. Tuskan. 2012. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. New Phytologist 196: 713–725.

Šmarda, P., P. Bureš, L. Horová, I. J. Leitch, L. Mucina, E. Pacini, L. Tichý, V. Grulich and O. Rotreklová. 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. Proceedings of the National Academy of Sciences of the United States of America 111: E4096–E4102.

Smith, A. M., L. E. Heisler, R. P. St. Onge, E. Farias-Hesson, I. M. Wallace, J. Bodeau, A. N. Harris, K. M. Perry, G. Giaever, N. Pourmand and C. Nislow. 2010. Highly-multiplexed barcode sequencing: An efficient method for parallel analysis of pooled samples. Nucleic Acids Research 38: e142.

Sollars, E. S. A., A. L. Harper, L. J. Kelly, C. M. Sambles, R. H. Ramirez-Gonzalez, D. Swarbreck, G. Kaithakottil, E. D. Cooper, C. Uauy, L. Havlickova, G. Worswick, D. J. Studholme, J. Zohren, D. L. Salmon, B. J. Clavijo, Y. Li, Z. He, A. Fellgett, L. V. McKinney, L. R. Nielsen, G. C. Douglas, E. D. Kjær, J. A. Downie, D. Boshier, S. Lee, J. Clark, M. Grant, I. Bancroft, M. Caccamo and R. J. A. Buggs. 2017. Genome

sequence and genetic diversity of European ash trees. Nature 541: 212–216.

Song, X., Y. Li, X. Cao and Y. Qi. 2019. MicroRNAs and their regulatory roles in plant–environment interactions. Annual Review of Plant Biology 70: 489–525.

Sprink, T., R. Wilhelm and F. Hartung. 2022. Genome editing around the globe: An update on policies and perceptions. Plant Physiology 190: 1579–1587.

Staton, M., C. Addo-Quaye, N. Cannon, J. Yu, T. Zhebentyayeva, M. Huff, N. Islam-Faridi, S. Fan, L. L. Georgi, C. D. Nelson, E. Bellis, S. Fitzsimmons, N. Henry, D. Drautz-Moses, R. E. Noorai, S. Ficklin, C. Saski, M. Mandal, T. K. Wagner, N. Zembower, C. Bodénès, J. Holliday, J. Westbrook, J. Lasky, F. V. Hebard, S. C. Schuster, A. G. Abbott and J. E. Carlson. 2020. A reference genome assembly and adaptive trait analysis of *Castanea mollissima* 'Vanuxem,' a source of resistance to chestnut blight in restoration breeding. Tree Genetics and Genomes 16: 57.

Sun, G., Y. Xu, H. Liu, T. Sun, J. Zhang, C. Hettenhausen, G. Shen, J. Qi, Y. Qin, J. Li, L. Wang, W. Chang, Z. Guo, I. T. Baldwin and J. Wu. 2018. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. Nature Communications 9: 2683.

Tahir, J., R. Crowhurst, S. Deroles, E. Hilario, C. Deng, R. Schaffer, L. Le Lievre, C. Brendolise, D. Chagne, S. E. Gardiner, M. Knaebel, A. Catanach, J. McCallum, P. Datson, S. Thomson, L. R. Brownfield, S. Nardozza and S. M. Pilkington. 2022. First chromosome-scale assembly and deep floral-bud transcriptome of a male kiwifruit. Frontiers in Genetics 13: 852161.

Tan, Q., S. Li, Y. Zhang, M. Chen, B. Wen, S. Jiang, X. Chen, X. Fu, D. Li, H. Wu, Y. Wang, W. Xiao and L. Li. 2021. Chromosome-level genome assemblies of five *Prunus* species and genome-wide association studies for key agronomic traits in peach. Horticulture Research 8: 213.

Tanaka, H., H. Hirakawa, S. Kosugi, S. Nakayama, A. Ono, A. Watanabe, M. Hashiguchi, T. Gondo, G. Ishigaki, M. Muguerza, K. Shimizu, N. Sawamura, T. Inoue, Y. Shigeki, N. Ohno, S. Tabata, R. Akashi and S. Sato. 2016. Sequencing and comparative analyses of the genomes of zoysiagrasses. DNA Research 23: 171–180.

The 1001 Genomes Consortium. 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell 166: 481–491.

The 3,000 Rice Genomes Project. 2014. The 3,000 rice genomes project. GigaScience 3: 7.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Thielen, P. M., A. L. Pendleton, R. A. Player, K. V. Bowden, T. J. Lawton and J. H. Wisecaver. 2020. Reference genome for the highly transformable *Setaria viridis* ME034V. G3: Genes, Genomes, Genetics 10: 3467–3478.

Tuskan, G. A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R. R. Bhalerao, R. P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G.-L. Chen, D. Cooper, P. M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Déjardin, C. Depamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson, C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J.-C. Leplé, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D. R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C.-J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer and D. Rokhsar. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604.

Valliyodan, B., S. B. Cannon, P. E. Bayer, S. Shu, A. V. Brown, L. Ren, J. Jenkins, C. Y.-L. Chung, T.-F. Chan, C. G. Daum, C. Plott, A. Hastie, K. Baruch, K. W. Barry, W. Huang, G. Patil, R. K. Varshney, H. Hu, J. Batley, Y. Yuan, Q. Song, R. M. Stupar, D. M. Goodstein, G. Stacey, H.-M. Lam, S. A. Jackson, J. Schmutz, J. Grimwood, D. Edwards and H. T. Nguyen. 2019. Construction and comparison of three reference-quality genome assemblies for soybean. The Plant Journal 100: 1066–1082.

Van Hoeck, A., N. Horemans, P. Monsieurs, H. X. Cao, H. Vandenhove and R. Blust. 2015. The first draft genome of the aquatic model plant *Lemna minor* opens the route for future stress physiology research and biotechnological applications. Biotechnology for Biofuels 8: 188.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannen-

halli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Deslattes Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu. 2001. The sequence of the human genome. Science 291: 1304–1351.

Wang, J., S. Tian, X. Sun, X. Cheng, N. Duan, J. Tao and G. Shen. 2020. Construction of pseudomolecules for the Chinese chestnut (*Castanea mollissima*) genome. G3: Genes, Genomes, Genetics 10: 3565–3574.

Wang, J., W. Wang, R. Li, Y. Li, G. Tian, L. Goodman, W. Fan, J. Zhang, J. Li, J. Zhang, Y. Guo, B. Feng, H. Li, Y. Lu, X. Fang, H. Liang, Z. Du, D. Li, Y. Zhao, Y. Hu, Z. Yang, H. Zheng, I. Hellmann, M. Inouye, J. Pool, X. Yi, J. Zhao, J. Duan, Y. Zhou, J. Qin, L. Ma, G. Li, Z. Yang, G. Zhang, B. Yang, C. Yu, F. Liang, W. Li, S. Li, D. Li, P. Ni, J. Ruan, Q. Li, H. Zhu, D. Liu, Z. Lu, N. Li, G. Guo, J. Zhang, J. Ye, L. Fang, Q. Hao, Q. Chen, Y. Liang, Y. Su, A. San, C. Ping, S. Yang, F. Chen, L. Li, K. Zhou, H. Zheng, Y. Ren, L. Yang, Y. Gao, G. Yang, Z. Li, X. Feng, K. Kristiansen, G. K.-S. Wong, R. Nielsen, R. Durbin, L. Bolund, X. Zhang, S. Li, H. Yang and J. Wang. 2008. The diploid genome sequence of an Asian individual. Nature 456: 60–65.

Wang, W., G. Haberer, H. Gundlach, C. Gläßer, T. Nussbaumer, M. Luo, A. Lomsadze, M. Borodovsky, R. Kerstetter, J. Shanklin, D. W. Byrant, T. C. Mockler, K. J. Appenroth, J. Grimwood, J. Jenkins, J. Chow, C. Choi, C. Adam, X.-H. Cao, J. Fuchs, I. Schubert, D. Rokhsar, J. Schmutz, T. P. Michael, K. F. X. Mayer and J. Messing. 2014. The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nature Communications 5: 3311.

Wang, Y., F. Li, Q. He, Z. Bao, Z. Zeng, D. An, T. Zhang, L. Yan, H. Wang, S. Zhu and T. Liu. 2021a. Genomic analyses provide comprehensive insights into the domestication of bast fiber crop ramie (*Boehmeria nivea*). The Plant Journal 107: 787–800.

Wang, Y., H. Xin, P. Fan, J. Zhang, Y. Liu, Y. Dong, Z. Wang, Y. Yang, Q. Zhang, R. Ming, G.-Y. Zhong, S. Li and Z. Liang. 2021b. The genome of Shanputao (*Vitis amurensis*) provides a new insight into cold tolerance of grapevine. The Plant Journal 105: 1495–1506.

Weber, J. L. and E. W. Myers. 1997. Human whole-genome shotgun sequencing. Genome Research 7: 401–409.

Weisenfeld, N. I., V. Kumar, P. Shah, D. M. Church and D. B. Jaffe. 2017. Direct determination of diploid genome sequences. Genome Research 27: 757–767.

Widanagama, S. D., J. R. Freeland, X. Xu and A. B. A. Shafer. 2022. Genome assembly, annotation, and comparative analysis of the cattail *Typha latifolia*. G3: Genes, Genomes, Genetics. 12: jkab401.

Williams, L. J. S., D. G. Tabbaa, N. Li, A. M. Berlin, T. P. Shea, I. MacCallum, M. S. Lawrence, Y. Drier, G. Getz, S. K. Young, D. B. Jaffe, C. Nusbaum and A. Gnirke. 2012. Paired-end sequencing of Fosmid libraries by Illumina. Genome Research 22: 2241–2249.

Wu, H., D. Yao, Y. Chen, W. Yang, W. Zhao, H. Gao and C. Tong. 2020. *De novo* genome assembly of *Populus simonii* further supports that *Populus simonii* and *Populus trichocarpa* belong to different sections. G3: Genes, Genomes, Genetics 10: 455–466.

Xie, M., C. Y.-L. Chung, M.-W. Li, F.-L. Wong, X. Wang, A. Liu,

Z. Wang, A. K.-Y. Leung, T.-H. Wong. S.-W. Tong, Z. Xiao, K. Fan, M.-S. Ng, X. Qi, L. Yang, T. Deng, L. He, L. Chen, A. Fu, Q. Ding, J. He, G. Chung, S. Isobe, T. Tanabata, B. Valliyodan, H. T. Nguyen, S. B. Cannon, C. H. Foyer, T.-F. Chan and H.-M. Lam. 2019. A reference-grade wild soybean genome. Nature Communications 10: 1216.

Xing, Y., Y. Liu, Q. Zhang, X. Nie, Y. Sun, Z. Zhang, H. Li, K. Fang, G. Wang, H. Huang, T. isseling, Q. Cao and L. Qin. 2019. Hybrid *de novo* genome assembly of Chinese chestnut (*Castanea mollissima*). GigaScience 8: giz112.

Xu, J., Y. Chu, B. Liao, S. Xiao, Q. Yin, R. Bai, H. Su, L. Dong, X. Li, J. Qian, J. Zhang, Y. Zhang, X. Zhang, M. Wu, J. Zhang, G. Li, L. Zhang, Z. Chang, Y. Zhang, Z. Jia, Z. Liu, D. Afreh, R. Nahurira, L. Zhang, R. Cheng, Y. Zhu, G. Zhu, W. Rao, C. Zhou, L. Qiao, Z. Huang, Y.-C. Cheng and S. Chen. 2017. *Panax ginseng* genome examination for ginsenoside biosynthesis. GigaScience 6: gix093.

Yin, M., S. Zhang, X. Du, R. G. Mateo, W. Guo, A. Li, Z. Wang, S. Wu, J. Chen, J. Liu and G. Ren. 2021. Genomic analysis of *Medicago ruthenica* provides insights into its tolerance to abiotic stress and demographic history. Molecular Ecology Resources 21: 1641–1657.

Yuan, Y., X. Jin, J. Liu, X. Zhao, J. Zhou, X. Wang, D. Wang, C. Lai, W. Xu, J. Huang, L. Zha, D. Liu, X. Ma, L. Wang, M. Zhou, Z. Jiang, H. Meng, H. Peng, Y. Liang, R. Li, C. Jiang, Y. Zhao, T. Nan, Y. Jin, Z. Zhan, J. Yang, W. Jiang and L. Huang. 2018. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. Nature Communications 9: 1615.

Zhang, H., N. Hall, L. R. Goertzen, B. Bi, C. Y. Chen, E. Peatman, E. K. Lowe, J. Patel and J. S. McElroy. 2019. Development of a goosegrass (*Eleusine indica*) draft genome and application to weed science research. Pest Management Science 75: 2776–2784.

Zhang, H.-B. and C. Wu. 2001. BAC as tools for genome sequencing. Plant Physiology and Biochemistry 39: 195–209.

Zhao, T., W. Ma, Z. Yang, L. Liang, X. Chen, G. Wang, Q. Ma and L. Wang. 2021. A chromosome-level reference genome of the hazelnut, *Corylus heterophylla* Fisch. GigaScience 10: giab027.

Zimin, A., K. A. Stevens, M. W. Crepeau, A. Holtz-Morris, M. Koriabine, G. Marçais, D. Puiu, M. Roberts, J. L. Wegrzyn, P. J. de Jong, D. B. Neale, S. L. Salzberg, J. A. Yorke and C. H. Langley. 2014. Sequencing and assembly of the 22-Gb loblolly pine genome. Genetics 196: 875–890.

Zou, Y.-P., X.-H. Hou, Q. Wu, J.-F. Chen, Z.-W. Li, T.-S. Han, X.-M. Niu, L. Yang, Y.-C. Xu, J. Zhang, F.-M. Zhang, D. Tan, Z. Tian, H. Gu and Y.-L. Guo. 2017. Adaptation of *Arabidopsis thaliana* to the Yangtze River basin. Genome Biology 18: 239.

Züst, T., S. R. Strickler, A. F. Powell, M. E. Mabry, H. An, M. Mirzaei, T. York, C. K. Holland, P. Kumar, M. Erb, G. Petschenka, J.-M. Gómez, F. Perfectti, C. Müller, J. C. Pires, L. A. Mueller and G. Jander. 2020. Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). Elife 9: e51712.