the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) **118,** 2296–2317 doi: 10.1111/tpj.16708

TECHNICAL ADVANCE

Degenerate oligonucleotide primer MIG-seq: an effective PCR-based method for high-throughput genotyping

Kazusa Nishimura[1,2,*], Hiroyuki Kokaji[1], Ko Motoki[1,2], Akira Yamazaki[3], Kyoka Nagasaka[1], Takashi Mori[1], Rihito Takisawa[4], Yasuo Yasui[1], Takashi Kawai[2], Koichiro Ushijima[2], Masanori Yamasaki[5], Hiroki Saito[6], Ryohei Nakano[1] and Tetsuya Nakazaki[1,*]

[1]*Graduate School of Agriculture, Kyoto University, 4-2-1, Shiroyamadai, Kizugawa City, Kyoto 619-0218, Japan,*
[2]*Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, 1-1-1 Tsushima-naka, Kita-ku, Okayama City 700-8530 Okayama, Japan,*
[3]*Faculty of Agriculture, Kindai University, 3327-204, Nakamachi, Nara City, Nara 631-8505, Japan,*
[4]*Faculty of Agriculture, Ryukoku University, 1-5 Yokotani, Seta Oe-cho, Otsu City, Shiga 520-2194, Japan,*
[5]*Graduate School of Science and Technology, Niigata University, 8050 Ikarashi 2 no-cho, Nishi-ku, Niigata City, Niigata 950-2181, Japan, and*
[6]*Tropical Agriculture Research Front, Japan International Research Center for Agricultural Sciences, 1091-1 Maezato-Kawara-baru, Ishigaki, Okinawa 907-0002, Japan*

Received 20 September 2023; revised 14 January 2024; accepted 14 February 2024; published online 8 March 2024.
*For correspondence (e-mail knishimura@okayama-u.ac.jp; nakazaki.tetsuya.4m@kyoto-u.ac.jp).

SUMMARY

Next-generation sequencing (NGS) library construction often involves using restriction enzymes to decrease genome complexity, enabling versatile polymorphism detection in plants. However, plant leaves frequently contain impurities, such as polyphenols, necessitating DNA purification before enzymatic reactions. To overcome this problem, we developed a PCR-based method for expeditious NGS library preparation, offering flexibility in number of detected polymorphisms. By substituting a segment of the simple sequence repeat sequence in the MIG-seq primer set (MIG-seq being a PCR method enabling library construction with low-quality DNA) with degenerate oligonucleotides, we introduced variability in detectable polymorphisms across various crops. This innovation, named degenerate oligonucleotide primer MIG-seq (dpMIG-seq), enabled a streamlined protocol for constructing dpMIG-seq libraries from unpurified DNA, which was implemented stably in several crop species, including fruit trees. Furthermore, dpMIG-seq facilitated efficient lineage selection in wheat and enabled linkage map construction and quantitative trait loci analysis in tomato, rice, and soybean without necessitating DNA concentration adjustments. These findings underscore the potential of the dpMIG-seq protocol for advancing genetic analyses across diverse plant species.

Keywords: next-generation sequencing library, oligonucleotide, polyphenols, plant leaves, polymerase Chain Reaction, technical advance.

INTRODUCTION

Variations in the genome sequences of plants produce biological diversity. Sequence polymorphisms can be used as genetic markers, regardless of whether they are directly responsible for trait differentiation. Since the development of PCR applications, genetic markers, such as random amplified polymorphic DNA, simple sequence repeats (SSR), and cleaved amplified polymorphic sequences (CAPS), have been used for genetic analysis (Amiteye, 2021; Dietrich et al., 1994; Konieczny & Ausubel, 1993; Williams et al., 1990). In recent years, the rapid development of next-generation sequencing (NGS) technology has made it possible to rapidly identify large numbers of polymorphisms in genomes. To reduce genome complexity, several NGS polymorphism-detection methods have been developed that do not require the entire genome to be read, including restriction site-associated DNA sequencing (RAD-seq), genotyping by sequence (GBS), double-digest RAD-seq (ddRAD-seq), multiplexed inter-simple sequence repeat (ISSR) genotyping by sequencing (MIG-seq), and

ISSR sequencing (Baird et al., 2008; Elshire et al., 2011; Peterson et al., 2012; Sinn et al., 2022; Suyama et al., 2022; Suyama & Matsuki, 2015; Figure 1).

RAD-seq, which was the NGS library construction method developed to reduce genome complexity (Baird et al., 2008), has been used for many years as a method for detecting single nucleotide polymorphisms (SNPs) and other polymorphisms from many individuals via selective sequencing at the loci adjacent to the restriction site. Several RAD-seq-based methods have been developed subsequently (Scheben et al., 2017), among which ddRAD-seq is widely used, especially for population structure analysis, linkage mapping, and genome-wide association studies, because it allows for variation in the number of loci depending on restriction enzyme combinations and provides highly reproducible loci detection (Peterson et al., 2012; Scheben et al., 2017; Shirasawa et al., 2016). The PCR-based NGS library construction method MIG-seq was developed for use in ecological studies and has the advantage of not requiring high-quality DNA, in contrast to ddRAD-seq (Suyama et al., 2022; Suyama & Matsuki, 2015). Using MIG-seq, a higher number of polymorphisms can be obtained for species with large genomes (Nishimura et al., 2022). ISSR sequencing is a method used to amplify

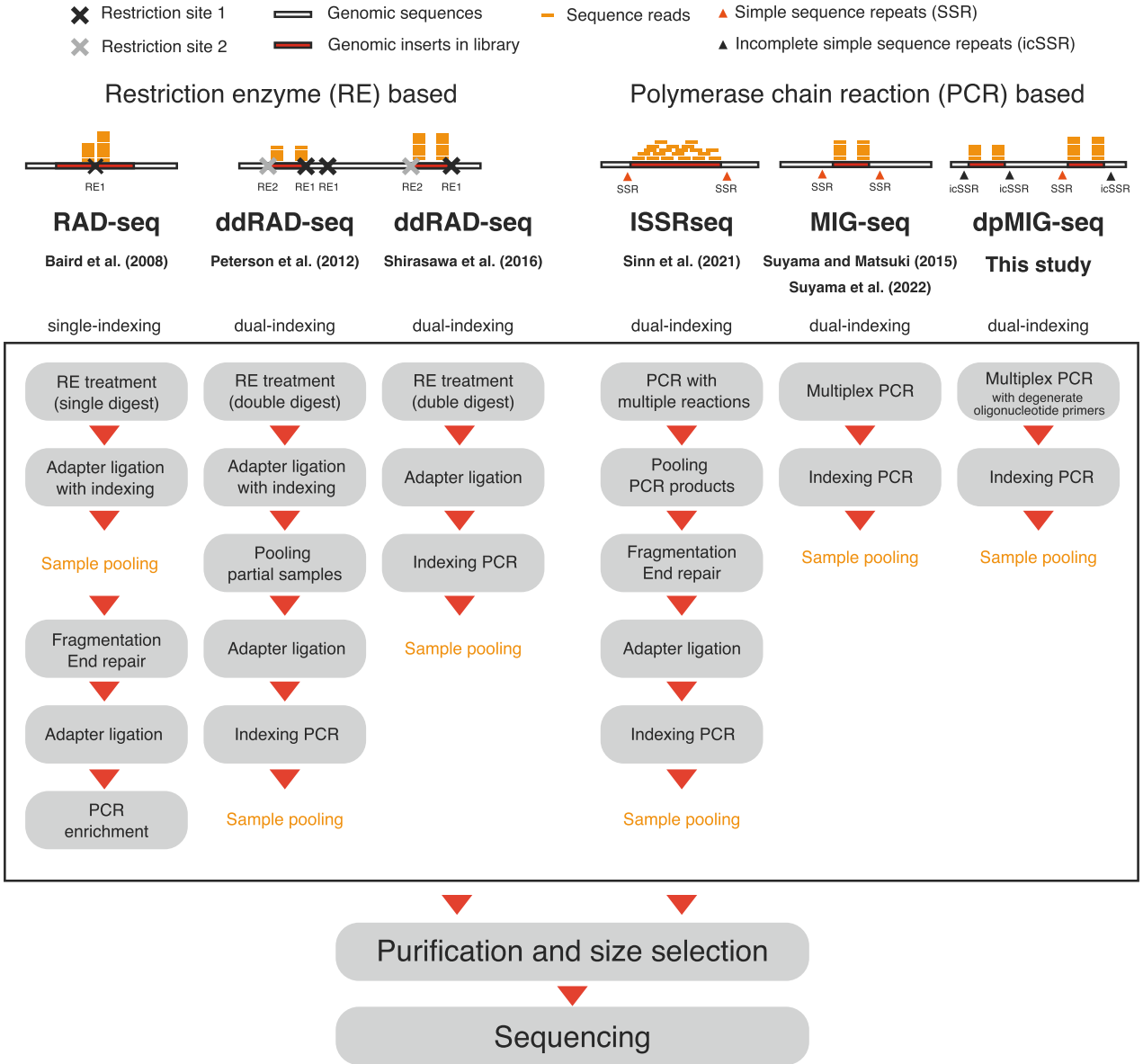


Figure 1. Comparison of next-generation sequencing library construction methods used to reduce genome complexity. The step involving the pooling of all samples is denoted as 'Sample pooling'.

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2024), **118**, 2296–2317

the ISSR region via PCR and is comparable to GBS and RAD-seq in terms of polymorphism detection; however, ISSR sequencing requires an additional fragmentation step during library construction and is more time-consuming than MIG-seq (Sinn et al., 2022).

Using NGS technology, it is now relatively easy to sequence the entire genomes of crop species (Hao et al., 2020; Lv et al., 2020; Ma et al., 2018; Yang et al., 2022). NGS technologies have made it possible to repeatedly analyze the relationship between genome information and various phenotypes in lines for which the genotypes can be maintained using cloning techniques (e.g., grafting and cuttings) and for which the genomes are completely fixed and can be maintained via self-propagation (Kajiya-Kanegae et al., 2021; Kumar et al., 2020; Tanaka et al., 2020, 2021).

In plant, agriculture, and breeding studies, genetic analysis is often performed using non-immortal lines, such as the F_2 and BC_1 populations, to which applying whole genome sequencing to obtain genotypes for one generation, when genotype maintenance via cloning is impossible or expensive, is costly and considered unfeasible. Thus, methods for constructing NGS libraries with reduced genome complexity, that is, ddRAD-seq and MIG-seq, are useful because of their low cost and application when whole genome information is not required for genetic analysis. In terms of their disadvantages, ddRAD-seq requires restriction enzyme processing in the first step and a certain amount of purified DNA, since plants and crops contain many impurities in their leaves that inhibit restriction enzyme reaction. The advantage of MIG-seq, which uses PCR as the first reaction in library construction and can produce libraries from small amounts of less clean DNA, can compensate for the disadvantages of ddRAD-seq, but MIG-seq produces fewer polymorphisms in species with small genome sizes, making it difficult to conduct genetic analysis in a flexible manner. Therefore, methods must be developed to overcome these disadvantages, such as PCR-based methods that allow the detectable number of polymorphisms to be freely adjusted.

Suyama et al. (2022) enhanced the MIG-seq protocol by lowering the annealing temperature from 48°C to 38°C, enabling more extensive annealing to various genomic regions. Building upon this concept, we aimed to create a method capable of amplifying genome-wide loci with variable polymorphism counts, akin to ddRAD-seq. Therefore, in the present study, we attempted to increase the number of sequenced regions in MIG-seq by introducing degenerate oligonucleotides (Ns: A, T, G, C) to certain primer positions (excluding the 3′ end) during the initial PCR (Figure 1). This approach, named degenerate oligonucleotide primer MIG-seq (dpMIG-seq; Figure 1), was comprehensively assessed. Initially, we examined the variation in sequenced loci in tomato (*Solanum lycopersicum* L.) using MIG-seq primers containing degenerate oligonucleotides at various positions. Subsequently, we assessed the applicability of dpMIG-seq across 11 crops with diverse genome sizes. We further examine the method's utility in the context of tetraploid wheat (*Triticum turgidum* L.), which has a substantial genome, assessing its effectiveness in selecting and validating near-isogenic lines (NILs). Additionally, we evaluated the advantages of dpMIG-seq over restriction enzyme-based protocols by investigating its compatibility with unpurified DNA for library generation. Finally, we conducted linkage mapping and quantitative trait loci (QTL) analysis in populations of rice (*Oryza sativa* L.), tomato, and soy (*Glycine max* L.) to demonstrate the efficacy of dpMIG-seq in advancing genetic analyses of crop plants.

RESULTS

Performance evaluation of dpMIG-seq using tomato

To validate dpMIG-seq, we developed 13 primer sets that included incomplete SSR regions by introducing degenerate oligonucleotides (N) into the SSR part of the MIG-seq primers originally developed by Suyama and Matsuki (2015) (Table 1). We also designed another MIG-seq primer set (which we named 'PS2') targeting SSRs that are not targeted by the MIG-seq primers developed by Suyama and Matsuki (2015) (Table 1; Table S1). The PS2 primers were used to determine whether replacing some of the bases with N in the SSR region of the primers had a generalized effect on the number of detectable polymorphisms. The primer sets for dpMIG-seq were named by combining (i) the original primer set name, (ii) an underscore, and (iii) the position at which the base was substituted with N. If two or more degenerate oligonucleotides were present, a range was indicated, for example, '3–4'. To compare the performance of each primer set in MIG-seq and dpMIG-seq, libraries were initially constructed and sequenced using two tomato cultivars: 'MPK-1' and 'Micro-Tom' (Tables S2 and S3). To ensure the accuracy of comparisons between the results of primer sets, the coverage depth (DP) was corrected to a value per 0.3 gigabase (Gb), and polymorphisms/regions with a DP of <10 were excluded from the analysis. The results for 17 primer sets are shown in Figure 2(a). The number of polymorphisms increased when only a single base N was introduced, regardless of the position of the base in the primer set used. In addition, the highest increase in the number of detectable polymorphisms was achieved when the degenerate oligonucleotides were introduced at the fourth and fifth bases from the 3′ end of the primers. The number of detectable polymorphisms also increased in the PS2 primers when the degenerate oligonucleotides were introduced at the fourth and fifth nucleotides. Conversely, when four or more degenerate oligonucleotides were introduced (e.g., PS1_4–

Table 1 Examples of the primers used for MIG-seq and dpMIG-seq

Primer set name	Primer name	Sequences (5′–3′)
PS1	ACT4TG-f	CGCTCTTCCGATCTCTGACTACTACTACTTG
PS1_3	ACT4TG-f_3	CGCTCTTCCGATCTCTGACTACTACTAC**N**TG
PS1_4	ACT4TG-f_4	CGCTCTTCCGATCTCTGACTACTACTA**N**TTG
PS1_5	ACT4TG-f_5	CGCTCTTCCGATCTCTGACTACTACT**N**CTTG
PS1_7	ACT4TG-f_7	CGCTCTTCCGATCTCTGACTACTA**N**TACTTG
PS1_10	ACT4TG-f_10	CGCTCTTCCGATCTCTGACTA**N**TACTACTTG
PS1_14	ACT4TG-f_14	CGCTCTTCCGATCTCTG**N**CTACTACTACTTG
PS1_3–4	ACT4TG-f_3–4	CGCTCTTCCGATCTCTGACTACTACTA**NN**TG
PS1_3–5	ACT4TG-f_3–5	CGCTCTTCCGATCTCTGACTACTACT**NNN**TG
PS1_4–5	ACT4TG-f_4–5	CGCTCTTCCGATCTCTGACTACTACT**NN**TTG
PS1_4–6	ACT4TG-f_4–6	CGCTCTTCCGATCTCTGACTACTAC**NNN**TTG
PS1_4–7	ACT4TG-f_4–7	CGCTCTTCCGATCTCTGACTACTA**NNNN**TTG
PS1_4–8	ACT4TG-f_4–8	CGCTCTTCCGATCTCTGACTACT**NNNNN**TTG
PS1_4–9	ACT4TG-f_4–9	CGCTCTTCCGATCTCTGACTAC**NNNNNN**TTG
PS2	TCA4AG-f	CGCTCTTCCGATCTCTGTCATCATCATCAAG
PS2_4	TCA4AG-f_4	CGCTCTTCCGATCTCTGTCATCATCAT**N**AAG
PS2_4–5	TCA4AG-f_4–5	CGCTCTTCCGATCTCTGTCATCATCA**NN**AAG

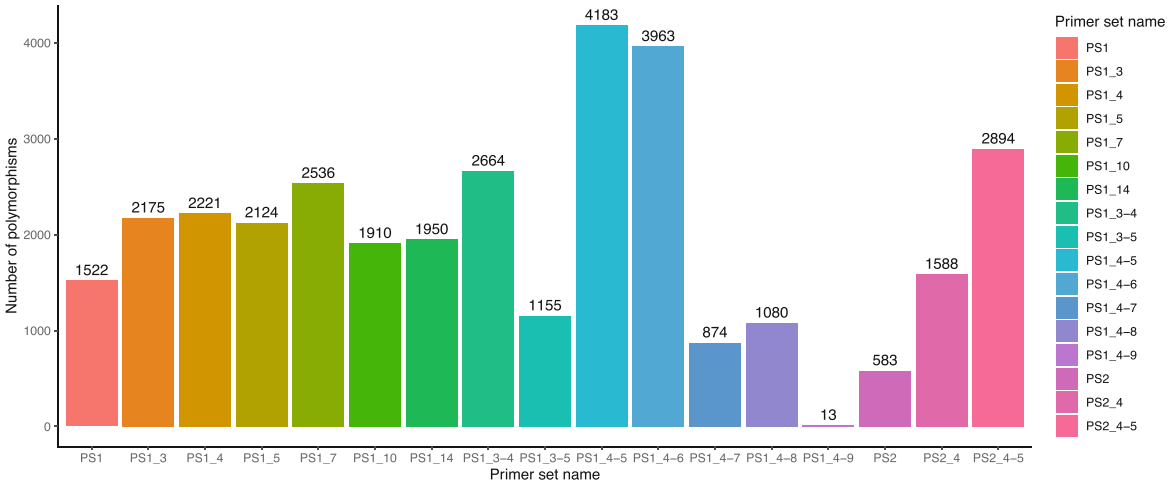
The primers used in degenerate oligonucleotide primer MIG-seq (dpMIG-seq) were the 16 multiplex primers reported by Suyama and Matsuki (2015) (named 'PS1' in the present study) or the 16 multiplex primers for MIG-seq developed in the present study (named 'PS2'), and the bases in the simple sequence repeats (SSR) region were changed to a degenerate oligonucleotide 'N' (A, T, G, or C). Only one of each multiplex primer is exemplified in this table; information on all primers is summarized in Table S1. Underlining indicates the SSR regions. Degenerate oligonucleotides are shown by a bold **N**. Primers are named according to Suyama and Matsuki (2015) in the order of SSR base, number of repeats, and anchor sequence at the 3′ end, followed by the position of N indicated by the underlining, that is, the position from the 3′ end. If more than one N is present, it is denoted by a range, for example, '3–4' in PS1_3–4.

7, PS1_4–8, and PS1_4–9), the number of polymorphisms decreased, especially in the case of the bases of PS1_4–9, for which almost no polymorphisms were detected. Mapping rates were calculated for each primer set of 'MPK-1' and 'Micro-Tom', and the average mapping rates of these cultivars were compared along with the number of polymorphisms. Interestingly, primer sets with lower mapping rates exhibited correspondingly lower numbers of polymorphisms (Figure 2b). Specifically, the mapping rate of PS1_4–9, which had the fewest other polymorphisms, was 5.856%, indicating that almost no reads were mapped. Electrophoretic images of the first PCR products of dpMIG-seq:PS1_4–9 showed limited amplification (Figure S1). Even in primer sets with relatively high mapping rates, variability in the number of polymorphisms was observed.

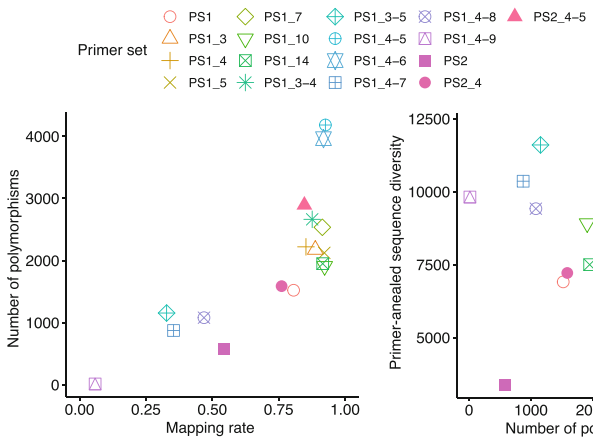
Subsequently, 14-base sequences representing the expected annealing sites in the first PCR were isolated, referring to the tomato reference genome for each read. The extraction frequency per million for each primer set was computed, and the diversity of 14-bp sequences appearing more than 10 times (referred to as 'primer-annealed sequence diversity' in this study) was compared with the count of detected polymorphisms for each primer set. Notably, dpMIG-seq exhibited a higher primer-annealed sequence diversity compared with MIG-seq (Figure 2c), indicating that the first PCR primer in dpMIG-seq annealed to a more diverse range of nucleotide sequences relative to that in MIG-seq. Primer sets with more detectable polymorphisms typically exhibited increased primer-annealed sequence diversity, except for PS1_3–5, PS1_4–7, PS1_4–8, and PS1_4–9, which displayed fewer 'detected polymorphisms, despite their higher primer-annealed sequence diversity (Figure 2c). This discrepancy is attributed to the low mapping rate associated with these primer sets (Figure 2b). Notably, dpMIG-seq primers consistently displayed a high frequency of

Figure 2. Evaluation of degenerate oligonucleotide primer MIG-seq (dpMIG-seq) performance. (a) Comparison of detectable polymorphisms between MIG-seq and dpMIG-seq in tomato, showing the number of polymorphisms with a data volume of 0.3 Gb and a coverage depth ≥10. (b) Mapping rates of MIG-seq and dpMIG-seq derived reads in tomato, and the number of polymorphisms in each primer set. (c) Relationship between primer-annealed sequence diversity and polymorphism count in tomato. Vertical axis represents the diversity of the sequences to which the primers annealed in the first PCR, whereas the horizontal axis shows the number of polymorphisms. (d) Ratio of common polymorphisms to total detected polymorphisms for each two primer set combination in tomato. (e) Error rate for different coverage depth thresholds in tomato variety 'MPK-1'. Error bars indicate standard deviation for three combinations (rep.1 versus rep. 2, rep. 2 versus rep. 3, and rep. 3 versus rep. 1). (f) Comparison of stably sequenced total base length between MIG-seq and dpMIG-seq in 11 crop species, with the vertical axis showing the number of bases to which reads were mapped with a coverage depth ≥10, and the horizontal axis representing the logarithm base 2 of genome size.

(a)

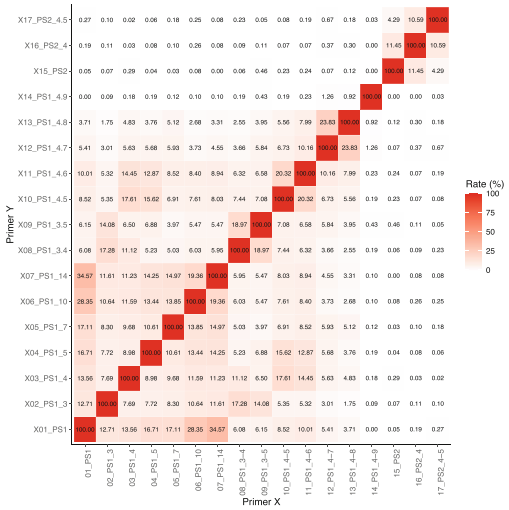


(b)

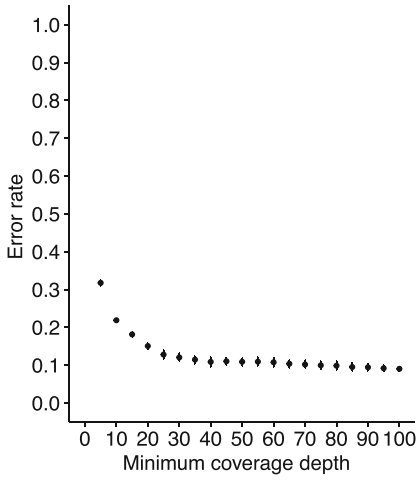


(c)

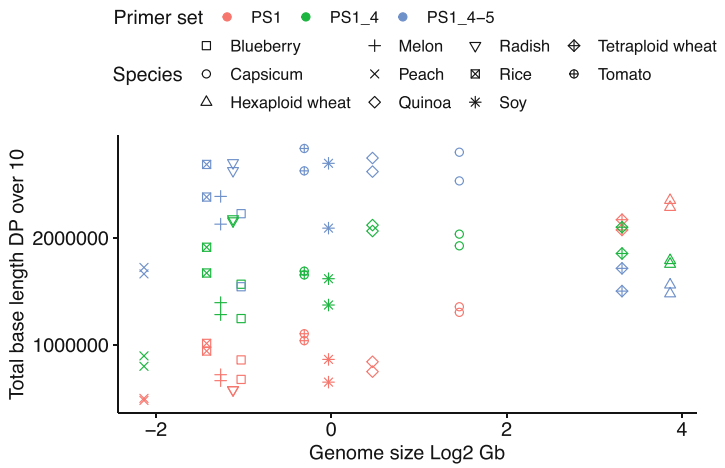
(d)



(e)



(f)



annealing to non-SSR sequences across all primer sets, a characteristic shared with MIG-seq (MIG-seq:PS1), where the most common sequence was an incomplete SSR (Data S1).

Next, the percentage of polymorphisms detectable using each primer set that were common between two primer sets was calculated. We found that many primer sets were able to sequence different loci (Figure 2d). For example, 34.57% was the highest percentage of common loci between PS1 and PS1_14 (Figure 2d).

To assess the error rate of dpMIG-seq, three library construction and sequencing runs were performed on tomato dpMIG-seq:PS1_4–5 for both 'Micro-Tom' and 'MPK-1'. Additionally, error rates (discrepancy rates between two iterations) were calculated. For replicates 1, 2, and 3, 'MPK-1' reads numbered 2 376 816, 4 107 942, and 3 626 666, whereas 'Micro-Tom' reads numbered 3 499 510, 2 257 316, and 3 050 984, respectively. With an increase in the coverage depth threshold, the error rate decreased (Figure 2e; Figure S2). Even with a DP threshold of 100, 'Micro-Tom' and 'MPK-1' genotype data contained averages of 7.987% and 9.035% errors, respectively (Figure S2). Given that increasing the coverage depth threshold can reduce the total number of polymorphisms, there was a need to appropriately establish filtering criteria, including minimum coverage depth, while considering the trade-off between error rate and polymorphism count (Figure S2).

Finally, the reproducibility of the first PCR amplification between the two tomato cultivars and the redundancy of amplification were investigated. Paired-end mapped read locations were identified, and their frequencies were converted to the number per million of paired-end mapped reads, with those exceeding 5 (defined as the frequency of an amplicon-mapped position) being extracted and compared between 'MPK-1' and 'Micro-Tom'. For all 17 primer sets, the frequency of amplicon-mapped positions between the two cultivars was significantly correlated ($P < 0.001$), indicating reproducible amplification efficiency for each locus in the first PCR (Figure S3). However, substantial variability in the frequency of amplicon-mapped positions was observed. For example, certain amplicons appeared over 50 000 times per million ('Micro-Tom' MIG-seq:PS2 data). Thus, some loci were over-sequenced when using MIG-seq or dpMIG-seq data.

Comparison of the number of loci that can be sequenced using MIG-seq and dpMIG-seq in 11 crop species

The results for tomato indicated that dpMIG-seq:PS1_4 and dpMIG-seq:PS1_4–5 were able to detect more polymorphisms than were detected by MIG-seq:PS1; thus, we used these primer sets to evaluate dpMIG-seq performance in other crop species. MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq:PS1_4–5 were used to perform dpMIG-seq on peach (*Prunus persica* L.), rice, melon (*Cucumis melo* L.), radish (*Raphanus sativas* L.), blueberry (*Vaccinium* spp.), soy, quinoa (*Chenopodium quinoa* Willd.), capsicum (*Capsicum* spp.), tetraploid wheat (*T. turgidum* L.), and hexaploid wheat (*Triticum aestivum* L.). Two genotypes were analyzed in each species (Tables S2 and S4). When the total base length with a DP ≥10 was compared for each method and each species after correcting the raw read data volume to 0.3 Gb, it was clear that the number of regions that could be sequenced by dpMIG-seq:PS1_4–5 was increased in all species except the wheat species (Figure 2f). Genome coverage by dpMIG-seq:PS1_4–5 was largest and smallest for peach (0.75%) and hexaploid wheat (0.01%), respectively (Table S5). These results confirm that an increase in the number of detectable polymorphisms is also observed in species with relatively small genomes, indicating the versatility of dpMIG-seq. However, in wheat, when the data volume was corrected to 0.3 Gb, the region sequenced by dpMIG-seq:PS1_4 and dpMIG-seq:PS1_4–5 did not increase relative to that sequenced by MIG-seq:PS1 (Figure 2f). We also found an increase in detectable polymorphisms in most crop species except those of wheat (Figure S4). In peach, rice, melon, blueberry, soy, radish, quinoa, and capsicum, dpMIG-seq:PS1_4–5 provided the highest number of polymorphisms, followed by dpMIG-seq:PS1_4 and then by MIG-seq. In addition, many polymorphisms obtained by dpMIG-seq:PS1_4–5 and dpMIG-seq:PS1_4 were derived from regions that differed from those sequenced by MIG-seq:PS1 in all 11 crop species (Figure S5). These results validate the use of dpMIG-seq:PS1_4 and dpMIG-seq:PS1_4–5 in a range of genome sizes, for example, from peach (approximately 0.227 Gb) to capsicum (approximately 2.753 Gb).

Analysis of GC content and genomic distribution of sequences amplified through first PCR of MIG-seq and dpMIG-seq

Genomic regions with elevated GC content generally harbor more protein-coding sequences. Additionally, such regions pose increased the difficulty for PCR amplification compared with regions with lower GC content. To investigate potential bias introduced by GC content in the first PCR amplification, we compared the GC content of amplicons from MIG-seq and dpMIG-seq with that of sequences randomly extracted from the reference genome. The GC content of the amplicons was calculated from sequences appearing at least five times per million mapped paired-end reads in both cultivars of each species by extracting sequences considered amplified by the first PCR from the paired-end mapped reads' positions on the reference genome. Given the peak around a GC content value of 0 in randomly extracted sequences, only those with a GC content of ≥5 were included in the analysis. For tomato MIG-seq and dpMIG-seq data, the average GC content of mapped amplicons was significantly higher for all 17 primer

sets than that of sequences randomly extracted from the tomato reference genome (Figure S6). Similar analyses of data from 10 different crops, obtained using MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq:PS1_4–5, yielded consistent results across all three primer sets in all species (Figure S7). These findings indicate that the first PCR of MIG-seq and dpMIG-seq may not exhibit a preferential bias toward amplifying sequences with low GC content.

To assess the genomic bias of the mapped reads, for MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq:PS1_4–5 data from 11 crops, paired-end reads meeting the GC content criteria were mapped to physical positions and visualized. Based on the physical positions of mapped paired-end reads, the first PCR of MIG-seq and dpMIG-seq was found to amplify genomic sequences across all chromosomes in all 11 crop species (Figure S8). Calculation of the maximum gap between amplicons revealed a 3 253 561 bp gap in MIG-seq:PS1 for blueberry, equivalent to 0.708% of the blueberry genome size (Table S6). Although bias was not statistically evaluated, we observed amplification from various regions in the genome via the first PCR.

Amount of data required for the effective use of dpMIG-seq in wheat

We hypothesized that a greater amount of raw read data than 0.3 Gb was required to sequence more loci in wheat using dpMIG-seq. As dpMIG-seq:PS1-4-5 sequenced more loci than dpMIG-seq:PS1_4 in numerous crops, this analysis exclusively compared dpMIG-seq:PS1_4–5 and MIG-seq:PS1. To determine the appropriate amount of data required for wheat genotyping with dpMIG-seq:PS1_4–5, we used data with >5.5 Gb of raw reads for each of two tetraploid wheat samples: the durum wheat cultivar 'Setodur' (*T. turgidum* L. ssp. *durum*) and the emmer wheat accession 'TN26' (*T. turgidum* L. ssp. *dicoccum*; Table S7). In our analysis of tetraploid wheat, dpMIG-seq:PS1_4–5 had a higher number of polymorphisms than that of MIG-seq:PS1 if around ≥2 Gb of data were obtained (Figure 3a). Even with the amount of raw reads obtained in this study, the number of polymorphisms obtained using MIG-seq:PS1 and dpMIG-seq:PS1_4–5 was not completely saturated. However, the slope of MIG-seq:PS1 began to decline as the number of raw reads increased, whereas the slope of dpMIG-seq:PS1_4–5 continued to increase at a constant rate (Figure 3a). Similar to the results for other crop species, many of the polymorphisms obtained by MIG-seq (PS1) and dpMIG-seq:PS1_4–5 were not common when the data volume was corrected to 2.0 Gb (Figure 3b).

Increased efficiency of selecting wheat NILs using MIG-seq and dpMIG-seq

Because dpMIG-seq makes it possible to vary the detectable number of polymorphisms in crops, including wheat, we investigated whether wheat NILs can be efficiently selected using different primer sets depending on the number of polymorphisms required in several experimental steps. In wheat, MIG-seq:PS1 can also detect a certain number of polymorphisms. Thus, we used dpMIG-seq:PS1_4–5 when a higher number of polymorphisms were needed for the analysis. In this experiment, we replaced a late-heading allele of *VRN-A3*, which is a wheat ortholog of *FLOWERING LOCUS T* in Arabidopsis (Nishimura et al., 2018, 2021; Yan et al., 2006) in 'Setodur', with an early-heading allele of *VRN-A3* in 'TN26'. F_1 individuals of 'Setodur' and 'TN26' were backcrossed with 'Setodur', and this step was repeated four times while confirming the heterozygosity at the *VRN-A3* locus. To confirm the allelic composition of each chromosome during the selection process, MIG-seq:PS1 was applied to BC_4F_2 individuals ($n = 16$), which were selfed BC_4 individuals with heterozygous *VRN-A3*. To validate the residual heterozygous regions on each chromosome, the segregation ratio of each polymorphism (1:2:1) was evaluated using a chi-squared test. The results indicated that four regions on chromosomes 1A, 1B, 3A, and 7A exhibited segregation with high *P*-values ($P > 0.5$; Figure S9, Data S2). From this BC_4F_2 population, we selected BC4F2_10, an individual with fewer heterozygous loci, no region fixed to the TN26-type, and a heterozygous allele of *VRN-A3*. Two CAPS markers were designed from the polymorphism information obtained using dpMIG-seq:PS1_4–5 in the heterozygous region of BC4F2_10 on chromosome 1B. Using the CAPS markers and the genetic marker for *VRN-A3*, we selected an individual from the BC_4F_3 generation ($n = 39$), namely BC4F3_32, presumed to be heterozygous only for the region near *VRN-A3*. To confirm the allele constitution of a selected NIL candidate, both MIG-seq:PS1 and dpMIG-seq:PS1_4–5 were applied to BC4F3_32. The number of raw reads was 4 976 366 and 17 607 262 when MIG-seq:PS1 and dpMIG-seq:PS1_4–5 were applied, yielding 9726 and 23 183 markers, respectively. Regions other than *VRN-A3* were mostly fixed, but heterozygous or TN26-type homozygous genotypes were detected in some regions, although only the region near *VRN-A3* was consistently found to be heterozygous via both MIG-seq:PS1 and dpMIG-seq:PS1_4–5 (Figure 3c; Data S3). The combined use of MIG-seq and dpMIG-seq yields a genotype resolution superior to each method used individually, as it can fill in gaps left by the use of MIG-seq alone. For example, the most extensive gap in MIG-seq:PS1 is 3B:79 078 439–120 553 067, with a length of 41 474 628. In this gap, 23 dpMIG-seq:PS1_4-5–derived polymorphisms were identified (Data S3). This occurrence may be attributed to the ability of MIG-seq and dpMIG-seq to sequence slightly different loci. The self-pollination of BC4F2_10 enabled the establishment of NILs for *VRN-A3*. Consequently, we demonstrated the efficient generation of

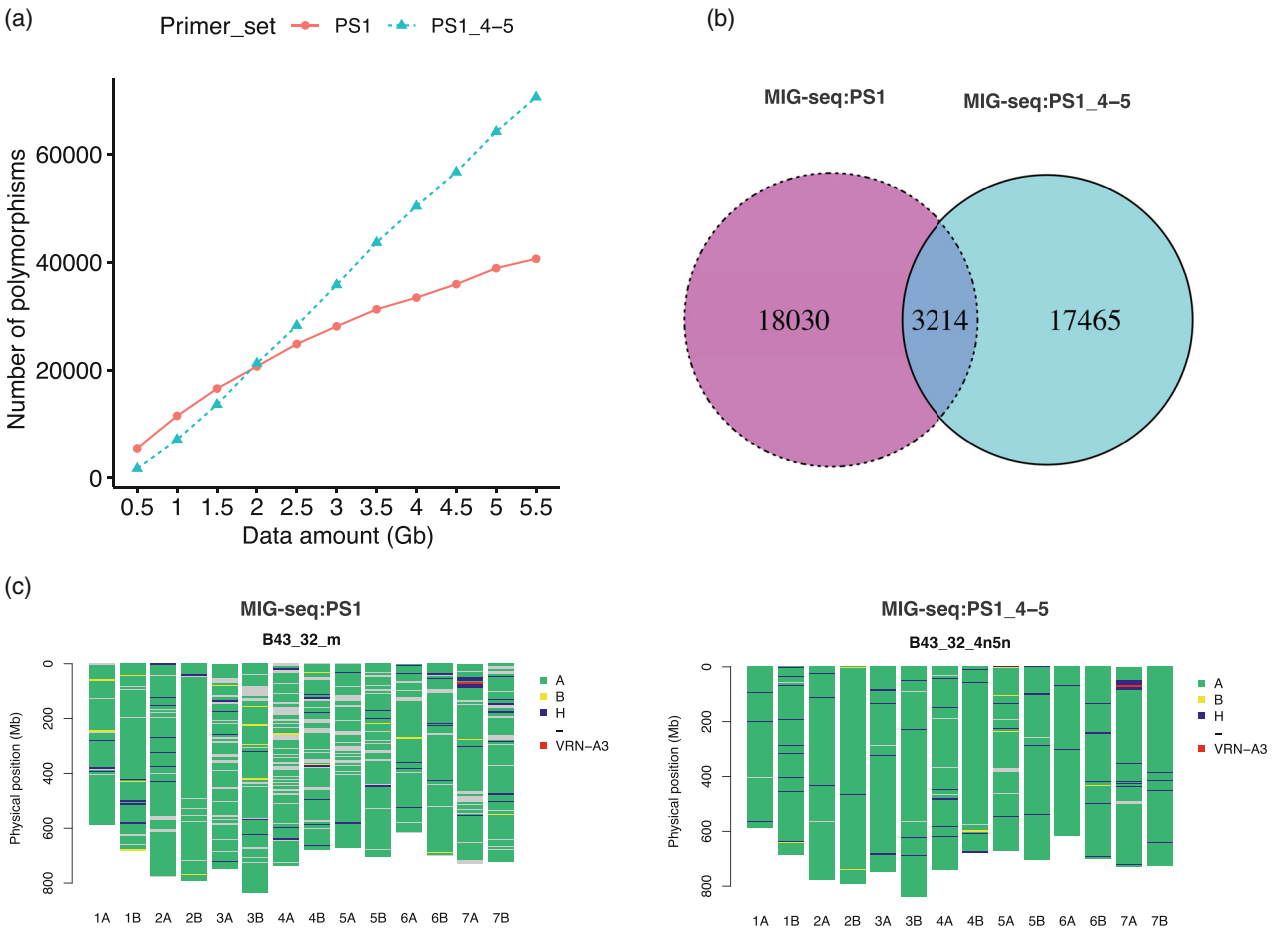


Figure 3. Applicability of degenerate oligonucleotide primer MIG-seq (dpMIG-seq) to tetraploid wheat, and graphical genotyping of wheat near-isogenic lines (NILs).
(a) Relationship between detectable polymorphisms via MIG-seq:PS1 and dpMIG-seq:PS1_4–5 and the raw read data volume, with the vertical axis indicating the number of polymorphisms between the two varieties, and the horizontal axis showing the raw read data amount. (b) Overlapping polymorphisms detected via MIG-seq and dpMIG-seq in tetraploid wheat, with a Venn diagram of polymorphisms detected via MIG-seq:PS1 and dpMIG-seq:PS1_4–5 with 2 Gb of data. (c) Graphical genotypes of tetraploid wheat NILs obtained via MIG-seq:PS1 and dpMIG-seq:PS1_4–5, where A, B, and H represent a recurrent parent (Setodur genotype), a donor parent (TN26 genotype), and the heterozygous genotype, respectively.

NILs using different primer sets depending on the situation: MIG-seq:PS1 for preliminary selection of individuals, dpMIG-seq:PS1_4–5 for detecting a higher number of polymorphisms for flexible CAPS marker design, and a combination of MIG-seq:PS1 and dpMIG-seq:PS1_4–5 for confirmation of higher resolution, surpassing the resolution achieved by MIG-seq alone, where the genetic background of the selected NILs is almost entirely replaced by the genetic background of the recurrent parent.

Construction of a dpMIG-seq library using unpurified DNA

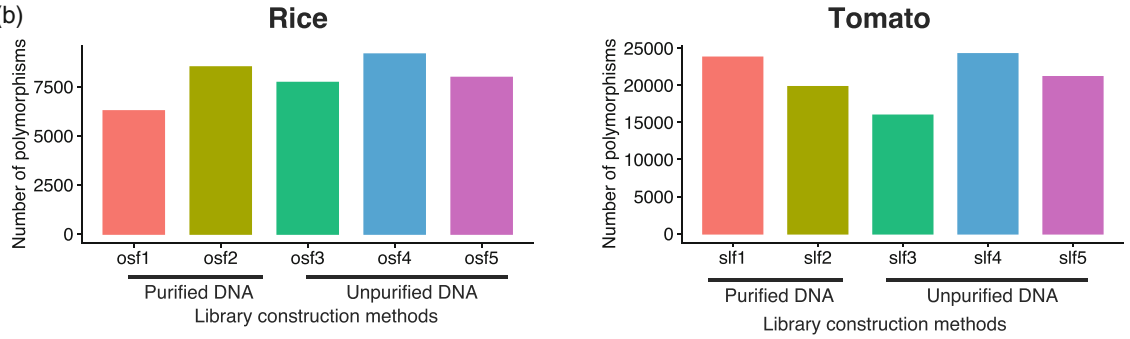
To develop a protocol for constructing libraries from unpurified DNA and validate the outcomes, we conducted the first PCR using a variety of combinations involving four DNA extraction methods and two different enzymes (Figure 4a; Figure S10). These methods encompassed the following: (i) purified DNA, (ii) the filter paper technique outlined by Jia et al. (2021), (iii) DNA extraction using the buffer detailed by Thomson and Henry (1995), and (iv)

Figure 4. Comparison of degenerate oligonucleotide primer MIG-seq (dpMIG-seq) library construction methods. (a) Comparison of each library's name, DNA extraction method, and the first PCR enzyme. (b) Comparison of detected polymorphism numbers among library construction methods, with the vertical axis showing the number of polymorphisms without loss in eight rice or eight tomato lines. (c) Comparison of mapping rates to reference genomes among library construction methods, with boxplots of five rice and five tomato libraries shown. (d) Comparison of principal component analysis using polymorphisms obtained from purified and unpurified DNA-derived libraries, showing results for osf01 and slf01 for purified DNA and osf05 and slf05 for unpurified DNA. (e) Mapping rates for libraries created using unpurified DNA.

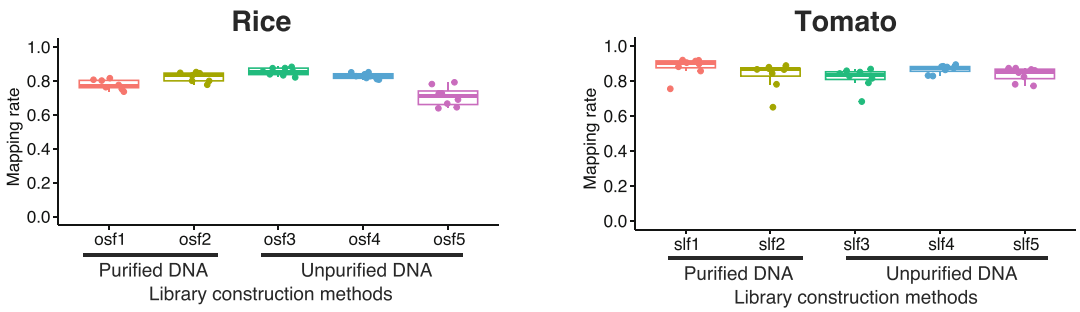
(a)

Library name	osf1 slf1	osf2 slf2	osf3 slf3	osf4 slf4	osf5 slf5
DNA extraction method	Maxwell® RSC Plant DNA Kit (Promega ,USA)	Maxwell® RSC Plant DNA Kit (Promega ,USA)	Filter paper based Jia et al. (2021)	One step buffer Thomas and Henry (1995)	One step buffer Thomas and Henry (1995) with PVPP
PCR enzyme	Multiplex PCR Assay Kit Ver.2 (Takara, Japan)	KOD one (Toyobo, Japan)	KOD one (Toyobo, Japan)	KOD one (Toyobo, Japan)	KOD one (Toyobo, Japan)

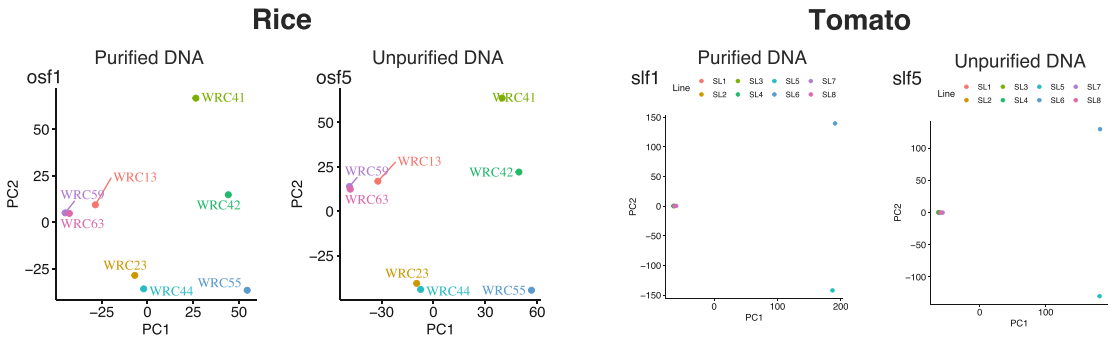
(b)



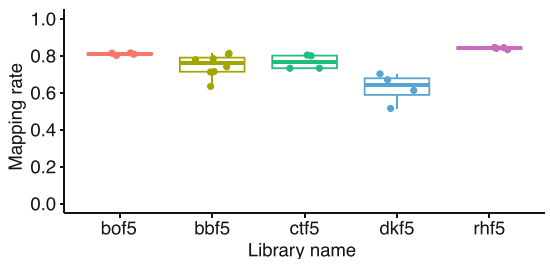
(c)



(d)



(e)



the buffer of Thomson and Henry (1995) with polyvinylpolypyrrolidone (PVPP) added to adsorb polyphenols. Two kinds of PCR enzyme were used: (i) that in the Multiplex PCR Assay Kit ver2 (Takara Co. Ltd., Kusatsu, Japan), previously used for MIG-seq, and (ii) that in KOD One PCR Master Mix (Toyobo Co. Ltd., Osaka, Japan). In both rice and tomato, the first PCR amplification was extensively verified, yet the filter paper method led to unstable amplification when used with the Multiplex PCR Assay Kit ver2 (Figure S10).

From the first PCR products, characterized by smears in Figure S10, we successfully constructed five distinct libraries of eight rice and eight tomato dpMIG-seq:PS1_4–5 (Tables S8 and S9). Figure 4(a) shows the library names, sequence analysis locations, and DNA extraction method and PCR enzyme combination (osf01–05 in rice and slf01–05 in tomato). The number of obtained polymorphisms remained relatively consistent across all methods (Figure 4b). Although mapping rates exhibited slight variations among methods and species, on average, >70% of reads consistently aligned to the reference genome across all libraries (Figure 4c; Table S10). Principal component analysis (PCA) analysis, based on the obtained polymorphisms for osf01, osf05, slf01, and slf05, yielded similar results among methods in both rice and tomato (Figure 4d).

Similarly, we explored combinations of DNA extraction method and PCR enzyme for eight blueberry cultivars, a fruit tree known for relatively high polyphenol content in its leaves. Stable first PCR amplification was only achieved when PVPP was introduced into the KCl buffer (Figure S11). This buffer's effectiveness was confirmed using four citrus (*Citrus* spp.), four cabbage (*Brassica oleracea* var. *capitata*), four persimmon (*Diospyros kaki* L.), and four rose (*Rosa hybrida* L.) varieties as plant materials, which all exhibited stable first PCR amplification (Figure S12, Tables S8 and S9). The mapping rate to the reference genome was slightly lower for persimmon, averaging 62.45%, yet exceeded 70% for the remaining species (Figure 4e; Table S10), validating the species origin of the obtained reads.

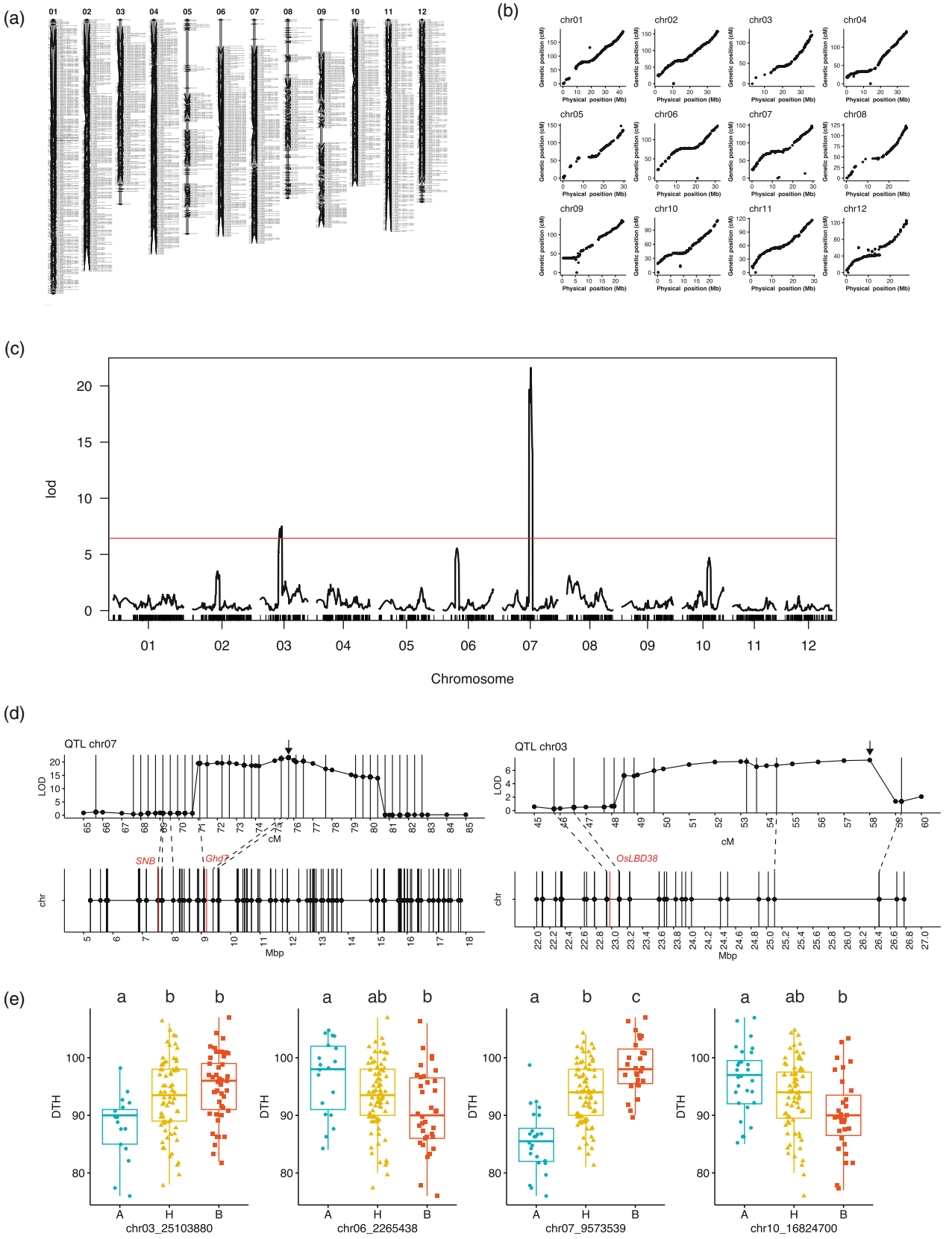
Linkage map construction and QTL analysis of the rice F_2 population using dpMIG-seq

To validate the effectiveness of dpMIG-seq for genetic analysis in rice with a relatively small genome, dpMIG-seq:PS1_4–5 was used with the F_2 population derived from a cross between the japonica rice cultivars 'Takanari' (TK: *O. sativa* L. ssp. *indica*) and 'Taichung 65' (T65: *O. sativa* L. ssp. *japonica*) to construct a linkage map. Using this linkage map, we performed QTL analysis for 'days from sowing to heading' (DTH). For TK and T65, 7 983 772 and 13 130 604 raw reads were obtained, respectively. For the 130 individuals of the F_2 population, 425 171 856 raw reads were obtained in total, with an average of 3 270 553 reads per individual, using dpMIG-seq:PS1_4–5 (Table S11). A linkage map of 1299 markers, consisting of 5048 SNPs/indels, was generated (Figure 5a; Table S12). The total distance of the linkage map was 1661.2 cM, and the average distance between two markers in the linkage map was 1.28 cM (Table S12). Although the marker order of the linkage map did not correspond perfectly with the physical and genetic maps for some markers (Figure 5b), genome-wide markers were obtained for each marker in the rice reference genome (Figure S13a, Data S4).

The mean DTH values of TK and T65 were 91.9 and 91.2, respectively, and the DTH of the F_2 population was 76–107, indicating transgressive segregation (Figure S13b). QTL analysis for DTH revealed two QTLs that had peaks with LOD scores of 7.940 and 23.186 on chromosomes 3 and 7, respectively, when the LOD score threshold was 6.640 after 1000 permutations (Figure 5c). The rice flowering suppressor gene *Ghd7* (Xue et al., 2008) was located near the peak of the QTL on chromosome 7 (Figure 5d), whereas no genes known to affect heading were located near the peak of the QTL on chromosome 3 (Figure 5d). The LOD score threshold was not exceeded, but peaks with LOD scores of 5.593 and 5.035 were observed on chromosomes 6 and 10, respectively.

The peak of chromosome 6 is located within a dense cluster of rice flowering genes, potentially corresponding to *Hd17* (Matsubara et al., 2012), *RFT1*, or *Hd3a* (Komiya et al., 2008; Figure S13c). Similarly, the peak on chromosome 10 may correspond to *Ehd1* (Doi et al., 2004; Figure S13c). A Tukey–Kramer comparison, analyzing the difference in mean DTH values between population alleles using markers closest to the position of the LOD peak, revealed significant differences between the alleles of both parental homozygotes for all four markers (Figure 5e).

Figure 5. Construction of linkage map and quantitative trait loci (QTL) analysis for days to heading in rice F_2 populations via degenerate oligonucleotide primer MIG-seq (dpMIG-seq). (a) Linkage map of the rice F_2 population constructed via dpMIG-seq, where the rice F_2 population originated from a cross between 'Takanari' and 'Taichung 65'. (b) Location of each marker on the physical and linkage maps, with the *x*-axis and *y*-axis representing the position on the physical map and genetic map, respectively. (c) LOD score for QTL analysis based on days to heading, where the red line indicates the LOD score threshold. (d) Comparison of locus positions and candidate genes on physical and genetic maps of QTLs on chromosomes 3 and 7 for the heading date of the rice F_2 population, with dotted lines denoting the same markers on the physical map and genetic map, and marker correspondence by dotted lines only pertaining to the vicinities of the QTL peak or known flowering genes. (e) QTL effect for days from transplant to heading in the rice F_2 population derived from a cross between 'Takanari' and 'Taichung 65'. Days from transplant to heading are categorized by markers close to the detected QTL peaks or LOD score peaks, with A, B, and H indicating the homozygous 'Taichung 65' allele, homozygous 'Takanari' allele, and heterozygous allele, respectively. Different lowercase letters indicate significant differences.



Re-analysis by dpMIG-seq of populations analyzed by tomato genotype array

We aimed to confirm that dpMIG-seq could yield results similar to those obtained through genotyping arrays. We used DNA from the tomato F_2 population, for which a tomato linkage map had been previously established via genotyping array and QTL analysis to gauge parthenocarpic levels for library construction (Takisawa et al., 2018). This F_2 population was derived from a cross between the parthenocarpic tomato cultivar 'MPK-1' and the tomato cultivar 'Micro-Tom'. We leveraged phenotypic data regarding parthenocarpic levels, as used by Takisawa et al. (2018) (Figure S14a). Subsequently, a dpMIG-seq:PS1_4–5 library was constructed and sequenced using Illumina HiSeq X (Table S13). 'MPK-1' and 'Micro-Tom' yielded 10 111 424 and 8 807 810 raw reads, respectively. For the F_2 population's 94 individuals, a total of 357 692 102 raw reads were obtained, averaging 3 805 235 reads per individual, using dpMIG-seq:PS1_4–5. The linkage map was established using detected SNPs/indels, resulting in a total length of 1165.59 cM with an average marker spacing of 2.03 cM (Figure 6a; Table S14). Markers were sourced from the entire genome, yet exhibited bias (Figure 6a,b; Figure S14b, Table S14, Data S5). Notably, genotyping array results also indicated marker bias across chromosomes (Table S14), likely attributed to sequencing differences between methods. However, QTL analysis revealed a peak with an LOD score of 20.37 on chromosome 1, consistent with the previously identified *Pat-k* locus (Figure 6c,d). Tukey–Kramer comparisons confirmed significance at the peak markers (Figure 6e), suggesting that dpMIG-seq yielded results aligning with those of the genotyping array.

Construction of the dpMIG-seq library and QTL analysis performed without DNA purification in soybean recombinant inbred lines

Finally, we investigated the potential for constructing a linkage map and conducting QTL analysis for flowering time using unpurified DNA in soybean recombinant inbred lines (RILs) derived from a hybrid between 'Misuzudaizu' and 'Moshidou Gong503' (Table S15). Phenotypic data from 1999, registered in the Legume base NBRP (https://legumebase.nbrp.jp/legumebase/top.jsp) were employed for the QTL analysis. Notably, transgressive segregation was observed in the flowering time of these RILs (Figure S15a). Illumina HiSeq X sequencing yielded 10 743 320 and 12 481 870 raw reads for 'Misuzudaizu' and 'Moshidou Gong503', respectively (Table S15). Library construction revealed an average of 2 460 481 reads, with only four samples falling below 1 million reads (563 390–791 754 reads). The remaining samples produced read counts of 1 051 098–4 482 752 (Table S15). Linkage map construction resulted in a map containing 1765 markers encompassing 5109 SNP/indel markers (Figure 7a; Table S16). The total length of this linkage map was 2114.52 cM, with an average intermarker distance of 1.94 cM. Markers were distributed genome-wide, and physical and genetic distances generally corresponded (Figure 7b; Figure S15b, Data S6). QTL analysis for flowering time revealed LOD score peaks at the *E1* and *E2* loci (Tsubokura et al., 2014; Watanabe et al., 2011; Xu et al., 2015), with specific LOD scores of 28.60 and 10.06, respectively (Figure 7c,d). An analysis of variance, using only homozygous *E1* and *E2* nearest markers, indicated significant differences in mean flowering time between genotypes (Figure 7e).

DISCUSSIONS

In this study, we introduced dpMIG-seq, a PCR-based method for constructing NGS libraries using multiplexing primers, where the degenerate oligonucleotide replaces the SSR region found in the MIG-seq primers used for the first PCR. Performance evaluation of dpMIG-seq in tomato revealed an expanded loci coverage due to versatile primer annealing. Suyama et al. (2022) previously altered the annealing temperature of the first PCR of MIG-seq from 48°C to 38°C for similar enhancement, and we adopted 38°C as the annealing temperature in this study, leading to a stable increase in the number of polymorphisms. This is because if the annealing temperature is much lower than the Tm value of the primers, even slight sequence variations can be annealed sufficiently to achieve stable PCR amplification and genome-wide loci amplification simultaneously. However, the primer set PS1_4–9, incorporating numerous degenerate oligonucleotides, failed to validate the first PCR amplification, resulting in a low mapping rate. Therefore, it is imperative to appropriately define the position and number of degenerate oligonucleotides. In this study, the introduction of the degenerate oligonucleotide was positioned on the SSR of the MIG-seq primer. Nevertheless, similar effects could be expected if a primer with a comparable structure is designed using a sequence other than the SSR (such as a primer where one of the three bases from the 3′ end is replaced with degenerate oligonucleotides). Further investigation in this area is warranted.

In our assessment of 11 crops, dpMIG-seq demonstrated superior locus sequencing compared with conventional MIG-seq:PS1 for plant species with a genome size of ≤2.753 Gb at a data volume of 0.3 Gb. However, for tetraploid and hexaploid wheat, the performance of dpMIG-seq was comparable with MIG-seq:PS1 or yielded fewer SNPs. When additional data was obtained, more polymorphisms were detected by dpMIG-seq:PS1_4–5 than by MIG-seq:PS1, suggesting that the inability to detect wheat polymorphisms with a data volume of 0.3 Gb was due to limited raw reads. For low-individual, high-data scenarios, such as CAPS marker design and genetic constitution confirmation

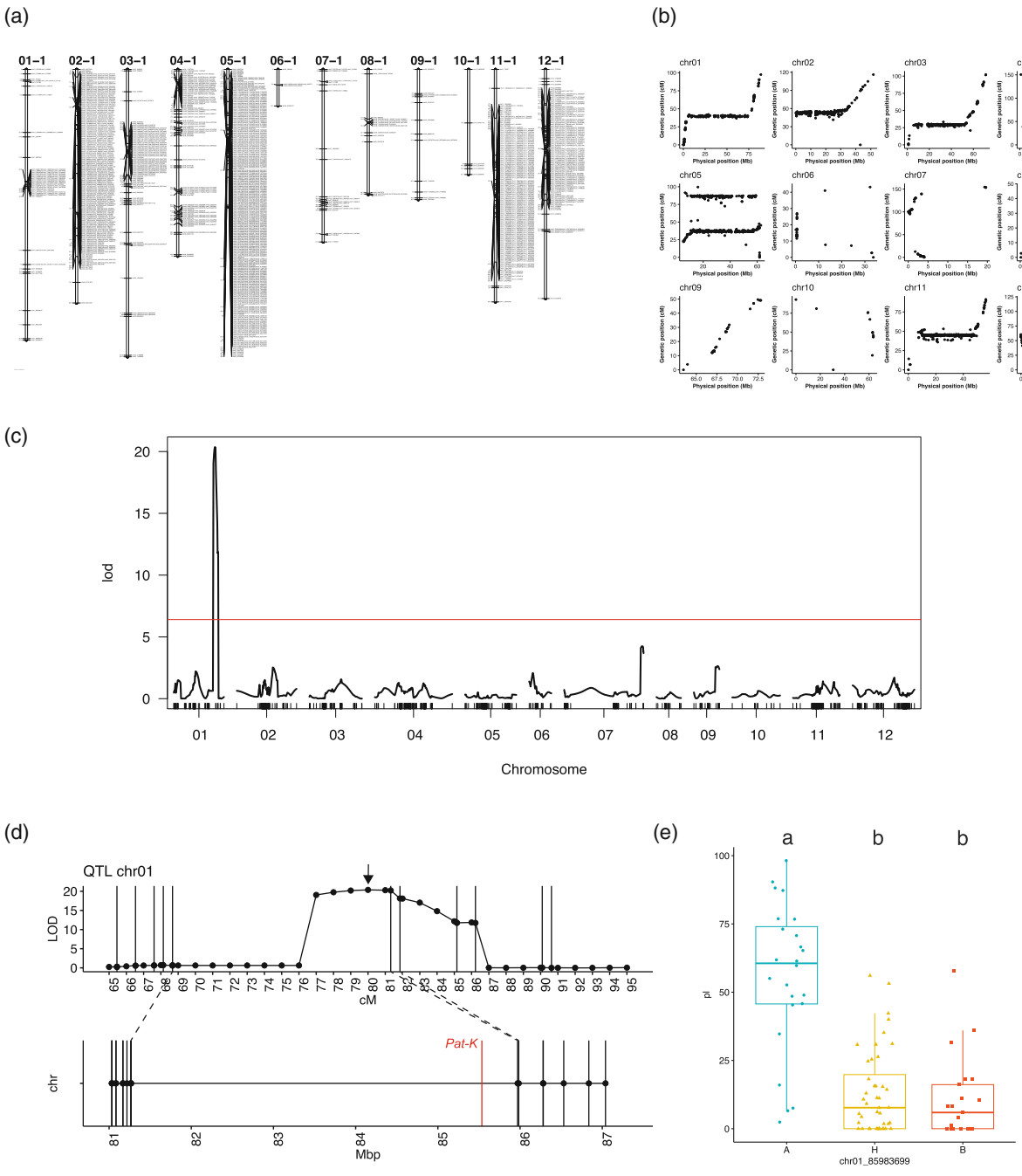


Figure 6. Construction of linkage map and quantitative trait loci (QTL) analysis for parthenocarpy level in tomato F_2 populations via dpMIG-seq. (a) Linkage map of the tomato F_2 population constructed via dpMIG-seq, where the tomato F_2 population resulted from a cross between 'MPK-1' and 'Micro-Tom'. (b) Location of each marker on the physical and linkage maps, with the *x*-axis and *y*-axis representing the position on the physical map and genetic map, respectively. (c) LOD score for QTL analysis based on parthenocarpy level, with the red line indicating the LOD score threshold. (d) Comparison of locus positions and candidate genes on physical and genetic maps of QTLs on chromosome 1 for parthenocarpic levels of the tomato F_2 population, with the dotted lines representing the same markers on the physical map and genetic map, and marker correspondence by dotted lines being confined to the vicinities of the QTL peak or known parthenocarpic genes. (e) QTL effect for parthenocarpic levels in the tomato F_2 population. Parthenocarpic level (pl) was classified using markers close to the detected QTL peaks, with A, B, and H indicating the homozygous 'MPK-1' allele, homozygous 'Micro-Tom' allele, and heterozygous allele, respectively. Different lowercase letters indicate significant differences.

in large-genome wheat, dpMIG-seq can be an effective method. However, for genetic analysis of >100 individuals, costs might be restrictive. Therefore, for wheat and its relatives, GBS/ddRAD-seq (Li et al., 2015; Yang et al., 2017), GRAS-Di (Miki et al., 2020), DArTseq (He et al., 2022; Semagn et al., 2022), and MIG-seq (Nishimura

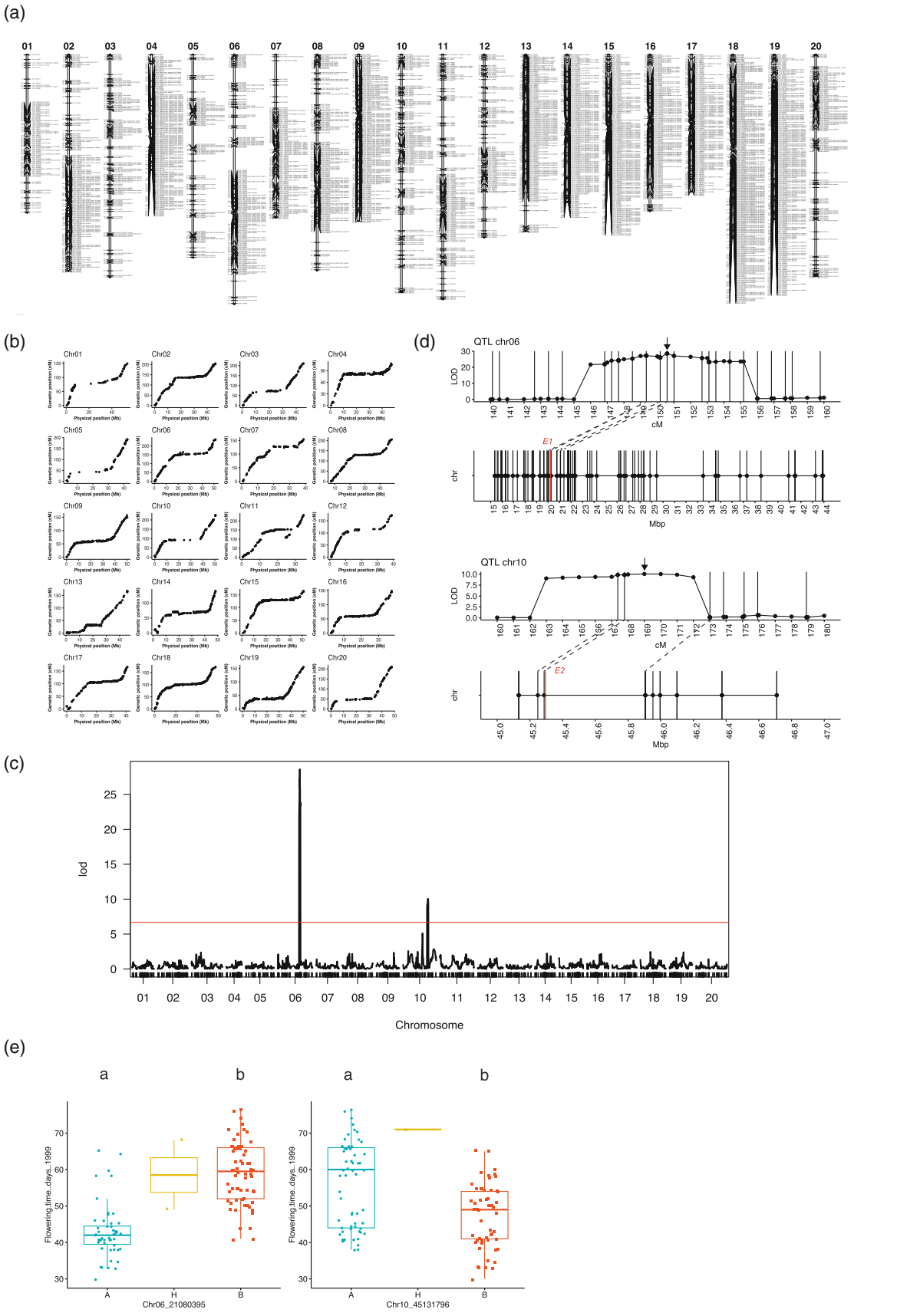


Figure 7. Construction of linkage map and quantitative trait loci (QTL) analysis for flowering time in soy recombinant inbred lines (RILs) via dpMIG-seq without DNA purification. (a) Linkage map of soy RILs generated via dpMIG-seq, where soy RILs were derived from a cross between 'Misuzudaizu' and 'Moshideu-Gong 503'. (b) Location of each marker on the physical and linkage maps, with the *x*-axis and *y*-axis representing the position on the physical map and genetic map, respectively. (c) LOD score for QTL analysis based on flowering time, with the red line indicating the LOD score threshold. (d) Comparison of locus positions and candidate genes on physical and genetic maps of QTLs on chromosomes 6 and 10 for the flowering stage of soy RILs, with dotted lines showing the same markers on the physical map and genetic map, and marker correspondence by dotted lines being limited to the vicinities of the QTL peak or known flowering genes. (e) QTL effect for the flowering stage of soy RILs categorized by markers close to the detected QTL peaks, with A, B, and H indicating the homozygous 'Moushidou-Gong503' allele, homozygous 'Misuzudaizu' allele, and heterozygous allele, respectively. Heterozygous genotypes were excluded from the statistical test owing to low numbers of individuals. Different lowercase letters indicate significant differences.

et al., 2022) remain viable options for genetic analysis. However, owing to the simplicity of dpMIG-seq and MIG-seq library construction, dpMIG-seq may be favored for high-density wheat linkage maps if NGS costs decrease.

In this study, we confirmed that dpMIG-seq, akin to MIG-seq, is prone to errors. Therefore, it is essential to account for these errors when using MIG-seq or dpMIG-seq polymorphism detection for genetic analysis. For example, during the construction of linkage maps, markers in the F_2 population can be readily filtered based on their adherence to the 1:2:1 segregation ratio. Moreover, as highlighted by Suyama and Matsuki (2015), a sequencing approach involving two iterations and exclusively utilizing data from matched loci could also be effective.

Generating MIG-seq-based linkage maps in small-genome crop species was challenging; however, in rice, tomato, and soy, dpMIG-seq effectively generated linkage maps with a substantial number of markers: 5142, 3678, and 5109, respectively. The successful construction of linkage maps validated the consistent sequencing of the same locus across the samples. Furthermore, previously reported genes were found in close proximity to markers exhibiting peak LOD scores, affirming the capability of dpMIG-seq for high-density, accurate genotyping conducive to QTL analysis. However, in a previous study, a linkage map consisting of 9303 markers was generated using ddRAD-seq in a rice population derived from an indica and japonica cross (de Leon et al., 2020). Although the number of polymorphisms in dpMIG-seq of this study was lower than in that of ddRAD-seq, there is room for further improvement to detect polymorphisms comparable to those in ddRAD-seq, such as comparison of the amount of data to be acquired and the position of degenerate oligonucleotide substitution, which has not yet been tested. Since sufficient polymorphisms were detected in this analysis of rice, there is an advantage of genotyping by dpMIG-seq for some materials, since DNA concentration measurement and purification are unnecessary. Additionally, the PCR-based nature of dpMIG-seq makes it potentially applicable to fungi and animals, similar to ddRAD-seq (Peterson et al., 2012), widening its immediate applicability across various research domains.

Our investigation revealed the feasibility of dpMIG-seq library construction using unpurified DNA, indicating that it does not require high-quality DNA, which is a hallmark of MIG-seq, as originally described by Suyama and Matsuki (2015). MIG-seq has been used in several ecological and taxonomic studies since its development (Cho et al., 2021; Hirota et al., 2021, 2022; Hoshino et al., 2021; Nakajima et al., 2021; Saito et al., 2022; Toji et al., 2022; van Ngoc et al., 2021; Yahara et al., 2021a, 2021b), although <1000 SNPs were available in some studies. The dpMIG-seq may serve as a useful alternative to conventional MIG-seq in future ecological and taxonomic studies requiring a higher marker count, especially for clonal identification or species with small genomes.

In this study, we developed an innovative methodology with a rapid protocol for constructing libraries with adjustable sequenced regions, similar to ddRAD-seq, and demonstrated its properties using various crops. Notably, the capacity of dpMIG-seq to generate libraries from unpurified DNA promises to streamline plant genome analysis, allowing for unprecedented speed and ease in genetic research across a diverse spectrum of plants, including crucial crop species.

EXPERIMENTAL PROCEDURES

Plant materials

Two tomato cultivars, 'MPK-1' and 'Micro-Tom', were used to assess the performance of dpMIG-seq (Tables S2 and S3) with diverse multiplex primer sets (Table S3). For broader dpMIG-seq performance evaluation, we used two rice cultivars, two soy accessions, two tetraploid wheat accessions, two hexaploid wheat cultivars, two quinoa accessions, two capsicum cultivars, two peach cultivars, two melon cultivars, two radish cultivars, and two blueberry cultivars (Tables S2 and S4). The durum wheat cultivar 'Setodur' and emmer wheat accession 'TN26' were used to (i) compare MIG-seq and dpMIG-seq under high raw read data volume, and (ii) generate *VRN-A3* gene (Nishimura et al., 2018; Yan et al., 2006) NILs (Table S7).

The feasibility of constructing dpMIG-seq libraries using unpurified DNA and altering the first PCR enzyme was assessed using eight rice and eight tomato varieties (Tables S8 and S9). To extend the application of the first PCR method to multiple crop species without DNA purification, eight blueberry, four citrus, four cabbage, four persimmon, and four rose varieties were used (Tables S6 and S7).

Linkage map construction and QTL analysis for DTH, TK, and T65 rice cultivars and 130 F_2 individuals from their cross were used (Table S11). Linkage map construction and QTL analysis for

parthenocarpic level were performed using 'Micro-Tom' and 'MPK-1' tomato cultivars, along with 94 F_2 individuals from their cross (Table S13). This F_2 population was previously analyzed by Takisawa et al. (2018) for QTL evaluation using the Axiom® Tomato Genotyping Array (TAKARA Bio Co. Ltd.). Linkage map construction and QTL analysis for flowering date in soy were performed using 'Misuzudaizu' and 'Moshidou Gong 503' cultivars, along with 121 RILs from their cross (Table S15).

Cultivation conditions for rice F_2 population

Seeds of TK, T65, and the 130 F_2 individuals were germinated at 20°C for 24 h, followed by 30°C for 48 h in darkness. On May 17, 2021, the seeds were sown in pots (15.0 × 12.5 cm) filled with soil and placed in a greenhouse. On June 8, 2021, seedlings were transplanted to a paddy field at the Kizu Experimental Farm of Kyoto University at Kizugawa, Japan (34°73′ N, 135°84′ E). Seedlings were transplanted with 30 cm spacing between rows and plants. Rice material heading dates were recorded for QTL analysis. Temperature and day length data obtained during the growing season are provided in Figure S16.

Primer design for dpMIG-seq

To increase the number of sequenced regions in MIG-seq, some primer sequences in the first PCR were altered with Ns (A, T, G, C), creating primer sets of dpMIG-seq. Given that Suyama and Matsuki (2015) designated the two bases at the 3′ end of the MIG-seq primers as anchor sequences, these two sequences were not replaced with N in the present study to ensure PCR stability. Although primers with Ns have slightly varied sequences, the 38°C annealing temperature in the first PCR ensured stable amplicon production, despite sequence differences. Validation of dpMIG-seq involved designing 13 primer sets with incomplete SSR regions, introducing Ns into MIG-seq primers originally developed by Suyama and Matsuki (2015) (Table 1; Table S1). Additionally, another MIG-seq primer set (named 'PS2') were designed to target SSRs not targeted by the MIG-seq primers of Suyama and Matsuki (2015) (Table 1, Table S1). PS2 primers were used to determine whether replacing SSR bases with Ns had a generalized effect on the number of detectable polymorphisms. Naming conventions for dpMIG-seq primer sets combined (i) the original primer name, (ii) an underscore, and (iii) the substituted base position. Ranges were used for multiple Ns, for example, '3–4'.

Library construction method for MIG-seq and dpMIG-seq

For MIG-seq and dpMIG-seq library construction, we followed a slightly modified protocol based on the MIG-seq library construction of Nishimura et al. (2022). Libraries were constructed using 16 multiplex primer sets (Table S1), Multiplex PCR Assay Kit ver. 2 (TAKARA Bio Co. Ltd.), and DNA extracted from each sample. DNA extraction methods for the 11 crops are outlined in Tables S2 and S7. First PCR was performed under the following conditions: 94°C for 1 min; 25 cycles at 94°C for 30 sec, 38°C for 1 min, and 72°C for 1 min; and a final extension at 72°C for 10 min. Prime Star GXL DNA Polymerase (TAKARA Bio Co. Ltd.) was used with the first PCR product diluted 50-fold and the second PCR primers, that is, those developed by Nishimura et al. (2022) (Table S17), which facilitated the second PCR under the following conditions: 98°C for 30 sec; 20 cycles of 98°C for 10 sec, 54°C for 15 sec, and 68°C for 30 sec; and a final extension at 72°C for 10 min. Second PCR products were pooled, purified using AMPure XP (Beckman Coulter, Inc., USA), and subjected to reconditioning PCR (eliminating heterodimeric strands in a single PCR cycle) for accurate size selection under the following conditions: 98°C for 40 sec, 54°C for 15 sec, 68°C for 30 sec, and a final extension at 72°C for 10 min. Reconditioning PCR was conducted using Prime Star GXL DNA Polymerase (TAKARA Bio Co. Ltd.), Illumina Primer P1 (5′-AATGATACGGCGACCACCGA-3′), and Illumina Primer P2 (5′-CAAGCAGAAGACGGCATACGA-3′). Following reconditioning PCR, the library was purified using AMPure XP and size-selected using SPRIselect (Beckman Coulter, Inc., Brea, CA, USA). To remove small and large fragments (right and left side selection, respectively), the library sample to SPRIselect ratios were 1.00:0.75 and 1.00:0.56, respectively. For all constructed libraries, an Illumina HiSeq X generated 151-bp paired-end reads with 20% Illumina PhiX control libraries. FASTQ files from multiple library constructions of the same sample were merged for analysis (Tables S3, S4 and S7).

For dpMIG-seq library construction of the rice F_2 population and its parents (TK and T65), DNA extraction from leaves was achieved following the method of Zheng et al. (1995). The dpMIG-seq library was constructed using the above-described method with PS1_4–5 primers. Regarding the tomato F_2 population and its parents, DNA was extracted using the DNeasy Plant Mini Kit (Qiagen GmbH, Germany) as described by Takisawa et al. (2018).

The procedures for library construction, including multiple DNA extraction methods and enzymes for the first PCR, were evaluated using materials from tomato, rice, and blueberry (Table S8). Based on the method of Jia et al. (2021), filter paper was dried after soaking in SDS and Tris-EDTA buffer, and then used to rub leaf blades for liquid absorption. In the two one-step buffer-based methods, crushed leaves in liquid nitrogen mixed with buffer (200 μl) were heated to 90°C for 10 min, the lysate was centrifuged, and the supernatant was diluted tenfold using sterile distilled water for use as a template in first PCR. For the first PCR, PS1_4–5 was used as described above, but 2% Tween 20 was included in the solution. The experimental steps used after the first PCR were identical to those described above. For soy RILs, DNA was extracted using a single-step protocol buffer with PVPP, and first the PCR was performed using the primer set PS1_4–5. For dpMIG-seq using KOD One® PCR Master Mix (Toyobo Co. Ltd.), the first PCR conditions were as follows: 98°C for 1 min; 25 cycles at 98°C for 10 sec, 38°C for 10 sec, and 68°C for 10 sec; and a final extension at 68°C for 5 min. Library construction conditions following the first PCR were the same as those described above. Library construction of dpMIG-seq:PS1_4–5 using a one-step buffer-based approach with PVPP was conducted in persimmon, citrus, rose, and cabbage, following the procedure described above.

Bioinformatics pipeline

Because the initial 17 bases at the 5′ end of each raw read originated from the first PCR primer of MIG-seq and dpMIG-seq, Trimmomatic v.2.0 (Bolger et al., 2014) was used for trimming and filtering, employing the following parameters: 'HEADCROP:17 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15' (Illumina adapter sequences are contained in a FASTA format file at https://github.com/timflutre/trimmomatic/blob/master/adapters/TruSeq3-PE-2.fa). Using Burrows–Wheeler Aligner maximal exact match (Li & Durbin, 2009), trimmed reads of the 11 plant species were mapped to the respective reference genomes (Appels et al., 2018; Colle et al., 2019; Garcia-Mas et al., 2012; Hosmani et al., 2019; Jarvis et al., 2017; Kawahara et al., 2013; Kim et al., 2014; Maccaferri et al., 2019; Schmutz et al., 2010; Verde et al., 2013; Zhang et al., 2021; Table S18). SAM files were converted to BAM and sorted using samtools version

1.9 (Li et al., 2009). Variant calling was achieved using samtools 'mpileup' command (Li et al., 2009) with the '-d 0' option for analyses based on the 11 crops, tetraploid wheat cultivars 'Setodur' and 'TN26' and graphical genotyping of tetraploid wheat NILs for *VRN-A3*.

Up to creating variant call format (VCF) files, the bioinformatics pipeline was the same as that described above for experiments using different enzymes and DNA extraction for first PCR. For linkage map construction, the raw reads of rice, tomato, and soy populations, including parental samples, underwent filtering and alignment to the respective reference genomes of each species (Hosmani et al., 2019; Kawahara et al., 2013; Schmutz et al., 2010) using the aforementioned method. Variant calling for these populations was achieved using the samtools 'mpileup' command.

Performance assessment of dpMIG-seq in crop plants

To compare polymorphism numbers between dpMIG-seq and MIG-seq, the coverage depth (DP) of each polymorphism was divided by the amount of raw sequencing data (Gb) and then multiplied by 0.3 as an adjustment to calculate DP per 0.3 Gb. This computation was performed using a custom R script. In both 'MPK-1' and 'Micro-Tom', only polymorphisms with corrected DP ≥10 were analyzed. Sequenced bases were calculated in a similar manner, considering DP ≥10 per 0.3 Gb. Common polymorphisms between MIG-seq and dpMIG-seq were determined by calculating the percentage of overlapping polymorphisms from each primer set pair. To calculate the mapping rate for each primer set, unmapped reads from the BAM file were examined using the SAMtools view command with the -f 4 option. Mapped reads were identified by calculating the number of mapped reads from the FASTQ row count. The mapping rate was then determined by dividing the number of mapped reads by the total number of reads. To determine whether the sequences to which the primers annealed differed among the first PCR primer sets, a GTF file specifying the region to which the primers were supposed to anneal was created using a custom-made R script and the sorted BAM files. Using this GTF file, SeqKit subseq (Shen et al., 2016) was applied to extract 14 bp of sequences from the 5′ end of the region to which each read mapped from the reference genome for each read from the FASTA file of the tomato reference genome. The types of sequences of these 14 bases were cross-tabulated. For the types of sequences that appeared more than 10 times per million reads, the average values of 'MPK-1' and 'Micro-Tom' were calculated and used in the analysis.

To evaluate the usefulness of dpMIG-seq in the 11 plant species, samtools 'depth' command (Li et al., 2009) was used to extract locus DP data from sorted BAM files using the parameter '-d 0'. Nucleotides with ≥10 DP were considered sequences that could be obtained stably using MIG-seq and dpMIG-seq.

Error estimation of dpMIG-seq:PS1_4–5 of tomato

The parents of populations used to construct the tomato linkage map were sequenced multiple times (Table S13), and data from three iterations with sufficient data were used for estimating error rates in 'Micro-Tom' and 'MPK-1'. For each of these iterations, variant calling was performed using SAMtools mpileup and bcftools calls, with filtering and mapping to the reference sequence performed as described above. Genotype discrepancy rates were calculated three times (replicate 1 versus replicate 2, replicate 2 versus replicate 3, and replicate 3 versus replicate 1) for each minimum coverage depth using a custom R script.

Reproducibility and redundancy were determined for MIG-seq and dpMIG-seq amplifications in tomato as follows. Genomic positions of paired-end mapped reads were extracted from the BAM file, and the frequency of paired-end reads mapped positions appearing at least five times per million for both cultivars was tabulated using a custom R script.

Analysis of GC content and genomic distribution of sequences via MIG-seq and dpMIG-seq

To investigate the amplification bias of the GC content in the first PCR during library construction, we analyzed MIG-seq and dpMIG-seq data using 17 primer sets in tomato, as well as MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq:PS1_4–5 for 11 different crop species (including tomato). GC content of the first PCR amplicons for MIG-seq and dpMIG-seq was calculated by extracting positions of forward and reverse reads mapped at paired ends from BAM files. To extract putative amplicon sequences of the first PCR from the reference genome, GTF files were generated via a custom R script for paired-end reads appearing at least five times out of 1 million read pairs in both cultivars or lines Seqkit subseq was employed to extract the sequences from the reference genome, with the output saved as FASTA format files. GC content was calculated for these FASTA files using Seqkit fx2tab with the '-nlg' option. Subsequently, an equivalent number of sequences was randomly extracted from the reference genome to match the same insert size of the observed amplicon sequences. Their GC content was extracted using Seqkit fx2tab and compared through Student's *t*-test analysis.

For the sequences amplified via first PCR, illustrating the dispersion in the genome, the locus positions of these amplicons (i.e., the positions of locations where paired reads were mapped with the smallest physical distance values) were depicted using custom R scripts. The size of the largest gap between adjacent amplicons was tabulated for each of the 11 crop species.

Applicability analysis of dpMIG-seq in tetraploid wheat

To estimate data needs for dpMIG-seq:PS1_4–5 in wheat analysis, 'Setodur' and 'TN26' tetraploid wheat accessions were subjected to MIG-seq:PS1 and dpMIG-seq:PS1_4–5. SAMtools 'mpileup' command was used to generate VCF files, and DP per data volume was calculated for all polymorphisms at 0.5 Gb intervals from 0.5 to 5.5 Gb using a custom R script. Polymorphisms with DP ≥10 were tallied for each data volume to determine the correlation between polymorphism count and data volume.

Wheat NIL selection using genome-wide polymorphisms from MIG-seq and dpMIG-seq

MIG-seq was applied to the BC_4F_2 generation (n = 16) to identify candidate NILs (Data S2). Using dpMIG-seq:PS1_4–5 data between 'Setodur' and 'TN26', two CAPS markers (1B_450454175_AluI and 1B_575563548_MspI) were created for selecting individuals from the BC_4F_3 population in the later generation of BC4F2_10. Forward and reverse primer sequences for 1B_450454175_AluI were 5′-TTTATGCCCCTAGTTGTGTCCC-3′ and 5′-TTGGCCATATCACATCACACGA-3′, and for 1B_575563548_Msp1 were 5′-TGCTCTATGGTAAACACGGCAT-3′ and 5′-TCATCTCCTTCTGGTCCATCCT-3′, respectively. NILs were chosen using a marker set developed for *VRN-A3* promoter region insertion–deletion polymorphisms by Nishimura et al. (2021). BC_4F_3 DNA was extracted using the method of Jia et al. (2021), utilizing filter paper fragments as 'PCR templates. PCR for the *VRN-A3* marker was performed using KOD One® PCR Master Mix (Toyobo Co. Ltd.) under the following conditions: 98°C for 2 min; 35 cycles at 98°C for 5 sec, 60°C for 5 sec, and 68°C for 5 sec; and a final extension at 68°C for 5 min. *VRN-*

A3 PCR products were separated using 4% agarose gel. CAPS marker PCR followed the same *VRN-A3* conditions. PCR products underwent 3-h digestion with AluI and MspI restriction enzymes at 37°C, followed by separation on 2% agarose gel.

Assessment of dpMIG-seq performance using a combination of DNA extraction methods and PCR enzymes

Five combinations of DNA extraction methods and the first PCR enzymes were compared based on the number of detected polymorphisms. First, unmapped reads from the bam file were assessed using the samtools view command with the -f 4 option. Mapped reads were determined by calculating the number of mapped reads from the FASTQ row count. The mapping rate was calculated by dividing the number of mapped reads by the total number of reads. Subsequently, for the eight accessions' VCF files, only polymorphisms with DP ≥10 and no missing data across all individuals were tabulated among the five methods. Extracted polymorphisms were also subjected to PCA, treating heterozygous genotypes in rice accessions as missing values. For eight blueberry, four citrus, four cabbage, four persimmon, and four rose, after DNA extraction using a one-step buffer with PVPP, libraries were constructed using dpMIG-seq:PS1_4–5 primer set, and the mapping rate to the reference genome (Akagi et al., 2020; Colle et al., 2019; Parkin et al., 2014; Raymond et al., 2018; Wang et al., 2017; Table S18) was calculated using the same method as above.

Linkage map construction and QTL analysis in rice, tomato, and soy mapping populations

Linkage maps for rice, tomato, and soy populations were established using Lep-MAP3 (Rastas, 2017). In the rice F_2 population, VCF files underwent initial filtering through VCFtools (Danecek et al., 2011) using the following parameters: '--max-missing 0.95, --minDP 10, --maf 0.25, --minQ 20, --min-alleles 2, --max-alleles 2'. The VCF files were separated by chromosome, and the polymorphisms in these files were grouped using the SeparateChromosome2 module of Lep-MAP3. The linkage group with the most markers in each group defined the corresponding chromosome's linkage group for subsequent analysis. QTL analysis for DTH involved using the composite interval mapping (CIM) method in R/qtl (Broman et al., 2003), with logarithm of odds (LOD) score thresholds established through 1000 permutation tests.

Similarly, in the tomato F_2 population, VCF files underwent initial VCFtools (Danecek et al., 2011) filtering with the following parameters: '—max-missing 0.95, --minDP 10, --maf 0.25, --minQ 20, --min-alleles 2, --max-alleles 2'. Chromosome-specific segregation and grouping of polymorphisms in the separated VCF files were performed using the SeparateChromosome2 module of Lep-MAP3, utilizing the linkage group with the most markers for each chromosome. QTL analysis for DTH was conducted using the CIM method in R/qtl (Broman et al., 2003), with LOD score threshold determination via 1000 permutation tests.

For the soy RILs, VCF files underwent initial VCFtools (Danecek et al., 2011) filtering with the following parameters: '—max-missing 0.95, --minDP 10, --maf 0.25, --minQ 20, --min-alleles 2, --max-alleles 2'. Segmentation of VCF files by chromosome and subsequent polymorphism grouping were executed using the SeparateChromosome2 module of Lep-MAP3, designating the linkage group with the highest marker count as the chromosome's linkage group. QTL analysis for the flowering stage was performed using the CIM method in R/qtl (Broman et al., 2003). Flowering stage data from 1999, used as phenotypes for QTL analysis, were obtained from Legume base NBRP (https://legumebase.nbrp.jp/legumebase/top.jsp). LOD score thresholds were determined using 1000 permutation tests.

Statistical analysis

The significance of the correlation coefficient was evaluated using the cor.test function from the base package in R. To determine the superiority of the difference in GC content between sequences amplified via MIG-seq or dpMIG-seq and randomly sampled sequences from reference genome, a Student's *t*-test was conducted via the t.test function in the R base package. The loci segregating in the BC_4F_2 generation were examined using a chi-square test through the 'chi.test' function in the R 'base' package. Tukey–Kramer tests were conducted using the 'glht' function from the 'multcomp' package in R. ANOVA was conducted using the 'aov' function from the 'base' package in R. Statistically significant results were defined as those with *P*-values ≤0.05.

ACCESSION NUMBERS

All sequence data were submitted to DDBJ, and corresponding accession numbers are provided below. Further details on individual samples and accession numbers can be found in Supplementary Tables. The DRA accession numbers assigned are DRA014713, DRA014715, DRA014716, DRA016895, DRA016896, DRA016897, and DRA016898.

AUTHOR CONTRIBUTIONS

Kazusa Nishimura performed primer design and library construction. Kazusa Nishimura, HK, KM, Kyoka Nagasaka, AY, RT, TK, KU, RN, and TN developed the bioinformatics analysis. Kazusa Nishimura primarily performed the analysis with assistance from the other authors. KM and Kazusa Nishimura led the filter paper-based and crude sample-based DNA extraction experiment. HK cultivated the rice F_2 population and acquired the heading date. HK, HS, MY, and Kazusa Nishimura obtained and organized information on rice flowering genes located within the detected QTLs. RT and Kazusa Nishimura performed tomato-related library construction and QTL analysis. TM and Kazusa Nishimura conducted soybean cultivation and experiments, including DNA extraction and QTL analysis. TN supervised all experiments. Kazusa Nishimura wrote the first draft of the manuscript. All authors agreed to the final version of the manuscript.

ACKNOWLEDGEMENTS

We extend our gratitude to the National BioResource Project-Wheat, which received partial support from the National BioResource Project of MEXT, Japan, for providing the 'Langdon' wheat accession. The durum wheat cultivar 'Setodur' was graciously provided by the Western Region Agricultural Research Center, NARO. NARO Genebank contributed the two hexaploid wheat, radish accessions and seven tomato varieties. Rice cultivars 'Takanari' and 'Taichung 65' were provided by NARO and Kyushu University, respectively. We thank Dr. Kenji Kato (Okayama University) for the melon accessions and the Germplasm Resources Information Network of the US Department of Agriculture for the two quinoa

accessions. Legume Base (http://www.legumebase.brc.miyazaki-u.ac.jp/) at the NBRP in Japan provided the soy RIL seeds. We are deeply thankful to the paddy field team at the Experimental Farm of Kyoto University, especially Mr. Hisashi Kagata, for their invaluable support in rice cultivation.

Financial support for this work was provided by Grant-in-Aid for Early-Career Scientists (20K15502 and 23K13929 to Kazusa Nishimura; 20K15518 to KM), Grant-in-Aid for Scientific Research (C) (22K05630 to Kyoka Nagasaka), and Grant-in-Aid for Scientific Research (A) (22H00368 to RN), all from the Japan Society for the Promotion of Science. This research was also facilitated by CREST (grant number: JPMJCR17O3) from the Japan Science and Technology Agency, and the Science and Technology Research Partnership for Sustainable Development (grant number: JPMJSA1907) from the Japan Science and Technology Agency/Japan International Cooperation Agency.

CONFLICT OF INTEREST

The primer set developed in this study, as well as the obtained tomato, peach, and capsicum data, was included in a Japanese patent application (application number: JP2022-099233 in Japan). TN, Kazusa Nishimura, RN, and KM applied for this patent. The authors declare that there are no other potential conflicts of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Electrophoretic images of the first PCR of MIG-seq and dpMIG-seq in tomato.

Figure S2. Number of polymorphisms detected using dpMIG-seq: PS1_4–5 in tomato and the error rate. (a) Error rate per minimum coverage depth. (b) Number of polymorphisms per minimum coverage depth. In (a) and (b), error bars indicate standard deviation.

Figure S3. Logarithm of frequency per locus for reads mapped via MIG-seq and dpMIG-seq paired-end reads in tomato. Horizontal and vertical axes represent data for 'MPK-1' and 'Micro-Tom', respectively. Logarithmic values are presented on each axis, with a base of 10, for those appearing >5 times, corrected to the frequency per million read pairs of mapped reads. *** indicates a *P*-value for the correlation coefficient of <0.001.

Figure S4. Relationship between the number of polymorphisms detected by MIG-seq and dpMIG-seq in 11 crop species and the minimum coverage depth for variant calls. Vertical and horizontal axes represent the number of polymorphisms and the minimum coverage depth (DP) for variant calls, respectively.

Figure S5. Venn diagram of polymorphisms obtained via MIG-seq and dpMIG-seq in 11 crop species. First PCR was performed using PS1, PS1_4, and PS1_4–5.

Figure S6. GC content of sequences considered amplified through first PCR of MIG-seq and dpMIG-seq, and GC content of sequences randomly extracted from reference genomes in tomato. *** indicates a *P*-value of <0.001 based on Student's *t*-test.

Figure S7. GC content of sequences considered amplified through first PCR of MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq: PS1_4–5 in 10 crop species, and GC content of sequences randomly extracted from reference genomes. *** indicates a *P*-value of <0.001 based on Student's *t*-test.

Figure S8. Distribution of amplicons on the genome in 11 crop species via MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq: PS1_4–5. In each species, the position of amplicons mapped at least 5 times per million paired reads mapped is shown as black lines.

Figure S9. Chi-square test results for markers obtained using MIG-seq in the BC_4F_2 generation derived from a 'Setodur' and 'TN26' cross. The *y*-axis and *x*-axis indicate the *P*-value for conformity to the 1:2:1 segregation ratio according to the chi-square test and the position of each polymorphism on the physical map of the chromosome, respectively.

Figure S10. Electrophoresis images from the investigation of the first PCR conditions of dpMIG-seq in rice and tomato. 'osf01'–'osf08' and 'slf01'–'slf08' indicate eight rice and eight tomato accessions, respectively. Red characters indicate library names. The first PCR was performed using a combination of DNA extraction methods, PCR enzymes, and with or without 2% TWEEN20. 'Thomson buffer' denotes the simple-step DNA extraction method reported by Thomson and Henry (1995). 'One-step buffer with PVPP' denotes a DNA extraction method using the buffer developed by Thomson and Henry (1995) with polyvinylpolypyrrolidone (PVPP). + and − symbols indicate that 2% TWEEN20 was or was not added, respectively.

Figure S11. Electrophoresis images for investigation of 1st PCR conditions of dpMIG-seq in blueberry. Red character indicates library name sequenced in this study. Experiments were performed using a combination of DNA extraction methods, enzymes, and with or without 2 percent TWEEN20 used for 1st PCR.

Figure S12. Electrophoretic images of the first PCR products derived from unpurified DNA from citrus, persimmon, cabbage, and rose. Red characters indicate sequenced library names. Experiments were performed using a combination of DNA extraction methods, enzymes, and with or without 2% TWEEN20 used for first PCR.

Figure S13. Preliminary information on rice QTL analysis via dpMIG-seq. (a) Frequency distribution of markers on the physical map. Dotted red lines indicate chromosomal terminals. (b) Frequency distribution of days to heading in the rice F_2 population. (c) Comparison of locus positions and candidate genes on physical and genetic maps of QTLs on chromosomes 6 and 10 for the heading date of the rice F_2 population. Dotted lines show the same markers on the physical and genetic maps, and marker correspondence according to dotted lines depicts only the vicinities of the QTL peak and/or known flowering genes.

Figure S14. Preliminary information on tomato QTL analysis via dpMIG-seq. (a) Frequency distribution of parthenocarpic level in the tomato F_2 population. (b) Frequency distribution of markers on the physical map. Dotted red lines indicate chromosomal terminals.

Figure S15. Preliminary information on soy QTL analysis via dpMIG-seq. (a) Frequency distribution of flowering time in soy RILs. (b) Frequency distribution of markers on the physical map. Dotted red lines indicate chromosomal terminals.

Figure S16. Temperature and daylength data obtained during the cultivation of the rice F_2 population. (a) Temperature data obtained during rice F_2 population cultivation. *y*-axis and *x*-axis indicate temperature (°C) and the date in 2021, respectively. (b) Daylength data obtained during the rice F_2 population cultivation. *y*-axis and *x*-axis indicate daylength (h) and the date in 2021, respectively. Daylength data were obtained from the National Astronomical Observatory of Japan (https://eco.mtk.nao.ac.

jp/koyomi/topics/). In both graphs, the black vertical line indicates the date of transplanting.

Table S1. First PCR primers for MIG-seq and dpMIG-seq.

Table S2. Summary of the 11 agricultural plant species used for comparison of MIG-seq and dpMIG-seq performance.

Table S3. Summary of NGS data for tomato accessions based on MIG-seq and dpMIG-seq.

Table S4. Summary of NGS data for 10 crop species other than tomato.

Table S5. Genome coverage obtained using MIG-seq:PS1, dpMIG-seq:PS1_4, and dpMIG-seq:PS1_4–5 in 11 crop species.

Table S6. Maximum gap between mapped amplicons of MIG-seq: PS1, dp MIG-seq:PS1_4, and dpMIG-seq:PS1_4–5 in 11 crop species.

Table S7. Summary of NGS data for tetraploid wheat materials used in the analysis evaluating dpMIG-seq and NILs construction.

Table S8. Plant materials employed to develop and validate dpMIG-seq protocols using unpurified DNA.

Table S9. NGS library information for the development of crude sample-based dpMIG-seq.

Table S10. Mapping rates of reads obtained via different first PCR methods using dpMIG-seq.

Table S11. Summarized data for rice F_2 populations and their parents for linkage map construction as well as QTL analysis.

Table S12. Summary of the rice F_2 population linkage map.

Table S13. Summarized data for tomato F_2 populations and their parents for linkage map construction as well as QTL analysis.

Table S14. Summary of the tomato F_2 population linkage map via dpMIG-seq:PS1_4–5 and marker number of linkagemap in the Takisawa et al. (2018).

Table S15. Summarized data for soy RILs and their parents for linkage map construction as well as QTL analysis.

Table S16. Summary of the soy RIL linkage map.

Table S17. Second PCR primers used for MIG-seq and dpMIG-seq.

Table S18. Summary of the reference genomes used in this study.

Data S1. Frequency of sequences assumed to anneal to primers per million in each primer set.

Data S2. Genotypes obtained via MIG-seq in the BC4F2 generation derived from a cross between 'Setodur' and 'TN26'.

Data S3. Genotypes obtained via MIG-seq and dpMIG-seq in 'Setodur', 'TN26', and BC4F3_32.

Data S4. Genotypic information of the rice F_2 population.

Data S5. Genotypic information of the tomato F_2 population.

Data S6. Genotypic information of soy RILs.

REFERENCES

Akagi, T., Shirasawa, K., Nagasaki, H., Hirakawa, H., Tao, R., Comai, L. *et al.* (2020) The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. *PLoS Genetics*, **16**, e1008566.

Amiteye, S. (2021) Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon*, **7**, e08093.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**, e3376.

Bolger, A.M., Lohse, M. & **Usadel, B.** (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Broman, K.W., Wu, H., Sen, Ś. & **Churchill, G.A.** (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889–890.

Cho, M.S., Takayama, K., Yang, J.Y., Maki, M. & **Kim, S.C.** (2021) Genome-wide single nucleotide polymorphism analysis elucidates the evolution of *Prunus takesimensis* in Ulleung Island: the genetic consequences of anagenetic speciation. *Frontiers in Plant Science*, **12**, 706195.

Colle, M., Leisner, C.P., Wai, C.M., Ou, S., Bird, K.A., Wang, J. *et al.* (2019) Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *GigaScience*, **8**, giz012.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dietrich, W.F., Miller, J.C., Steen, R.G., Merchant, M., Damron, D., Nahf, R. *et al.* (1994) A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics*, **7**, 202–245.

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z. *et al.* (2004) Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes & Development*, **18**, 926–936.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A. & **Kawamoto, K.** (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, 19379.

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V.M. *et al.* (2012) The genome of melon (*Cucumis melo* L.). *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 11872–11877.

Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y. *et al.* (2020) Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. *Molecular Plant*, **13**, 1733–1751.

He, X., Rezaul Kabir, M., Roy, K.K., Marza, F., Chawade, A., Duveiller, E. *et al.* (2022) Genetic dissection for head blast resistance in wheat using two mapping populations. *Heredity*, **128**, 402–410.

Hirota, S.K., Yahara, T., Fuse, K., Sato, H., Tagane, S., Fujii, S. *et al.* (2022) Molecular phylogeny and taxonomy of the Hydrangeaserrata complex (Hydrangeaceae) in western Japan, including a new subspecies of H.acuminata from Yakushima. *PhytoKeys*, **188**, 49–71.

Hirota, S.K., Yasumoto, A.A., Nitta, K., Tagane, M., Miki, N., Suyama, Y. *et al.* (2021) Evolutionary history of Hemerocallis in Japan inferred from chloroplast and nuclear phylogenies and levels of interspecific gene flow. *Molecular Phylogenetics and Evolution*, **164**, 107264.

Hoshino, M., Hiruta, S.F., Croce, M.E., Kamiya, M., Jomori, T., Wakimoto, T. *et al.* (2021) Geographical parthenogenesis in the brown alga *Scytosiphon lomentaria* (Scytosiphonaceae): Sexuals in warm waters and parthenogens in cold waters. *Molecular Ecology*, **30**, 5814–5830.

Hosmani, P.S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L.V., Schijlen, E.G. *et al.* (2019) An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv*, 767764.

International Peach Genome Initiative, Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S. *et al.* (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, **45**, 487–494.

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

Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmöckel, S.M., Li, B., Borm, T.J.A. *et al.* (2017) The genome of *Chenopodium quinoa*. *Nature*, **542**, 307–312.

Jia, Z., Ding, M., Nakano, M., Hong, K., Huang, R., Becker, D. *et al.* (2021) Letter to the Editor: DNA purification-free PCR from plant tissues. *Plant & Cell Physiology*, **62**, 1503–1505.

Kajiya-Kanegae, H., Nagasaki, H., Kaga, A., Hirano, K., Ogiso-Tanaka, E., Matsuoka, M. *et al.* (2021) Whole-genome sequence diversity and association analysis of 198 soybean accessions in mini-core collections. *DNA Research*, **28**, dsaa032.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S. *et al.* (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, **6**, 4.

Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A. *et al.* (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature Genetics*, **46**, 270–278.

Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S. & **Shimamoto, K.** (2008) Hd3a and RFT1 are essential for flowering in rice. *Development*, **135**, 767–774.

Konieczny, A. & **Ausubel, F.M.** (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant Journal*, **4**, 403–410.

Kumar, A., **Kumar, S.**, **Singh, K.B.**, **Prasad, M.**, **Thakur, J.K.** & **Kbm, S.** (2020) Designing a mini-core collection effectively representing 3004 diverse rice accessions. *Plant Communications*, **1**, 100049.

de Leon, T.B., **Pruthi, R.**, **Jampala, B.**, **Borjas, A.H.** & **Subudhi, P.K.** (2020) Genetic determinants for agronomic and yield-related traits localized on a GBS-SNP linkage map from a japonica × indica cross in rice. *Plant Gene*, **24**, 100249.

Li, H. & **Durbin, R.** (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, H., **Handsaker, B.**, **Wysoker, A.**, **Fennell, T.**, **Ruan, J.**, **Homer, N.** ***et al.*** (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, H., **Vikram, P.**, **Singh, R.P.**, **Kilian, A.**, **Carling, J.**, **Song, J.** ***et al.*** (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics*, **16**, 216.

Lv, Q., **Li, W.**, **Sun, Z.**, **Ouyang, N.**, **Jing, X.**, **He, Q.** ***et al.*** (2020) Resequencing of 1,143 indica rice accessions reveals important genetic variations and different heterosis patterns. *Nature Communications*, **11**, 4778.

Ma, Z., **He, S.**, **Wang, X.**, **Sun, J.**, **Zhang, Y.**, **Zhang, G.** ***et al.*** (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nature Genetics*, **50**, 803–813.

Maccaferri, M., **Harris, N.S.**, **Twardziok, S.O.**, **Pasam, R.K.**, **Gundlach, H.**, **Spannagl, M.** ***et al.*** (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature Genetics*, **51**, 885–895.

Matsubara, K., **Ogiso-Tanaka, E.**, **Hori, K.**, **Ebana, K.**, **Ando, T.** & **Yano, M.** (2012) Natural variation in Hd17, a homolog of Arabidopsis ELF3 that is involved in rice photoperiodic flowering. *Plant and Cell Physiology*, **53** (4), 709–716.

Miki, Y., **Yoshida, K.**, **Enoki, H.**, **Komura, S.**, **Suzuki, K.**, **Inamori, M.** ***et al.*** (2020) GRAS-Di system facilitates high-density genetic map construction and QTL identification in recombinant inbred lines of the wheat progenitor Aegilops tauschii. *Scientific Reports*, **10**, 21455.

Nakajima, S., **Sueyoshi, M.**, **Hirota, S.K.**, **Ishiyama, N.**, **Matsuo, A.**, **Suyama, Y.** ***et al.*** (2021) A strategic sampling design revealed the local genetic structure of cold-water fluvial sculpin: a focus on groundwater-dependent water temperature heterogeneity. *Heredity*, **127**, 413–422.

van Ngoc, N., **Binh, H.T.**, **Nagahama, A.**, **Tagane, S.**, **Toyama, H.**, **Matsuo, A.** ***et al.*** (2021) Morphological and molecular evidence reveals three new species of Lithocarpus (Fagaceae) from Bidoup-Nui Ba National Park, Vietnam. *PhytoKeys*, **186**, 73–92.

Nishimura, K., **Handa, H.**, **Mori, N.**, **Kawaura, K.**, **Kitajima, A.** & **Nakazaki, T.** (2021) Geographical distribution and adaptive variation of VRN-A3 alleles in worldwide polyploid wheat (Triticum spp.) species collection. *Planta*, **253**, 132.

Nishimura, K., **Moriyama, R.**, **Katsura, K.**, **Saito, H.**, **Takisawa, R.**, **Kitajima, A.** ***et al.*** (2018) The early flowering trait of an emmer wheat accession (*Triticum turgidum* L. ssp. dicoccum) is associated with the cis-element of the Vrn-A3 locus. *Theoretical and Applied Genetics*, **131**, 2037–2053.

Nishimura, K., **Motoki, K.**, **Yamazaki, A.**, **Takisawa, R.**, **Yasui, Y.**, **Kawai, T.** ***et al.*** (2022) MIG-seq is an effective method for high-throughput genotyping in wheat (Triticum spp.). *DNA Research*, **29**, dsac011.

Parkin, I.A.P., **Koh, C.**, **Tang, H.**, **Robinson, S.J.**, **Kagale, S.**, **Clarke, W.E.** ***et al.*** (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biology*, **15**, R77.

Peterson, B.K., **Weber, J.N.**, **Kay, E.H.**, **Fisher, H.S.** & **Hoekstra, H.E.** (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**, e37135.

Rastas, P. (2017) Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*, **33**, 3726–3732.

Raymond, O., **Gouzy, J.**, **Just, J.**, **Badouin, H.**, **Verdenaud, M.**, **Lemainque, A.** ***et al.*** (2018) The Rosa genome provides new insights into the domestication of modern roses. *Nature Genetics*, **50**, 772–777.

Saito, R., **Kondo, N.I.**, **Nemoto, Y.**, **Kumada, R.**, **Nakajima, N.** & **Tamaoki, M.** (2022) Genetic population structure of Wild Boars (Sus scrofa) in Fukushima Prefecture. *Animals*, **12**, 491.

Scheben, A., **Batley, J.** & **Edwards, D.** (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal*, **15**, 149–161.

Schmutz, J., **Cannon, S.B.**, **Schlueter, J.**, **Ma, J.**, **Mitros, T.**, **Nelson, W.** ***et al.*** (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Semagn, K., **Iqbal, M.**, **Crossa, J.**, **Jarquin, D.**, **Howard, R.**, **Chen, H.** ***et al.*** (2022) Genome-based prediction of agronomic traits in spring wheat under conventional and organic management systems. *Theoretical and Applied Genetics*, **135**, 537–552.

Shen, W., **Le, S.**, **Li, Y.** & **Hu, F.** (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, **11**, e0163962.

Shirasawa, K., **Hirakawa, H.** & **Isobe, S.** (2016) Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. *DNA Research*, **23**, 145–153.

Sinn, B.T., **Simon, S.J.**, **Santee, M.v.**, **DiFazio, S.P.**, **Fama, N.M.** & **Barrett, C.F.** (2022) ISSRseq: An extensible method for reduced representation sequencing. *Methods in Ecology and Evolution*, **13**, 668–681.

Suyama, Y., **Hirota, S.K.**, **Matsuo, A.**, **Tsunamoto, Y.**, **Mitsuyuki, C.**, **Shimura, A.** ***et al.*** (2022) Complementary combination of multiplex high-throughput DNA sequencing for molecular phylogeny. *Ecological Research*, **37**, 171–181.

Suyama, Y. & **Matsuki, Y.** (2015) MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Scientific Reports*, **5**, 16963.

Takisawa, R., **Nakazaki, T.**, **Nunome, T.**, **Fukuoka, H.**, **Kataoka, K.**, **Saito, H.** ***et al.*** (2018) The parthenocarpic gene Pat-k is generated by a natural mutation of SlAGL6 affecting fruit development in tomato (*Solanum lycopersicum* L.). *BMC Plant Biology*, **18**, 72.

Tanaka, N., **Shenton, M.**, **Kawahara, Y.**, **Kumagai, M.**, **Sakai, H.**, **Kanamori, H.** ***et al.*** (2020) Whole-genome sequencing of the NARO World Rice Core Collection (WRC) as the basis for diversity and association studies. *Plant & Cell Physiology*, **61**, 922–932.

Tanaka, N., **Shenton, M.**, **Kawahara, Y.**, **Kumagai, M.**, **Sakai, H.**, **Kanamori, H.** ***et al.*** (2021) Investigation of the genetic diversity of a rice core collection of Japanese Landraces using whole-genome sequencing. *Plant & Cell Physiology*, **61**, 2087–2096.

Thomson, D. & **Henry, R.** (1995) Single-step protocol for preparation of plant tissue for analysis by PCR. *BioTechniques*, **19**, 394–397.

Toji, T., **Hirota, S.K.**, **Ishimoto, N.**, **Suyama, Y.** & **Itino, T.** (2022) Intraspecific independent evolution of floral spur length in response to local flower visitor size in Japanese Aquilegia in different mountain regions. *Ecology and Evolution*, **12**, e8668.

Tsubokura, Y., **Watanabe, S.**, **Xia, Z.**, **Kanamori, H.**, **Yamagata, H.**, **Kaga, A.** ***et al.*** (2014) Natural variation in the genes responsible for maturity loci E1, E2, E3 and E4 in soybean. *Annals of Botany*, **113**, 429–441.

Wang, X., **Xu, Y.**, **Zhang, S.**, **Cao, L.**, **Huang, Y.**, **Cheng, J.** ***et al.*** (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genetics*, **49**, 765–772.

Watanabe, S., **Xia, Z.**, **Hideshima, R.**, **Tsubokura, Y.**, **Sato, S.**, **Yamanaka, N.** ***et al.*** (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics*, **188**, 395–407.

Williams, J.G., **Kubelik, A.R.**, **Livak, K.J.**, **Rafalski, J.A.** & **Tingey, S.V.** (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**, 6531–6535.

Xu, M., **Yamagishi, N.**, **Zhao, C.**, **Takeshima, R.**, **Kasai, M.**, **Watanabe, S.** ***et al.*** (2015) The soybean-specific maturity gene E1 family of floral repressors controls night-break responses through down-regulation of FLOWERING LOCUS T orthologs. *Plant Physiology*, **168**, 1735–1746.

Xue, W., **Xing, Y.**, **Weng, X.**, **Zhao, Y.**, **Tang, W.**, **Wang, L.** ***et al.*** (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nature Genetics*, **40**, 761–767.

Yahara, T., **Hirota, S.K.**, **Fuse, K.**, **Sato, H.**, **Tagane, S.** & **Suyama, Y.** (2021a) A new subspecies of Stellariaalsine (Caryophyllaceae) from Yakushima, Japan. *PhytoKeys*, **187**, 177–188.

Yahara, T., **Hirota, S.K.**, **Fuse, K.**, **Sato, H.**, **Tagane, S.** & **Suyama, Y.** (2021b) Validation of Hostaalata (Asparagaceae) as a new species and its phylogenetic affinity. *PhytoKeys*, **181**, 79–93.

Yan, L., **Fu, D.**, **Li, C.**, **Blechl, A.**, **Tranquilli, G.**, **Bonafede, M.** ***et al.*** (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19581–19586.

Yang, C., **Yan, J.**, **Jiang, S.**, **Li, X.**, **Min, H.**, **Wang, X.** ***et al.*** (2022) Resequencing 250 soybean accessions: new insights into genes associated with agronomic traits and genetic networks. *Genomics, Proteomics & Bioinformatics*, **20**, 29–41.

Yang, Z., **Chen, Z.**, **Peng, Z.**, **Yu, Y.**, **Liao, M.** & **Wei, S.** (2017) Development of a high-density linkage map and mapping of the three-pistil gene (Pis1) in wheat using GBS markers. *BMC Genomics*, **18**, 567.

Zhang, X., **Liu, T.**, **Wang, J.**, **Wang, P.**, **Qiu, Y.**, **Zhao, W.** ***et al.*** (2021) Pan-genome of Raphanus highlights genetic variation and introgression among domesticated, wild, and weedy radishes. *Molecular Plant*, **14**, 2032–2055.

Zheng, K., **Subudhi, P.**, **Domingo, J.**, **Magpantay, G.** & **Huang, N.** (1995) Rapid DNA isolation for marker-assisted selection in rice breeding. *Rice Genetics Newsletter*, **12**, 255–257.