Legume Science

WILEY

Legume Science Open Access

TECHNICAL NOTE OPEN ACCESS

Optimization of Genotyping-by-Sequencing (GBS) for Germplasm Fingerprinting and Trait Mapping in Faba Bean

Hailin Zhang[1] | Lavinia I. Fechete[2] | Axel Himmelbach[1] | Anja Poehlein[3] | Ulrike Lohwasser[1] | Andreas Börner[1] | Fouad Maalouf[4] | Shiv Kumar[4] | Hamid Khazaei[5] | Nils Stein[1,6] | Murukarthick Jayakodi[1]

[1]Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) OT Gatersleben, Seeland, Germany | [2]Department of Molecular Biology and Genetics, Aarhus University, Aarhus C, Denmark | [3]Institute of Microbiology and Genetics Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Georg-August-University Göttingen, Grisebachstr. 8, Göttingen, Germany | [4]Biodiversity and Crop Improvement program, International Center for Agricultural Research in the Dry Areas (ICARDA), Beirut, Lebanon | [5]Production Systems, Natural Resources Institute Finland (Luke), Helsinki, Finland | [6]Institute of Agricultural and Nutritional Sciences, Martin-Luther-University of Halle-Wittenberg, Halle (Saale), Germany

Correspondence: Murukarthick Jayakodi (jayakodi@ipk-gatersleben.de)

Received: 3 June 2024 | **Revised:** 17 July 2024 | **Accepted:** 30 July 2024

Funding: This research was supported by the German Leibniz Association to the Leibniz Junior Research Groups (J118/2021/REPLACE).

Keywords: genetic diversity | genotyping | GWAS | *Vicia faba*

ABSTRACT

Faba bean (*Vicia faba* L.) is an important pulse crop with a wide range of agroecological adaptations. The development of genomic tools and a comprehensive catalog of extant genetic diversity are crucial for developing improved faba bean cultivars. The lack of a cost-effective genotyping platform limits the characterization of large germplasm collections, understanding of genetic diversity across populations, and implementing breeder's tools like genomic selection. Genotyping-by-sequencing (GBS) offers high-resolution genotyping for both model and crop plant species, even without a reference genome sequence. The genome fragments targeted by GBS depend substantially on the restriction enzyme (RE) used for the complexity reduction step. Species with complex genomic architecture require optimization of GBS with proper RE to realize the full potential of GBS. Here, we evaluated various REs in the GBS method and identified that the combination of *Ape*KI/*Mse*I proved to be the most appropriate for faba bean based on the best library quality, a high number of genomic loci spread across chromosomes, and high enrichment loci associated with the gene space. With the new optimized protocol, we constructed a genetic map using a recombinant inbred line (RIL) population and identified a QTL for seed hilum color on Chromosome 1. In addition, we also genotyped a diversity panel and performed a genome-wide association studies (GWAS) for important agronomic traits, including plant height (PH), flowering time (FT), and number of pods per plant (PPP). We identified six SNP markers significantly associated with these traits and listed potential candidate genes. The optimized faba bean-specific GBS procedure will facilitate access to the untapped genetic diversity for genetic research and breeding and may facilitate functional genomics.

1 | Introduction

Faba bean (*Vicia faba*), belonging to the tribe Fabeae in the family Fabaceae, is a widely grown cool-season grain legume supplying dietary protein for humans and livestock. Originating in Southwest Asia (Gopher, Lev-Yadun, and Abbo 2021), the faba bean is one of the founder crops of agriculture and is rich in starch, dietary fiber, and micronutrients (Khazaei and Vandenberg 2020). The great ability of faba bean to fix atmospheric nitrogen (Herridge, Peoples, and Boddey 2008), adapt to diverse agroclimatic conditions, support farmland biodiversity (Beyer et al. 2020), and increase cereal crop yield in crop-rotation

systems (Angus et al. 2015) offers unprecedented opportunities to build a sustainable farming system for food and nutrition with fewer environmental footprints. Maximization of genetic diversity, either by capturing natural diversity or induced by mutagenesis, is a key to sustainable faba bean breeding. Currently, the wild progenitor of the faba bean is still unknown, but a wide range of diversity was observed within the cultivated gene pool (Mulugeta et al. 2021). International genebanks hold thousands of faba bean genetic resources, particularly the genebanks of the International Center for Agricultural Research in the Dry Areas (ICARDA), Lebanon, and of IPK Gatersleben, Germany, which hold a total of 9654 and 1860 individual accessions, respectively. Characterizing the extant diversity and corresponding phenotypes would provide an impactful resource for faba bean improvement. However, the lack of a cost-effective genotyping method in faba bean has hampered so far the systematic exploration of genetic diversity from large germplasm collections, high-resolution trait mapping, and genomic selection or prediction.

The faba bean has a very large genome (~13 Gb) with six chromosome pairs ($2n = 12$). In the past, due to a lack of a reference genome sequence, diverse transcriptome datasets were generated to develop genotyping arrays such as "Vfaba_v2" containing 24,929 single nucleotide polymorphism (SNP) markers derived from 15,846 gene loci (Khazaei et al. 2021). This provided an important tool for genetic mapping studies. For the evaluation of broader yet uncharacterized diversity, however, the use of such a SNP array derived from limited genetic diversity would introduce strong ascertainment biases to the analysis. Recently, a high-quality chromosome-level reference genome was generated for faba bean (Jayakodi et al. 2023), and this information was used to design 90,000 oligonucleotide probes to perform single primer enrichment technology (SPET) assay-based genotyping. In addition to known polymorphisms used for assay design, SPET also allows the discovery of novel variation; hence, it is relatively better suited for exploring unknown diversity than SNP-Chip assays. Nevertheless, these probes were generated from a single reference genome, which could also lead to a biased estimation of diversity due to single reference biases (Jayakodi et al. 2021). Particularly, the genotyping cost per sample limits its application in the genotyping of a large number of accessions, such as genebank collections, and is less affordable to smaller labs working with a mapping population or a diversity panel comprising a few hundred genotypes. Therefore, a cost-effective and unbiased genotyping method is highly needed to accelerate faba bean research and breeding.

Genotyping-by-sequencing (GBS) enables the simultaneous discovery and genotyping of a large number of SNPs (Elshire et al. 2011). In crops, GBS was successfully applied to genotyping large genebank collections (Milner et al. 2019), genome-wide association studies (GWAS) (Arruda et al. 2016), linkage map construction (Hussain et al. 2017), quantitative trait locus (QTL) mapping (Mathivathana et al. 2019), and genomic selection (GS) (Battenfield et al. 2016). The GBS method uses restriction enzymes (REs) to target less complex genomic loci for developing genome-wide polymorphic markers, and optimization of proper REs concerning species or genomic architecture is imperative. Improper selection of REs leads to biases toward sequencing exonic or intergenic regions with high proportions (López et al. 2023; Nguyen et al. 2018). In general, methylation-sensitive enzymes (also known as rare cutters that recognize genomic loci occurring only rarely in the genome) such as *Ape*KI or a combination of a frequent cutter that recognizes commonly occurring sites in the genome and methylation-sensitive enzymes like *Msp*I and *Pst*I are often used in most cereal crops (Milner et al. 2019; Wang et al. 2020) and forage legumes (Julier et al. 2021). These enzymes tend to target genic or low-copy regions and thus reduce the representation of repetitive regions in GBS sequencing. Nevertheless, the complex and unique genome characteristics require the identification of appropriate REs to realize the full potential of GBS. Strikingly, faba bean has a large intergenic space (~330 kb) expanded by a significant proportion of satellite repeats and large-size retroelements (up to 35 kb) (Jayakodi et al. 2023). Selection of inappropriate REs might result in a library with incomplete digestions, enrichment for repeats or GC-rich regions, and biased marker coverage along the genome.

In this study, we used single RE (*Ape*KI and *Msl*I) and a combination of REs (*Ape*KI/*Mse*I, *Pst*I/*Ape*KI, *Pst*I/*Mse*I, and *Pst*I/*Msp*I) digestion to optimize the GBS procedure for faba bean. We identified the most optimal RE combination based on the sequencing library quality, highest number of loci, genome-wide coverage, and enrichment for gene space. In addition, we applied our best RE combination (*Ape*KI/*Mse*I) to genotype an advanced biparental mapping population developed from ILB 938/2 × Mélodie/2 and a diversity panel comprising 217 accessions to demonstrate the application of genetic diversity analysis and gene mapping.

2 | Materials and Methods

2.1 | DNA Isolation, GBS Library Construction, and Sequencing

The faba bean germplasm used in this study includes an advanced mapping population along with a diversity set. The mapping population consists of 177 recombinant inbred lines (RILs) developed from the biparental cross of the homozygous genotypes ILB 938/2 × Mélodie/2 (Table S1) (Khazaei et al. 2014) at F8:9 generation. In addition, we have selected a custom subset of 217 accessions covering broad geographical regions of 45 countries from the ICARDA genebank (Table S2). This panel includes mainly domesticated spring faba bean genotypes. Individual faba bean plants were grown outside in the soil using a plastic foil tunnel or pots containing compost soil in a greenhouse with controlled shade and light conditions (Zimmermann et al. 2006) for 2–3 weeks. Three leaf discs (1 cm in diameter) were collected, pooled, and processed for DNA isolation as described earlier (Milner et al. 2019). GBS libraries were constructed following essentially a previously described procedure (Wendler et al. 2014). Genomic DNA (200 ng) was cleaved using *Ape*KI, *Msl*I, *Ape*KI + *Mse*I, *Pst*I + *Mse*I, *Pst*I + *Ape*KI, and *Pst*I + *Msp*I (Table S3). After restriction digestion, samples were used for adapter ligation without purification as described (Wendler et al. 2014). Adapter mixes (P5 and P7, Table S3) compatible with the ends derived from the restriction digestions were used. Adapter ligation to the DNA fragments and SPRI (solid phase reversible

immobilization) purification using MagNa beads (Thermo Scientific Inc., Waltham, MA; Rohland and Reich 2012) were performed as published (Wendler et al. 2014).

DNA was eluted in 20 μL EB (10 mM Tris HCl pH 8.0), and the DNA/bead suspension ("with-bead" SPRI method; Fisher et al. 2011) was used for the adapter fill-in (Wendler et al. 2014). For the subsequent DNA clean-up, 1.8 volumes of PEG buffer (18% PEG 8000, 1 M NaCl, 10 mM Tris HCl pH 8.0, 1 mM EDTA, and 0.05% Tween 20) were added. The purification of the DNA using a magnet plate involved standard procedures. The DNA was eluted in 20 μL EB, and the remaining MagNa beads were discarded. The addition of unique dual indexes by indexing PCR, SPRI purification, quantification of the individual libraries, and equimolar pooling were conducted as described by Wendler et al. (2014). Products of adapter ligation, indexing PCR, and primer positions are shown in Figure S1. Pooled DNA was size-fractionated (targeted size range: 400–600 bp) using the preparative Blue Pippin electrophoresis system and 2% agarose gel cassettes (Marker V2) according to standard manufacturer's protocols (Sage Science, Beverly, MA, United States). The library was quantified by qPCR, as described by Mascher et al. (2013). For the initial comparison of GBS libraries derived from different restriction digestions, an equimolar library pool was prepared and low-pass sequenced by the Illumina MiSeq device (Illumina Inc., San Diego, CA, United States). Final deep sequencing of GBS libraries prepared from *Ape*KI and *Mse*I fragments was performed using the Illumina NovaSeq6000 device (S4 reagent kit: 200 cycles, v1.5 chemistry, standard Illumina sequencing primer; paired-end [PE]: 151 cycles [Read 1], 8 cycles [Index Read 1], 8 cycles [Index Read 2], and 71 cycles [Read 2]) according to the manufacturer's instructions (Illumina Inc., San Diego, CA, United States).

2.2 | GBS Analysis of Diverse Genotypes

The adapter sequences were removed using cutadapt (Martin 2011) before read alignment. The trimmed reads were mapped to the reference genome sequence of faba bean cv. "Headin/2" Version 2 using BWA-MEM v0.7.17 (Li 2013). The resulting binary alignment map (BAM) files were sorted by NovoSort v3.06.05. Variant (SNPs and INDELs) calling was performed with Bcftools v1.8/SAMtols v1.13 (Li 2011). We chose only biallelic sites that have a minimum mapping quality score (QUAL) of 40, as demonstrated by Milner et al. (2019). Principal component analysis (PCA) was conducted with filtered SNPs containing less than 10% missing data and less than 10% heterozygosity.

2.3 | Linkage Map Construction

The genetic map was built with 177 RILs developed from the cross ILB 938/2 × Mélodie/2 population (Khazaei et al. 2014). GBS reads from this population were mapped to the "Hedin/2" Version 2 genome using BWA-MEM v0.7.17 (Li 2013). Variants were called using Bcftools v1.8/SAMtools v1.13 (Li 2011) using a -d 10,000,000 and a minimum mapping quality of 20. The linkage map was constructed using R/qtl package v1.66 (Broman et al. 2003) in R v4.3.2.

2.4 | QTL Mapping of Seed Hilum Color

The seed hilum color, either black or white, scored for the 177 RILs. The QTL for hilum color was mapped using our linkage map and the "scanone" function in R/qtl v1.66 (Arends et al. 2010). The binary model and the marker regression parameters were used. In both cases, *p* values and significance thresholds were generated using a permutation test with 1000 replicates.

2.5 | GWAS Analysis

Our custom panel of 217 genotypes was used for GWAS analysis. The traits of plant height (PH), flowering time (FT), and number of pods per plant (PPP) were recorded in Terbol, Lebanon, for 2 years (2021 and 2022). Each trial was in alpha lattice design with two replicates. Phenotype scores were recorded in three plants of each plot and averaged for each trait. The best linear unbiased estimates (BLUEs) were computed by considering all accessions as fixed terms using Genstat (v2021) software with an incomplete spatial model. The missing genotype calls were imputed using Beagle v5 (Browning et al. 2021). PCA and the analysis of the kinship matrix were performed using the GAPIT v3.0 (Wang and Zhang 2021) function to determine any underlying population structure. GWAS analysis was done with BLINK (Huang et al. 2018) using the imputed SNP matrix. The candidate gene function annotation was identified according to the predicted protein-coding genes blasted against five databases of eggNOG5.0, NCBI NR, and SwissProt.

3 | Results

3.1 | Optimization of GBS for Faba Bean

We tested *Ape*KI and *Msl*I for single digestion and *Ape*KI/*Mse*I, *Pst*I/*Mse*I, *Pst*I/*Ape*KI, and *Pst*I/*Msp*I REs for double digestion method in two genotypes, including FAB 6502 and FAB 6477. We selected these single and double enzyme combinations based on our pilot experiments with various REs (data not shown). First, we evaluated the GBS library construction and observed good library quality, such as a diverse library, no bias, no overrepresentation of repetitive DNA, and an evenly distributed smear of DNA fragments on agarose gel for *Ape*KI, *Msl*I, and *Ape*KI/*Mse*I digestions (Figure S2). Notably, we identified poor library quality, that is, bias and overrepresentation of specific identical sequences, indicated by prominent bands of DNA fragments on agarose gel less data for analysis for libraries that used *Pst*I as one of the REs in double digestion (Figure 1 and Figure S2). Second, we aimed to obtain approximately 2–3 million (M) sequencing reads from each library (Table S4). Third, we mapped each library's reads to the faba bean reference genome "Hedin/2" and calculated the number of loci targeted by each library and depth per loci. Intriguingly, we found that the *Ape*KI/*Mse*I library showed the highest number of loci with a mean read depth of 1.8 (Figure 1 and Figure S3), which could compromise the number of variants (SNPs and INDELs). Furthermore, the distribution of read depth along chromosomes exhibited uniform coverage from

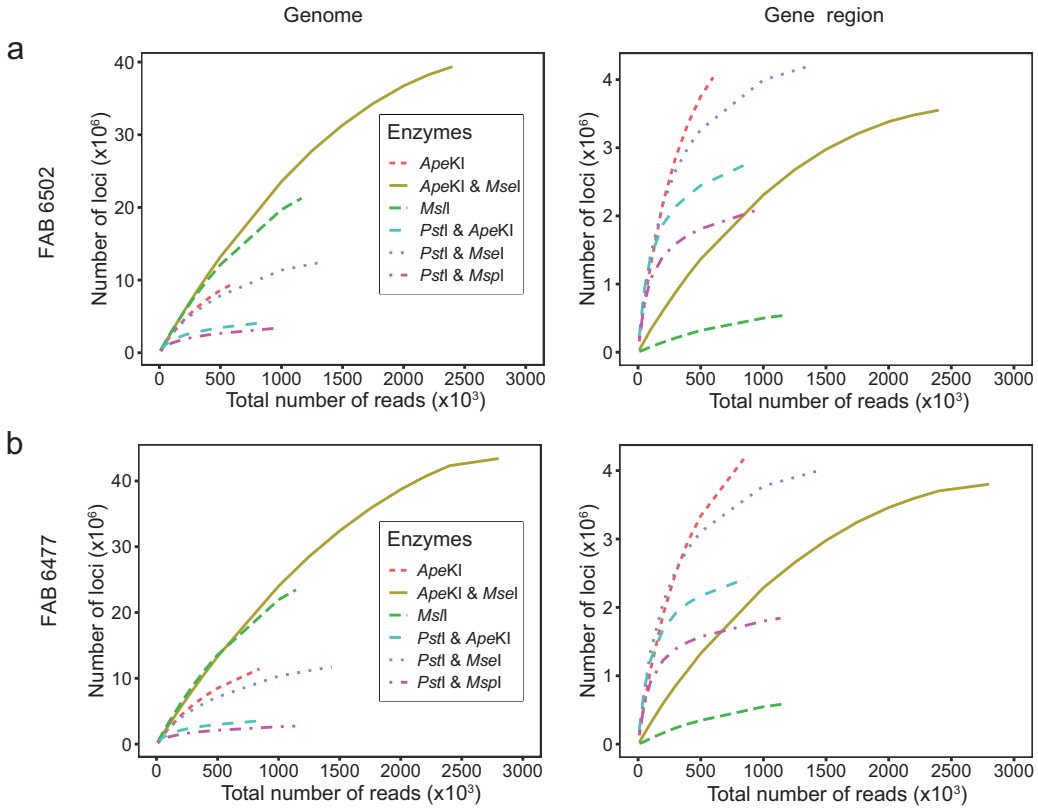

FIGURE 1 | Read depth of GBS reads covering genomic loci produced by the various restriction enzyme (RE) combinations. (a) Single (*Ape*KI and *Msl*I) and double (*Ape*KI/*Mse*I, *Pst*I/*Ape*KI, *Pst*I/*Mse*I, and *Pst*I/*Msp*I) digestion for GBS optimization in faba bean were conducted in FAB 6502 accession. (b) Similar single and double digestion GBS were done in another accession, FAB 6477. The lines represent the number of loci in the genome region covered after the corresponding GBS library.

the *Ape*KI/*Mse*I library (Figure 1 and Figure S3). We also identified the highest number of reads targeting gene space from the *Ape*KI/*Mse*I library. In contrast, *Pst*I/*Ape*KI and *Pst*I/*Msp*I covered fewer genomic loci with a high mean depth of between 6- and 7-fold. Instead of targeting entire chromosomes, these libraries were enriched for fewer genomic loci, thus limiting the development of genome-wide SNP markers. Overall, these results showed that the *Ape*KI/*Mse*I double digestion was the most appropriate for GBS library construction and sequencing to discover variants across the faba bean genome without potential biases in genome reduction. Then, we tested the GBS data yield with regard to SNP missingness. As expected, a lower number of reads in the samples led to higher missing data. However, we observed reduced missing data when the samples were sequenced over 6 M reads (Figure S4a,b), indicating an appropriate read yield required to avoid technical limitations in obtaining high-quality SNPs. Further, due to the high repeat nature of the faba bean genome, single-end GBS read alignment produced a limited number of high-quality alignments, which might lead to numerous false-positive variants. Therefore, we shifted to a new PE approach where Read 1 is 151 bp and Read 2 is 72 bp (151 × 71) (Figure S4c). With this strategy, we achieved better read mappability and numerous high-quality alignments (Figure S4d), and thereby, we expect high-quality variant discovery from faba bean germplasms.

3.2 | Genetic Mapping and QTL Analysis

A filtered set of 30,928 SNPs was generated from the GBS datasets from a ILB 938/2 × Mélodie/2 RIL population. The SNP dataset was then filtered for markers that were polymorphic in the two progenitors. Additionally, the genotype samples with a percentage of heterozygous counts higher than 5% were eliminated for the downstream analysis. After filtering, the resulting dataset contained 1742 markers across 101 individuals. After filtering and quality control, the linkage groups contained a total of 947 GBS markers across 101 individuals, genotyped at a percentage of 90.5%. Finally, we constructed six linkage groups containing 947 markers spanning 1395.2 cm, overlapping with six physical chromosomes of the faba bean (Figure 2a). The average distance between the markers is 1.5 cm, with a maximum distance of 12.6 cm (Table S5). The genetic size of the linkage groups ranged from 156.8 cm for Linkage Group 4 to 432.1 cm for Linkage Group 1. Overall, we observed good collinearity between the genetic map and the physical pseudomolecules of the faba bean genome assembly (Figure S5). Further, we performed QTL mapping for the economically important trait "hilum color." A significant QTL for hilum color was also mapped to Chromosome 1 (Figure 2b). Consistently, the QTL region overlapped with the previously identified candidate genes for seed hilum color (Jayakodi et al. 2023) (Table S6).

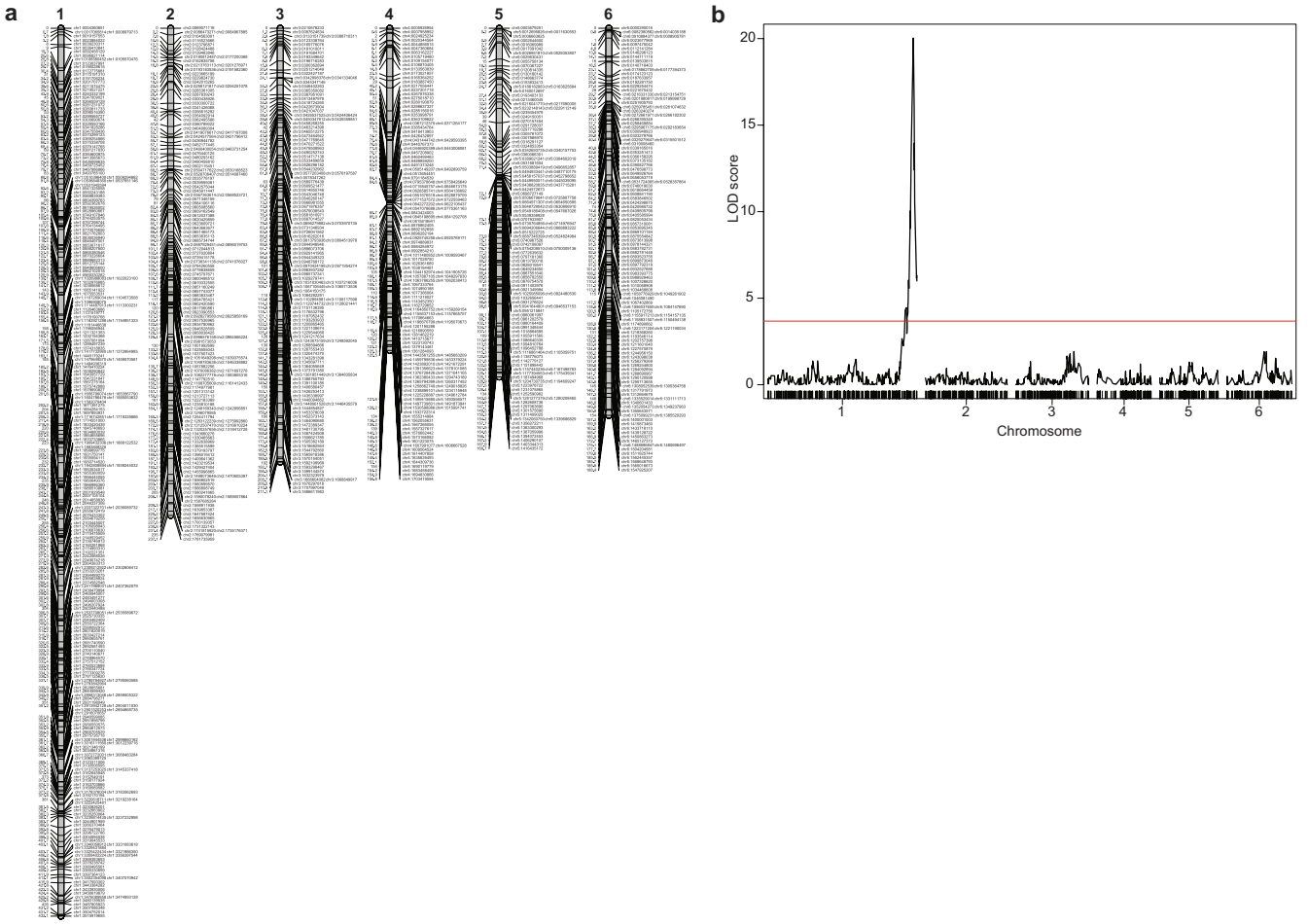

FIGURE 2 | Genetic map and QTL analysis. (a) Linkage map from the ILB 938/2×Mélodie/2 biparental population showing the position of SNPs across the six linkage groups. (b) LOD scores across the chromosomes for the seed hilum color trait. The red line indicates the LOD significance threshold calculated based on 1000 permutations at a 5% significance level. The red line indicates the LOD significance threshold calculated based on 1000 permutations at a 5% significance level.

3.3 | Genetic Diversity Analysis

We also performed a diversity analysis of our custom diversity panel (Figure S6a). After GBS experiments, the reads were mapped to the genome sequence of the faba bean cv. Hedin/2 for variant calling. We identified 39,928 SNPs after removing SNPs containing a missing rate (>10%) and heterozygosity (>10%) (Figure 3a). The number of high-quality SNPs per chromosome ranged from 5045 (chr5) to 11,534 (chr1). With these markers, a PCA identified the distribution of genotypes across PCA edges (Figure S6b and Table S2). In addition, pairwise identity-by-state (IBS) comparison analysis revealed genetic similarity between 88% and 90% (Figure S6c), exhibiting a wide diversity in the panel correlating with the geographic origin. Furthermore, these results indicate that these GBS-based informative markers provide a robust tool to assess the genetic diversity of faba bean.

3.4 | GWAS With Genebank Material

Our dense GBS-derived SNP markers are well suited to detect genetic loci linked to agronomic characters using GWAS. To demonstrate an association scan with our GBS data, we first imputed the missing genotype calls. Then, we performed GWAS with three agronomic traits: "PH," "FT," and "number of PPP." We identified significant maker-trait associations for each trait (Figure 3). Further, the quantile–quantile (QQ) plots for the GWAS supported a strong association between the observed and expected distributions of the p values. We defined the QTL regions based on the LD (r^2 cutoff=0.9) of the associated GWAS marker and listed the genes in those regions (Table S7). Like in numerous previous GBS studies, these results corroborate the utilization of our GBS method in genome-wide association studies in faba bean.

4 | Discussion

High-throughput genotyping platforms are required to advance plant genetic studies and gene cloning. The advent of next-generation sequencing (NGS) technologies led to the development of GBS, which generates a large number of SNP markers for germplasm characterization, genome diversity assessment, and various genetic analyses. RE is employed in GBS, and RE is sensitive to genomic composition; thus,

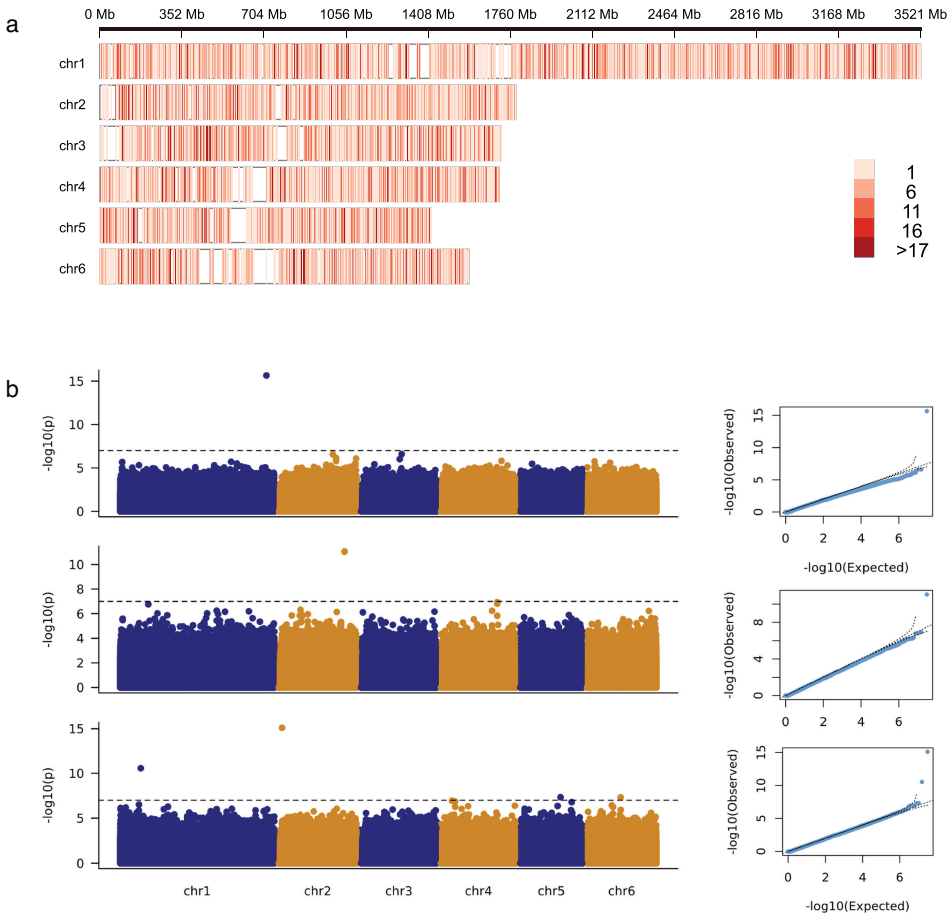

FIGURE 3 | Application of GBS markers in faba bean genetics. (a) Distribution of filtered GBS marker ($n = 39{,}928$) on the chromosomes of the faba bean. The heatmap indicates the marker density across chromosomes. (b) GWAS analysis with GBS makers. The top, middle, and bottom panels show the GWAS results for plant height (PH), flowering time (FT), and number of pods per plant (PPP), respectively. The corresponding QQ plots show the theoretical distribution of *p* values (*x*-axis) versus the observed distribution within the samples (*y*-axis).

species-specific optimization of the GBS procedure may be required in certain species. In principle, GBS is cheaper than medium- or high-coverage whole-genome shotgun sequencing. Heinrich et al. (2020) used *MslI* to obtain GBS data from 20 faba bean genotypes. However, in this study, we have identified an appropriate RE combination, *Ape*KI/*Mse*I. We selected this combination based on the reference genome composition, which showed heavy DNA methylation and a substantial proportion of AT-rich DNA sequences (Jayakodi et al. 2023). Thus, we choose enzymes to recognize the unmethylated portion of the genome and can cut AT-rich regions. In addition, we tested various approaches to increase GBS read mappability and reduce the missingness. Our key recommendations are PE sequencing (151 × 71 cycles) with a data coverage of greater than 5–7-fold per sample. As of 2024, we calculated that the cost of genotyping a single faba bean sample was 20.69 EUR, including library construction and VAT (19%), which is two-fold cheaper than previous genotyping platforms. The cost may also vary in the future based on new developments in library kits and sequencing platforms.

Further, we demonstrated the use of our GBS in QTL mapping and GWAS. The size of the genetic map generated in this study is comparable to other maps constructed using genotyping data from the same population, which range from 928 (Khazaei et al. 2014) to 1229.5 cm (Gela et al. 2022). The candidate genes controlling the hilum color in faba bean were previously identified as a cluster of polyphenol oxidase genes using GWAS and a diverse population (Jayakodi et al. 2023). Here, we remapped a major QTL for this trait in the same region of Chromosome 1 as the cluster of polyphenol oxidase genes. Similarly, the GWAS pinpointed two significant loci linked to the number of PPP traits on Chromosomes 1 and 2. Among them, the Chromosome 1 locus is proximal to the previously associated marker (AX-416763724) (*Vfaba.Hedin2.R2.1g001468.1/Vfaba.Hedin2.R1.1g069400.1* rab GTPase-activating Protein 22), highlighting an important locus for marker-based improvement of yield-related traits in faba bean (Gutierrez et al. 2024). A previous study used a RIL population and reported several QTLs related to FT in faba bean (Aguilar-Benitez et al. 2021). In this study, we used a diversity panel and reported a new locus for FT in the faba bean. Intriguingly, we identified a candidate gene-encoding enzyme, peptidyl-prolyl *cis-trans* isomerase (*Vfaba.Hedin2.R2.2g016227.1*), which was found to be a key regulator of FT in *Arabidopsis* (Wang et al. 2010). Therefore, aside from our GBS method, these markers are valuable to use in marker-assisted breeding.

5 | Conclusion

The availability of the faba bean reference genome enables the characterization of genetic diversity, high-resolution trait mapping, and efficient breeding. Here, we optimized the GBS protocol for broad utility in faba bean research and demonstrated its utility in genetic diversity and QTL mapping. Our GBS genotyping strategy will be a valuable tool to activate faba bean genetic resources in genebanks and accelerate population genomic studies in faba bean.

Author Contributions

Hailin Zhang: formal analysis, validation, visualization, writing – review and editing. **Lavinia I. Fechete:** formal analysis, writing – review and editing. **Axel Himmelbach:** conceptualization, methodology, writing – review and editing. **Anja Poehlein:** methodology, resources. **Ulrike Lohwasser:** data curation, resources. **Andreas Börner:** investigation, resources. **Fouad Maalouf:** phenotyping, resources. **Shiv Kumar:** phenotyping, resources. **Hamid Khazaei:** investigation, resources. **Nils Stein:** methodology, resources, supervision. **Murukarthick Jayakodi:** conceptualization, methodology, validation, visualization, supervision, writing – reveiw and editing.

Acknowledgments

This research was supported by the German Leibniz Association to the Leibniz Junior Research Groups (J118/2021/REPLACE). We also acknowledge Jacqueline Pohl and Katrin Trnka for providing technical support during our sequencing experiments. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All raw GBS sequence data collected in this study have been deposited at the European Nucleotide Archive (ENA). Accession codes for raw data are listed in Table S1 (RIL population) and Table S2 (diversity panel).

References

Aguilar-Benitez, D., I. Casimiro-Soriguer, F. T. Maalouf, and A. M. Torres. 2021. "Linkage Mapping and QTL Analysis of Flowering Time in Faba Bean." *Scientific Reports* 11: 13716.

Angus, J. F., J. A. Kirkegaard, J. R. Hunt, M. H. Ryan, L. Ohlander, and M. B. Peoples. 2015. "Break Crops and Rotations for Wheat." *Crop and Pasture Science* 66, no. 6: 523–552. https://doi.org/10.1071/CP14252.

Arends, D., P. Prins, R. C. Jansen, and K. W. Broman. 2010. "R/qtl: High-Throughput Multiple QTL Mapping." *Bioinformatics* 26, no. 23: 2990–2992. https://doi.org/10.1093/bioinformatics/btq565.

Arruda, M. P., P. Brown, G. Brown-Guedira, et al. 2016. "Genome-Wide Association Mapping of *Fusarium* Head Blight Resistance in Wheat Using Genotyping-By-Sequencing." *The Plant Genome* 9, no. 1: plantgenome2015.2004.0028. https://doi.org/10.3835/plantgenome2015.04.0028.

Battenfield, S. D., C. Guzmán, R. C. Gaynor, et al. 2016. "Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program." *The Plant Genome* 9, no. 2: plantgenome2016.2001.0005. https://doi.org/10.3835/plantgenome2016.01.0005.

Beyer, N., D. Gabriel, F. Kirsch, K. Schulz-Kesting, J. Dauber, and C. Westphal. 2020. "Functional Groups of Wild Bees Respond Differently to Faba Bean *Vicia faba* L. Cultivation at Landscape Scale." *Journal of Applied Ecology* 57, no. 12: 2499–2508. https://doi.org/10.1111/1365-2664.13745.

Broman, K. W., H. Wu, Ś. Sen, and G. A. Churchill. 2003. "R/qtl: QTL Mapping in Experimental Crosses." *Bioinformatics* 19, no. 7: 889–890. https://doi.org/10.1093/bioinformatics/btg112.

Browning, B. L., X. Tian, Y. Zhou, and S. R. Browning. 2021. "Fast Two-Stage Phasing of Large-Scale Sequence Data." *The American Journal of Human Genetics* 108, no. 10: 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

Elshire, R. J., J. C. Glaubitz, Q. Sun, et al. 2011. "A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species." *PLoS ONE* 6, no. 5: e19379. https://doi.org/10.1371/journal.pone.0019379.

Fisher, S., A. Barry, J. Abreu, et al. 2011. "A Scalable, Fully Automated Process for Construction of Sequence-Ready Human Exome Targeted Capture Libraries." *Genome Biology* 12, no. 1: R1. https://doi.org/10.1186/gb-2011-12-1-r1.

Gela, T. S., M. Bruce, W. Chang, et al. 2022. "Genomic Regions Associated With Chocolate Spot (*Botrytis fabae* Sard.) Resistance in Faba Bean (*Vicia faba* L.)." *Molecular Breeding* 42, no. 6: 35. https://doi.org/10.1007/s11032-022-01307-7.

Gopher, A., S. Lev-Yadun, and S. Abbo. 2021. "Breaking Ground: Plant Domestication in the Neolithic Levant: The "Core-Area One-event" Model." Emery and Claire Yass Publications in Archaeology, The Institute of Archaeology, Tel Aviv University.

Gutierrez, N., M. Pégard, I. Solis, et al. 2024. "Genome-Wide Association Study for Yield-Related Traits in Faba Bean (*Vicia faba* L.)." *Frontiers in Plant Science* 15: 1328690. https://doi.org/10.3389/fpls.2024.1328690.

Heinrich, F., M. Gültas, W. Link, and A. O. Schmitt. 2020. "Genotyping by Sequencing Reads of 20 Vicia Faba Lines With High and Low Vicine and Convicine Content." *Data* 5: 63. https://doi.org/10.3390/data5030063.

Herridge, D. F., M. B. Peoples, and R. M. Boddey. 2008. "Global Inputs of Biological Nitrogen Fixation in Agricultural Systems." *Plant and Soil* 311, no. 1: 1–18. https://doi.org/10.1007/s11104-008-9668-3.

Huang, M., X. Liu, Y. Zhou, R. M. Summers, and Z. Zhang. 2018. "BLINK: A Package for the Next Level of Genome-Wide Association Studies With Both Individuals and Markers in the Millions." *GigaScience* 8, no. 2: giy154. https://doi.org/10.1093/gigascience/giy154.

Hussain, W., P. S. Baenziger, V. Belamkar, et al. 2017. "Genotyping-by-Sequencing Derived High-Density Linkage Map and Its Application to QTL Mapping of Flag Leaf Traits in Bread Wheat." *Scientific Reports* 7, no. 1: 16394. https://doi.org/10.1038/s41598-017-16006-z.

Jayakodi, M., A. A. Golicz, J. Kreplak, et al. 2023. "The Giant Diploid Faba Genome Unlocks Variation in a Global Protein Crop." *Nature* 615, no. 7953: 652–659. https://doi.org/10.1038/s41586-023-05791-5.

Jayakodi, M., M. Schreiber, N. Stein, and M. Mascher. 2021. "Building Pan-Genome Infrastructures for Crop Plants and Their Use in Association Genetics." *DNA Research* 28, no. 1: dsaa030. https://doi.org/10.1093/dnares/dsaa030.

Julier, B., S. Blugeon, S. Delaunay, et al. 2021. "Optimisation of GBS Protocols for Efficient Genotyping of Forage Species." Paper presented at the Eucarpia-Section Fodder Crops and Amenity Grasses Meeting.

Khazaei, H., D. M. O'Sullivan, M. J. Sillanpää, and F. L. Stoddard. 2014. "Use of Synteny to Identify Candidate Genes Underlying QTL Controlling Stomatal Traits in Faba Bean (*Vicia faba* L.)." *Theoretical and Applied Genetics* 127, no. 11: 2371–2385. https://doi.org/10.1007/s00122-014-2383-y.

Khazaei, H., D. M. O'Sullivan, F. L. Stoddard, et al. 2021. "Recent Advances in Faba Bean Genetic and Genomic Tools for Crop Improvement." *Legume Science* 3, no. 3: e75. https://doi.org/10.1002/leg3.75.

Khazaei, H., and A. Vandenberg. 2020. "Seed Mineral Composition and Protein Content of Faba Beans (*Vicia faba* L.) With Contrasting Tannin Contents." *Agronomy* 10, no. 4: 511. https://www.mdpi.com/2073-4395/10/4/511.

Li, H. 2011. "A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation From Sequencing Data." *Bioinformatics* 27, no. 21: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs Eith BWA-MEM." arXiv: Genomics. https://doi.org/10.48550/arXiv.1303.3997.

López, A., C. Carreras, M. Pascual, and C. Pegueroles. 2023. "Evaluating Restriction Enzyme Selection for Reduced Representation Sequencing in Conservation Genomics." *Molecular Ecology Resources*. Published ahead of print, September 14, 2023. https://doi.org/10.1111/1755-0998.13865.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads." *EMBnet. Journal* 17, no. 1: 3. https://doi.org/10.14806/ej.17.1.200.

Mascher, M., T. A. Richmond, D. J. Gerhardt, et al. 2013. "Barley Whole Exome Capture: A Tool for Genomic Research in the Genus *Hordeum* and Beyond." *The Plant Journal* 76, no. 3: 494–505. https://doi.org/10.1111/tpj.12294.

Mathivathana, M. K., J. Murukarthick, A. Karthikeyan, et al. 2019. "Detection of QTLs Associated With Mungbean Yellow Mosaic Virus (MYMV) Resistance Using the Interspecific Cross of *Vigna radiata* × *Vigna umbellata*." *Journal of Applied Genetics* 60, no. 3: 255–268. https://doi.org/10.1007/s13353-019-00506-x.

Milner, S. G., M. Jost, S. Taketa, et al. 2019. "Genebank Genomics Highlights the Diversity of a Global Barley Collection." *Nature Genetics* 51, no. 2: 319–326. https://doi.org/10.1038/s41588-018-0266-x.

Mulugeta, B., K. Tesfaye, G. Keneni, and S. Ahmed. 2021. "Genetic Diversity in Spring Faba Bean (*Vicia faba* L.) Genotypes as Revealed by High-Throughput KASP SNP Markers." *Genetic Resources and Crop Evolution* 68, no. 5: 1971–1986. https://doi.org/10.1007/s10722-021-01110-x.

Nguyen, T. K., J. Yu, H. W. Choi, B. C. In, and J. H. Lim. 2018. "Optimization of Genotyping-by-Sequencing (GBS) in Chrysanthemums: Selecting Proper Restriction Enzymes for GBS Library Construction." *Horticultural Science and Technology* 36, no. 1: 108–114. https://doi.org/10.12972/KJHST.20180012.

Rohland, N., and D. Reich. 2012. "Cost-Effective, High-Throughput DNA Sequencing Libraries for Multiplexed Target Capture." *Genome Research* 22, no. 5: 939–946. https://doi.org/10.1101/gr.128124.111.

Wang, J., and Z. Zhang. 2021. "GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction." *Genomics, Proteomics & Bioinformatics* 19, no. 4: 629–640. https://doi.org/10.1016/j.gpb.2021.08.005.

Wang, N., Y. Yuan, H. Wang, et al. 2020. "Applications of Genotyping-by-Sequencing (GBS) in Maize Genetics and Breeding." *Scientific Reports* 10, no. 1: 16308. https://doi.org/10.1038/s41598-020-73321-8.

Wang, Y., C. Liu, D. Yang, H. Yu, and Y. C. Liou. 2010. "Pin1At Encoding a Peptidyl-Prolyl cis/Trans Isomerase Regulates Flowering Time in *Arabidopsis*." *Molecular Cell* 37: 112–122.

Wendler, N., M. Mascher, C. Nöh, et al. 2014. "Unlocking the Secondary Gene-Pool of Barley With Next-Generation Sequencing." *Plant Biotechnology Journal* 12, no. 8: 1122–1131. https://doi.org/10.1111/pbi.12219.

Zimmermann, G., H. Bäumlein, H. P. Mock, A. Himmelbach, and P. Schweizer. 2006. "The Multigene Family Encoding Germin-Like Proteins of Barley. Regulation and Function in Basal Host Resistance." *Plant Physiology* 142, no. 1: 181–192. https://doi.org/10.1104/pp.106.083824.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.